# National Forecasting with Large International Datasets – an Application to New Zealand*

**Sandra EICKMEIER**
**(Deutsche Bundesbank)**

**&**

**Tim NG**
**(Reserve Bank of New Zealand)**

19 March 2009 **(**preliminary)

**Abstract:**

We look at how large international datasets can improve forecasts of national activity and illuminate the nature of international influences on the domestic economy. We use the case of New Zealand, an archetypal small open economy. We apply several data-rich (factor and shrinkage) methods such as principal components, targeted predictors, weighted principal components partial least squares, elastic net, and ridge regression. We systematically assess the predictive content of international data for New Zealand GDP growth, and assess the type of international data that is most relevant for the New Zealand economy. We find that exploiting international information can improve national forecasts compared to smaller models, with largest gains at longer forecast horizons. However, the performance of data-rich methods differs widely with shrinkage methods and partial least squares performing particularly well in our international context. Historical decompositions of forecasts suggest economic interpretations of the drivers of the forecasts that accord with policymakers' narratives about the economic circumstances of the time. National business survey data seem to capture a substantial proportion of the relevant information in developments offshore.

**Key words:** Forecasting, principal components, targeted predictors, weighted principal components, partial least squares, ridge regression, elastic net, international business cycles

**JEL classifications:** C53, F47, C33

* Views expressed in this paper do not necessarily reflect views of the Reserve Bank of New Zealand and the Deutsche Bundesbank. This paper was written while the first author visited the Reserve Bank of New Zealand. We are grateful to Jörg Breitung, Mardi Dungey, Heinz Herrmann, Troy Matheson and Christian Schumacher for helpful suggestions and discussions and to Troy for sharing his national dataset with us. This paper also benefited from discussions with participants at seminars at University of Otago, the RBNZ, the Deutsche Bundesbank, and the Workshop on Modern Macroeconometric Methods and Applications at University of Tasmania and from Victoria Zhang's very valuable research assistance.

## 1. Introduction

Exploiting information from large datasets has been shown to improve forecasts. Some so-called "data-rich" methods such as principal components (PC) are now widely used in economic forecasting and analysis by academics and practitioners. In this paper, we explore PC and other lesser-known data-rich methods using the case of New Zealand GDP growth forecasting. We focus in particular on using large international datasets effectively.

We use New Zealand as an example because it is an archetypal case of a small, open economy with many trading partners across the world, suggesting *a priori* that large international datasets would be potentially useful. Structural vector-autoregressive (SVAR) analyses of New Zealand macroeconomic fluctuations such as Buckle *et al*. (2007) and Dungey and Fry (2008) find that internationally sourced shocks contribute substantially to New Zealand GDP fluctuations. Given New Zealand's economic profile, this is not particularly surprising. Exports account for around 30 percent of New Zealand GDP, with a large proportion of these accounted for by agricultural commodity products. New Zealand is thus highly exposed to the cycle in international commodity prices, such that it (along with Australia and Canada) is often viewed as a "commodity economy" (Chen and Rogoff 2003). Finally, New Zealand also has strong international links through financial markets. Its net foreign liability position is around 100 percent of GDP, while about 30 percent of the assets of the banking system, which is predominantly owned by large Australian parent banks, are funded through issuance in offshore wholesale money markets dominated by US, European and Japanese financial institutions. These trade and financial relationships along with confidence cover, at least in principle, all of the economic channels for the transmission of business cycles internationally that have been examined by the literature on the topic (see e.g. Imbs 2004, Kose *et al*. 2003).

This high dimensionality and complexity of the transmission mechanisms lend themselves to the use of data-rich approaches. Most PC applications use datasets that include some international variables (not very many, compared to what is available), but surprisingly few explicitly assess their specific predictive content. None systematically examines how best to exploit international information for domestic forecasting, or what these methods can reveal about the influence of the international economy on the domestic economy in a forecasting setup.

We tackle this opportunity by asking three main questions, as follows.

(1) By how much can international information improve forecasts of New Zealand GDP growth?

(2) Are data-rich techniques a feasible and useful means of handling the large quantity of available international data manageably, while retaining predictive power?
(3) How can data-rich methods provide a way of capturing transmission channels such that the evolution of the economy may be explained in terms of foreign economic and regional influences?

We look at the first question by comparing forecasting accuracy using national (i.e. New Zealand) data only with that using international data or both national and international data, against benchmark univariate models.

For the second question, we compare different data-rich approaches to capturing the predictive information in an international dataset of several hundred time series covering New Zealand's top 12 trading partners, which represent the three major world economic regions – North America (which accounts for 14 percent of New Zealand's exports and imports), Europe (17 percent) and Asia-Pacific (68 percent)[1]. Data-rich methods reduce the dimension of such large datasets by either shrinking the variance of certain parameters in estimated equations, or by summarising the information contained in many data into a few common factors which makes forecasting using large amounts of information feasible. Of the data-rich methods, factor analysis based on factors estimated by principal components (PC) has so far been most widely applied in economics. We also use other methods which have been suggested to cope with some of the limitations of PC in practice such as targeted predictors (TP) or weighted principal components (WPC), which are variants of PC, partial least squares (PLS), elastic net (EN) and, as a special case of EN, ridge regression (RR). We present these methods in detail below and discuss pros and cons. We compare the forecasts based on data-rich approaches to international data with those based on trade-weighted and other simple weighted-average approaches to summarising international data, which are often used by practitioners due to their simplicity and intuitive appeal. The data-rich methods have, in principle, several advantages in forecasting compared to the use of trade-weighted aggregates. In particular, they are able to handle many more time series, they usually derive weights from an explicit optimization problem, and they (implicitly) capture international transmission channels beyond direct trade linkages, such as indirect trade, financial markets, commodity prices, and confidence.

For the third question, we partition the international dataset broadly in two alternative ways – in terms of economic meaning, and in terms of region of origin. The subsets by economic meaning are real activity, prices, monetary and financial and commodity price data. The

[1] North America contains the US (which accounts for 14 percent in 2007) and Canada (2 percent), Europe the euro area as a whole (13 percent) and the UK (4 percent), and Asia-Pacific includes Australia, Japan, China, Singapore, Korea, Malaysia, Taiwan, and Hong Kong which account for 28, 12, 11, 6, 4, 3, 3, and 1 percent, respectively.

subsets by region are North America, Europe, Asia and Australia. We then assess the marginal predictive content and historical contributions to forecasts of these subsets. The partitioning exercise gives us some sense of the economic and regional nature of foreign spillovers to New Zealand over the sample. The intention of this exercise is to address the criticism sometimes levelled at large forecasting models with little imposed economic structure that they are "black boxes" (Forni *et al.* 2008) and that their output is difficult to interpret and communicate.

The rest of the paper is organised as follows. In section 2 we present the related literature, and in section 3 we describe the data. In section 4 we outline our approach to assessing relative forecasting accuracy, and introduce and discuss the data-rich methods. In section 5 we present our results on forecasting accuracy. In section 6 we look at how the partitioning exercise and data-rich techniques can illuminate the drivers of the forecasts. We conclude in section 7.

## 2. Related literature

Our paper is related to other papers using data-rich methods for national forecasting. Of the data-rich methods, factor analysis based on PC has so far been most widely applied. Many central banks now incorporate factor analysis in their regular forecasting process. In a number of seminal papers, Stock and Watson (2002a, 2002b, 2004) showed the benefits of factor models for economic forecasting. In a recent meta-analysis, Eickmeier and Ziegler (2008) overview 52 studies (including one, Matheson 2006, for New Zealand) that predict output growth and inflation using factor models and which compare factor forecasts with forecasts based on alternative methods. They find that, in general, factor models outperform small-scale models – though gains are sometimes small or not robust across specifications – but that other data-rich methods tend to outperform factor models.

These other methods are much less widely applied and most of them to US data. Boivin and Ng (2006) and Boivin and Ng (2008) suggest and use WPC and TP, respectively. TP is also applied by Schumacher (2009) to German data. De Mol *et al.* (2008) produce their forecasts based on RR and LASSO, another special case of EN. Groen and Kapetanios (2008) and Lin and Tsay (2006) apply RR as well as PLS.

While most papers using data-rich methods include a small set of foreign variables, only very few explicitly investigate the information content of foreign data. Banerjee et al. (2006) investigate the role of euro-area information for central and east European countries (CEECs). Gosselin and Tkacz (2001), Cheung and Demers (2007) and Brisson *et al.* (2003) assess the predictive content of US variables for Canadian real economic activity and inflation. Brisson *et al.* (2003) include other countries' data as well as US data. Finally, Schumacher (2009) considers the role of euro-area and remaining G7 countries for German GDP forecasts.

All these studies compare forecasts based on factors extracted from national datasets with those based on factors estimated either from datasets containing both national and foreign variables, or containing foreign variables only. Banerjee *et al.* (2006), in addition, use in their forecasting equation distinct national and international factors, i.e. factors estimated separately from national and international datasets, as well as factors estimated from the entire dataset. The forecasting accuracy results are, however, qualitatively similar. The Banerjee *et al.* (2006) study finds that euro-area information is of limited value in forecasting the CEECs, and attribute this result to the phase of low correlation between the CEECs and the euro area during convergence. In the studies for Canada, US variables generally appear to contain information beyond that in Canadian data. Brisson *et al.* (2003) find that data from countries outside North America provide useful information for forecasting Canadian inflation, but not Canadian GDP growth. Schumacher (2009) finds that international data is useful only when TP is applied to a joint national and international dataset.

There have been few data-rich approaches taken to New Zealand data in a forecasting context. Matheson (2006) estimated a large-scale factor model to predict GDP and other variables. He estimates factors from a dataset of almost 400 variables. Most variables included in this set are domestic variables but there are also some international variables such as Australian and US interest rates, world GDP growth and CPI inflation, and world oil prices. Bloor and Matheson (2008) use a large Bayesian VAR featuring almost one hundred variables to forecast GDP and other variables, finding its forecasting accuracy to be quite favourable for GDP and some other variables, compared to smaller models such as small- or medium-sized VARs. The large Bayesian VAR includes a handful of international variables, including the weighted-average foreign aggregates used in the RBNZ's forecasting process. However, the information content of the international variables for domestic forecasting is systematically examined in neither of these studies.

## 3. Data

We use quarterly data from 1990Q1 to 2007Q4. Our target variable is New Zealand GDP growth. Both the New Zealand and the international datasets are large. Most of our 446 national series are taken (and updated) from Matheson (2006).[2] Our international dataset

[2] For details on the national dataset see Matheson (2006).

contains 434 series from the 12 major export destinations for New Zealand goods (US, Canada, euro area, UK, Australia, China, Singapore, Hong Kong, Malaysia, Taiwan, Korea and Japan). The breakdown of our international dataset is shown in Table 1. The set comprises roughly 100 North American, 100 European, 200 Asian, and 42 Australian series. There are 206 real activity, 108 price and 130 monetary and financial series. The real block includes national accounts data, industrial production, productivity and labor market variables, orders, retail trade, and survey-based expectations about real economic activity. The price block contains consumer prices, producer prices, the GDP deflator and deflators of the main GDP components, unit labor costs and house prices. The monetary and financial variables block includes interest rates, money and credit aggregates, exchange rates and stock prices. In addition, 38 world commodity price series are included in the dataset. In building the dataset, we attempted to roughly balance the numbers of series in each region and by economic type, subject of course to availability.

We seasonally adjusted by Census X12 the data that were not already seasonally adjusted. We ensured stationarity by differencing where necessary. Following Stock and Watson (2005), we removed outliers, defined as observations of the stationary data with absolute deviations from the median larger than 6 times the interquartile range, replacing them with the median value of the preceding 5 observations. A detailed overview of the international data and the treatment is given in Table A.1.

# 4. Methodology

## 4.1. Forecasting setup

Our benchmark forecasting model for our target variable $y_{t+h,t}$ is the univariate model:

$$y_{t+h,t} = a' \mathrm{W}_t + e_{t+h,t}, \tag{1}$$

where $y_{t+h,t} = y_{t+h} - y_t$, the log difference of New Zealand GDP between period $t+h$ and $t$ and $\mathrm{W}_t$ is a constant or $\begin{bmatrix} 1 & y_{t,t-1} & \cdots & y_{t-p,t-p-1} \end{bmatrix}'$ with $y_{t,t-1} = y_t - y_{t-1}$, and $a$ is the scalar or the $(p+1)\times 1$ vector of associated parameter(s).

Our target variable may also be influenced by other variables, summarized in the $N \times 1$ vector $\mathrm{X}_t$. Introducing $\mathrm{X}_t$ in equation (1) yields

$$y_{t+h,t} = a' \mathrm{W}_t + \Gamma' \mathrm{X}_t + e_{t+h,t}, \tag{2}$$

where $\Gamma = [\Gamma_1 \quad \cdots \quad \Gamma_N]'$ has dimension $N \times 1$. If $N$ is too large, the forecast is not efficient because the sampling variability increases with $N$. If $N > T$, the forecast is even not feasible. In this case, the dimension of $X_t$ can be reduced by means of data-rich techniques which we present in detail below. These methods either shrink the $\Gamma$ parameters to or towards zero, or summarize $X_t$ in a few common factors, or both.

We consider forecasts based on $X_t$ where $X_t$ contains only national data, only international data and both national and international data. We do not estimate national and international factors from separate sets and include them in equation (2) to keep the number of results manageable. Moreover, and as noted, Banerjee et al. (2006) have found that results are qualitatively similar.

Our forecasting setup is as follows. We produce direct, recursive out-of-sample forecasts for horizons 1, 2, 4 and 8 quarters. We compute root mean squared errors (RMSE) between $y_{t+h,t}$ and its forecast $\hat{y}_{t+h,t} = \hat{a}' W_t + \hat{\Gamma}' X_t$, where '^' indicates here and in what follows estimates, and compare them across models. To make RMSEs (roughly) comparable across horizons, we divide them by $h$. In the case where $W_t$ contains lagged target variables, we consider up to 4 lags, i.e. $p = 0,...,3$. For each horizon and each $p$, we compute the corresponding RMSE over the forecast evaluation period and choose the lag length which yields the minimal RMSE for each horizon.[3] Following Matheson (2006) we start the forecast evaluation period in 2000Q1.

## 4.2. Constructing trade-weighted and other simple weighted-average aggregates

In this study, trade-weighted or other simple weighted-average aggregates of foreign variables are used as a benchmark means of treating international information for the purposes of forecasting. Such aggregates are used in, for example, the RBNZ's current forecasting process supporting monetary policy formulation. In that process, they are fed into a large calibrated structural economic model of New Zealand, the Forecasting and Policy System (FPS), to produce domestic forecasts (see Drew and Frith 1998 for an explanation of the RBNZ's forecasting process and core model). Here, we are interested in whether trade- or other simple weighted-average schemes for aggregating international data can be outperformed in our reduced-form forecasting context by data-rich approaches to international data. It is interesting in this context that another popular approach which handles large datasets to model international business cycles, the Global VAR model suggested by Pesaran *et al.* (2004), constructs foreign factors as trade-weighted averages of observable variables, in

[3] This does, of course, not correspond to a real-time forecast situation. Since estimation and forecast evaluation already take long, since it would take much longer to obtain results if we applied information criteria to set the lag length in each period and since this is not the focus of our analysis, we adopt a performance-based approach. For a thorough discussion on performance-based criteria versus information criteria see Schumacher (2007).

a way similar to that of many practitioners, before including them in a VAR, see Dées *et al.* (2007).

We use three variants of trade-weighted aggregates. In the first (*trade1*), we try to mimic the RBNZ's scheme as closely as possible. "Foreign GDP" is produced using two-year moving averages of goods export weights on the GDP of the top 12 trading partners, "foreign CPI" using weights calculated as two-year moving averages of goods import weights of the five trading partners (US, Australia, euro area, Japan and UK), and "foreign interest rates" using fixed 80:20 percent weights on US and Australian interest rates respectively. These weights reflect RBNZ estimates and judgements about the appropriate weights given the large structural model setting in which the international information is used.

In the second variant, we use total trade, measured as the sum of export and import values, weights for constructing foreign GDP, CPI and interest rate variables, following the GVAR applications which also use export plus import weights. Foreign GDP and CPI are calculated using trade weights on all 12 major trading partners, and long- and short-term interest rates using trade weights on the 5 trading partners in the TWI (due to data availability). In *trade2* we employ fixed trade weights averaged over the entire period (*trade2*). The third variant *trade3* is the same as *trade2* except that we use two-year moving averages of trade shares where the first two years are initialized to be the same.

In the forecasting exercise, we include all (output, inflation and interest rate) weighted-average aggregates simultaneously in our forecasting regression. In the section which analyses the predictive content of subsets of data, we also include them individually.

## 4.3. Data-rich methods

As noted, data-rich methods reduce the dimension of $X_t$ by either directly carrying out a regression of the target variable on the predictors where the parameters are shrunk towards or to zero (subsection 4.3.1.) or by summarising the information in a few common factors and regressing the target variable on the factors (subsection 4.3.2.).

### 4.3.1. Shrinkage

*Ridge regression*

The EN and the RR directly regress the target variable on the predictors. For convenience, we neglect here and when presenting the other data-rich methods $W_t$, but account for it at each recursion exploiting the Frisch-Waugh-Lovell theorem. Also, we standardize the predictors to have zero mean and unit variance at each recursion. These transformations are needed for the factor analysis, so we apply them for the other data-rich methods for comparability.

The ridge estimator solves the penalized regression problem

$$\min_{\Gamma} \boldsymbol{RSS} + \boldsymbol{\nu} \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \Gamma_{\boldsymbol{i}}^2 \tag{3}$$

and is given by

$$\hat{\Gamma}_{\boldsymbol{h}}^{\boldsymbol{RR}} = (\mathrm{X'X} + \boldsymbol{\nu} \mathrm{I}_{\boldsymbol{N}})^{-1} \mathrm{X'} \boldsymbol{y}^{\boldsymbol{h}}, \tag{4}$$

where $\mathrm{X} = \begin{bmatrix} \mathrm{X}_1 & \cdots & \mathrm{X}_{\boldsymbol{T}-\boldsymbol{h}} \end{bmatrix}'$, $\boldsymbol{y}^{\boldsymbol{h}} = \begin{bmatrix} \boldsymbol{y}_{\boldsymbol{h}+1,1} & \cdots & \boldsymbol{y}_{\boldsymbol{T},\boldsymbol{T}-\boldsymbol{h}} \end{bmatrix}'$, $\Gamma_{\boldsymbol{i}}$ is an element of $\Gamma$, and $\boldsymbol{\nu}$ is the scalar ridge or shrinkage parameter.[4] The RR estimator penalises large parameter estimates and shrinks the LS estimator towards (but almost never to) zero, thus reducing the variance of the estimators. With RR, all variables enter the forecasting equation. We produce forecasts for $\boldsymbol{\nu}$ = 0.25, $\boldsymbol{N}$, $5\boldsymbol{N}$ and $10\boldsymbol{N}$. $\boldsymbol{\nu}$ = $\boldsymbol{N}$, $5\boldsymbol{N}$ and $10\boldsymbol{N}$ has also been chosen by Groen and Kapetanios (2008). $\boldsymbol{\nu}$ = 0.25 reflect much less shrinkage but is consistent with the parameterisation for the EN and TPS we choose below and chosen by previous work.

*Elastic net*

EN is more general than RR and performs shrinkage and variable selection at the same time, i.e. unlike with RR some parameters may be shrunk to zero (Zou and Hastie 2005). It solves

$$\min_{\Gamma} \boldsymbol{RSS} + \boldsymbol{\nu}_1 \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \left| \Gamma_{\boldsymbol{i}} \right| + \boldsymbol{\nu}_2 \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} \Gamma_{\boldsymbol{i}}^2, \tag{5}$$

and $\hat{\Gamma}_{\boldsymbol{h}}^{\boldsymbol{EN}}$ is a restricted estimate of $\Gamma$.[5] The EN estimator can be found iteratively with least angle regression (LARS) which efficiently solves the optimization problem (5). For details, see Efron et al. (2004). Following Bai and Ng (2008), we set $\boldsymbol{\nu}_2$ to 0.25. As in their application, our results do not change much when we set $\boldsymbol{\nu}_2$ to 0.5 or even 1.[6] Choosing $\boldsymbol{\nu}_1$ is

[4] The ridge regression can be interpreted as a Bayesian regression with a Gaussian prior, see De Mol *et al.* (2008).

[5] EN combines shrinkage - which is achieved by RR - and variable selection - which is achieved by LASSO, a method we do not use here. In contrast to EN, the optimization problem in LASSO considers only the first penalization term of equation (5). While LASSO "tends to select only one variable from the group and does not care which one is selected" (Zou and Hastie 2005), the EN has the advantage that, under certain parameter constellations, it can select groups of (similar) variables. This is the case if $\boldsymbol{\nu}_2 / (\boldsymbol{\nu}_1 + \boldsymbol{\nu}_2) > 0$, i.e. if the EN penalty is a strict convex combination of the LASSO and the RR penalty terms. We have made sure that this is indeed the case for our choices of $\boldsymbol{\nu}_2$ and of the number of variables we keep in our datasets. EN should also select similar variables at each point in time, unlike LASSO. This should make results easier to interpret than with LASSO, and this is also the main reason why we use the EN here and not LASSO.

[6] A small value for $\boldsymbol{\nu}_2$ reflects the idea that with EN, the focus lies on variable selection (and LASSO shrinkage). We also set $\boldsymbol{\nu}_2$ to $\boldsymbol{N}$ or values larger than $\boldsymbol{N}$, as for the RR case. In this case, the ridge penalty term in equation (5) obtains a very large weight relative to the first (LASSO) penalty term. No noticeable forecast

equivalent to choosing the number of variables with non-zero coefficients which we set to 30, 100 and 200 for our forecasts.

### 4.3.2. Factor methods

*Principal components*

Approximate factor models (Chamberlain and Rothschild 1983, Stock and Watson 2002a, Forni et al. 2000, Bai and Ng 2002, Bai 2003) assume that the (typically large number of) variables in the $\boldsymbol{N} \times 1$ vector $\mathrm{X}_t = [\boldsymbol{x}_{1t} \quad \cdots \quad \boldsymbol{x}_{Nt}]'$ can be described as

$$\mathrm{X}_t = \Lambda \mathrm{F}_t + \Xi_t \,. \tag{6}$$

$\Lambda = [\boldsymbol{\lambda}_1' \quad \cdots \quad \boldsymbol{\lambda}_N']'$ is the $\boldsymbol{N} \times \boldsymbol{r}$ matrix of factor loadings and $\Xi_t$ is the $\boldsymbol{N} \times 1$ vector of idiosyncratic components which can be weakly mutually and serially correlated in the sense of Bai and Ng (2002). $\mathrm{F}_t$ is the $\boldsymbol{r} \times 1$ vector of common factors, and $\boldsymbol{r}$, the number of common factors, is typically small.

Estimation of $\mathrm{F}_t$ and $\Lambda$ with PC involves solving the following optimization problem

$$\min_{\Lambda,\mathrm{F}} \frac{1}{\boldsymbol{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \boldsymbol{\lambda}_i \,' \mathrm{F}_t)^2 \tag{7}$$

subject to $\sum_{t=1}^{T} \mathrm{F}_t \mathrm{F}_t ' / \boldsymbol{T} = \mathrm{I}_r$ and $\Lambda ' \Lambda$ being diagonal. This is equivalent to searching for the linear combinations of the data that maximize the share of the total variance explained for each number of combinations up to $\boldsymbol{N}$. A solution to this problem can be found by an eigenvalue-eigenvector decomposition of $\sum_{t=1}^{T} \mathrm{X}_t ' \mathrm{X}_t / \boldsymbol{T}$. PC estimates for $\Lambda$ and $\mathrm{F}_t$ are given by

$$\hat{\Lambda} = \hat{\mathrm{V}} \,, \tag{8}$$

and

$$\hat{\mathrm{F}}_t^{PC} = \hat{\mathrm{V}}' \mathrm{X}_t \,, \tag{9}$$

where $\hat{\mathrm{V}}$ is the $\boldsymbol{N} \times \boldsymbol{r}$ matrix of eigenvectors corresponding to the $\boldsymbol{r}$ largest eigenvalues of $\sum_{t=1}^{T} \mathrm{X}_t ' \mathrm{X}_t / \boldsymbol{T}$. We introduce $\hat{\mathrm{F}}_t^{PC}$ in equation (1) and estimate the LS regression

$$\boldsymbol{y}_{t+h,t} = \mathrm{b}' \hat{\mathrm{F}}_t^{PC} + \boldsymbol{e}_{t+h,t} \,. \tag{10}$$

improvements can be obtained compared to the RR case with larger shrinkage parameters, and, hence, we do not present results for EN with large values of $\boldsymbol{\nu}_2$ here.

Our factor-model-based forecasts use a fixed $\boldsymbol{r}$ = 1, 2, 3 and 5. $\boldsymbol{r} = 5$ is suggested by the Bai and Ng (2002) $\boldsymbol{IC}_{p1}$ for all three (national, international and combined national and international) datasets when applied to the whole sample period. In addition, we consider smaller numbers of (up to 3) factors, because previous studies have found that forecasts based on very few factors, often only one single factor, tend to yield better results than less sparsely parameterised factor models.

*Weighted principal components*

Factor forecasting based on PC is not always successful in practice. For instance, the PC estimator does not perform well when idiosyncratic components are highly cross-correlated and explain a larger part of the variation of the data. We adopt Rule 1c suggested by Boivin and Ng (2006) (WPCBN) which involves removing variables from the set which, if included in the PC estimate, are characterised by highly cross-correlated idiosyncratic components and a small commonality, i.e. variance explained by the common factors relative to the total variance. With the rule applied here, weights are either zero or one, i.e. variables are either in or out. We estimate in a first step common and idiosyncratic components, again, based on $\boldsymbol{r}$ = 1, 2, 3 and 5. In a second step, we consider the set of series whose error is most correlated with the error of some other series. Those series are dropped. If two series' idiosyncratic components are most correlated with each other, the series with the lower commonality is dropped in a third step. Last, PCs are estimated from the reduced dataset and included in the forecasting equation.

As we will show below, it will turn out that WPCBN reduces the amount of cross-correlation in the idiosyncratic components in our datasets but does not succeed in increasing the commonality. We therefore consider an additional rule (rule SWa as described in Boivin and Ng 2006, WPCSW) which weights down variables with important idiosyncratic components. This involves first applying PC to $X_t$ (based on $\boldsymbol{r}$ = 1, 2, 3 and 5), then multiplying each series with the inverse of its idiosyncratic component's standard deviation, and applying PC to the matrix of weighted series. As before, the factor estimates are introduced in the forecasting equation.

*Targeted predictors*

Another problem of simple PC for forecasting is that it finds the linear combinations of variables that maximize the variance of the total dataset explained, but which are not directly related to the target variable, and hence do not necessarily improve forecasts of the latter variable. TP, by contrast, takes into account the relation between the target variable and the predictors. The idea is to estimate PCs from a reduced set of variables that contains only variables closely related to the target variable. These PCs are then introduced in the forecasting equation. Insofar, TP combines variable selection and PC. Following Bai and Ng (2008), we select the variables in two ways.

Hard thresholding (TPH) uses a statistical criterion to decide which variables to exclude from the active set. The target variable is regressed on each predictor one-by-one, and only predictors with an absolute *t*-statistic exceeding a specified threshold are retained in the set.[7] We set a threshold of 1.65, the critical value of the two-sided 5% significance level. One weakness of TPH is that when making a decision on whether a variable is included in the active set of not, it does not take into account the other predictors. Consequently, TPH may end up selecting "too similar" variables, and we know that pooling over variables is effective only if these have distinctive information content from each other.

With soft-thresholding (TPS), by contrast, other predictors are taken into account. We select 30, 100 and 200 variables solving the EN problem with LARS. Our choice of a minimum of 30 is based on work by Boivin and Ng (2006) who show in simulations that factors can be estimated precisely from datasets containing 30 variables or more. Bai and Ng (2008) have also based their analysis on 30 variables. The parameter $\nu_2$ is set to 0.25, consistent with our parameterisation for EN.

*Partial least squares*

PLS which has been suggested by Wold (1982) and which has become prominent in sciences such as chemical engineering and chemometrics, basically finds the linear combinations of predictors that have maximum covariance with the target variable while being mutually orthogonal. We adopt the following algorithm[8]:

i. Set $\boldsymbol{u}_t = \boldsymbol{y}_{t+h,t}$ and $\boldsymbol{\vartheta}_{it} = \boldsymbol{x}_{it}$, $\boldsymbol{i} = 1,...,\boldsymbol{N}$. Set $\boldsymbol{j} = 1$.
ii. Determine the $\boldsymbol{N} \times 1$ vector of indicator variable weights or loadings $\boldsymbol{\omega}_j = \left[\boldsymbol{\omega}_{1j} \quad \cdots \quad \boldsymbol{\omega}_{Nj}\right]'$ by computing individual covariances: $\boldsymbol{\omega}_{ij} = \mathrm{Cov}(\boldsymbol{u}_t, \boldsymbol{\vartheta}_{it})$, $\boldsymbol{i} = 1,...,\boldsymbol{N}$. Construct the $\boldsymbol{j}$ th PLS factor by taking the linear combination given by $\boldsymbol{\omega}_j' \boldsymbol{\vartheta}_t$ and denote this factor by $\hat{\boldsymbol{f}}_{jt}^{PLS}$.
iii. Regress $\boldsymbol{u}_t$ and $\boldsymbol{\vartheta}_{it}$, $\boldsymbol{i} = 1,...,\boldsymbol{N}$ on $\hat{\boldsymbol{f}}_{jt}^{PLS}$. Denote the residuals of these regressions by $\tilde{\boldsymbol{u}}_t$ and $\tilde{\boldsymbol{\vartheta}}_{it}$, respectively.
iv. If $\boldsymbol{j} = \boldsymbol{k}$, where $\boldsymbol{k}$ is the number of PLS factors, stop, else set $\boldsymbol{u}_t = \tilde{\boldsymbol{u}}_t$ and $\boldsymbol{\vartheta}_{it} = \tilde{\boldsymbol{\vartheta}}_{it}$, $\boldsymbol{i} = 1,...,\boldsymbol{N}$ and $\boldsymbol{j} = \boldsymbol{j} + 1$ and go to step ii.

---

[7] When we analyse subsets of our data, in a few cases, *t*-statistics associated with only less than *r* variables exceed the critical values, and we cannot estimate *r* principal components. In these cases, we include the *r* variables with the largest *t*-statistics in the active set.
[8] A good overview is given in Groen and Kapetanios (2008), and we follow the authors closely.

$\hat{F}_t^{PLS,h} = \left[ \hat{f}_{1t}^{PLS} \quad \cdots \quad \hat{f}_{kt}^{PLS} \right]'$ is the $k \times 1$ matrix of estimated common PLS factors for horizon $h$ which are included in the forecasting equation (1). We set $k$ to 1, 2, 3 and 5.

Table 2 gives an overview of the specifications we apply.

## 4.4. Strength and weaknesses of the various approaches

It is not clear *a priori* which of the approaches adopted in this paper will lead to the best results. Each method has its strengths and weaknesses, and the success of each method will depend on the particular forecasting problem at hand, and on the particular characteristics of the data.

As noted, the data-rich methods have several advantages over the trade-weighted (and other simple weighted-average) aggregates. First, they implicitly take into account not only direct trade linkages, but additional linkages which may be relevant as well, such as indirect trade, financial market and confidence linkages. Those linkages are more difficult to measure directly: data are either not available, or they are very volatile (in the case of financial data), or it is not obvious how to translate them into weights. For these reasons, explicit calculation of weights which reflect these other linkages are less common. A second obvious advantage is that data-rich methods can cope with more information compared to the trade-weighted averaging approach. For example, not only foreign countries' GDP growth, but other real economic variables such as productivity, capacity utilization, consumption, investment, industrial production etc. (which may be more directly relevant influences on domestic developments than foreign GDP) can be taken into account. Third, the data-rich methods (generally) derive weights from an explicit statistical optimisation problem, whereas trade-weights do not.

However, trade-weighted (and other simple weighted-average) aggregates also have their appeal. They are often preferred by practitioners, since they provide an intuitive summary of foreign data and may thus make communication with the users of forecasts easier (Robertson 2000, Drew and Frith 1998). In contrast to the data-rich methods (as applied here), trade-weighting can accommodate changing weights, whereas PC and variants of PC, PLS, RR and EN approaches are based on constant factor loadings and/or weights.[9] This may be important since globalisation has increased the trade and financial markets integration of New Zealand in the global economy.

[9] Del Negro and Otrok (2008) suggest a factor model with time-varying loadings which is, however, computationally more demanding than the constant parameter approaches.

As noted, a disadvantage of PC and WPC compared to other methods is that PC and WPC do not take into account the relationship between the predictors and the target variable. This may be one reason why PC, despite of its ability to use lots of data, does not systematically outperform small-scale models and why other data-rich methods which take the predictors' relationship with the target variable into account tend to outperform factor models as noted above (Boivin and Ng 2006, Eickmeier and Ziegler 2008).

Whether PC, PLS and RR on the one hand or EN on the other hand perform better will depend on whether some coefficients are actually exactly zero, i.e. whether some predictors are in fact irrelevant. If this is the case, including all variables in the forecasting model (as done with PC, PLS and RR) increases sampling uncertainty, but will not improve forecasts, and forecasters may be better off setting these coefficients to zero (as done with EN).

Finally, if the conditions that guarantee consistent estimation of the factor space in approximate factor models are not satisfied, PC (and TP) will lead to poor forecasts compared to WPC, RR and PLS. The conditions which are relevant here are non-trivial loadings of the factors on a infinite number of time series (so called "strong" factors)[10] and weak serial and cross-correlation of the idiosyncratic components (e.g. Boivin and Ng 2006). In addition, Boivin and Ng (2006) have shown that factor estimates are worsened with the cross-section dispersion of the commonality. Groen and Kapetanios (2008) have, for example, demonstrated in Monte Carlo simulations that PC is outperformed by PLS and RR if factors are "weak". These issues may be particularly relevant in our international context where there may be regional besides global cycles, which would lead to highly correlated idiosyncratic components and to small loadings if some countries' variables are barely correlated with all other countries' variables.[11]

## 4.5. Characteristics of the dataset

Here we summarise the statistical characteristics of our dataset, to illuminate some of the above issues with the various techniques and to guide the choice of the technique. First, we compute *correlations between the target variable* $y_{t+1,t}$ *and each predictor* $x_{it}$ contained in our national and international datasets. We show in Figure 1, panel (a) box plots of absolute correlations for the full national, international and combined (national and international) datasets as well as for datasets which were reduced with either hard- or soft-thresholding. There are more national than international variables being highly correlated with $y_{t+1,t}$. The figure also nicely illustrates the effect of pre-selection based on both hard- and soft-thresholding: the share of variables which are relatively highly correlated with the target

[10] Formally, Onatski (2007) defines "weak factors" as factors that have bounded instead of increasing with $N$, cumulative effects on the cross-sectional units.
[11] See Kose et al. (2008) for evidence of global divergence and regional convergence.

variable rises in most cases. Hard thresholding and soft thresholding with 30 variables are most successful in increasing the correlation between predictors and target variable, whereas, surprisingly, the correlation does not change much with soft thresholding with 100 or 200 variables. The dispersion between individual variables' correlations with $y_{t+1,t}$ is smallest with hard-thresholding which probably reflects that it tends to select rather similar variables.

We next fit a factor model to the national, the international and the combined datasets. In Table 3, we report the *correlations between the factors extracted from* $X_t$ *and the target variable* $y_{t+1,t}$. Three observations are worth highlighting. First, with PC, the first factor is meaningfully correlated with $y_{t+1,t}$ when extracted from the national and the combined datasets with correlation coefficients of 0.5, whereas the correlation coefficient is lower, at 0.3, when the factors are estimated based on international data only. Second, the correlation does not decline monotonically for lower-ranked PCs. For example, the third PC of the international dataset is most highly correlated with $y_{t+1,t}$ (correlation coefficient: 0.32), followed by the first (0.28) and the fourth PCs (0.27). This is because the optimisation problem which yields the PC factors does not take into account the relation between target and predictors. Third, pre-selection unsurprisingly clearly helps to extract factors which are more closely linked to the target variable. For all datasets, the correlations of the first PC with the target variable estimated with TPH exceed those estimated with simple PC, and correlations increase further when variable selection is made with TPS and 30 variables. As noted before, TPS with 100 or 200 variables is less successful in raising the correlation. Correlations based on TPS and 30 variables are comparable with or even larger than correlations between the first PLS factors and the target variable.[12]

We then decompose each variable into a common component and an idiosyncratic component based on $r = 3$ and $r = 5$ and examine, along the lines of Boivin and Ng (2008), to what extent *assumptions of the approximate factor model* are satisfied for our datasets. Since results for $r = 3$ and $r = 5$ are qualitatively very similar, we present in the remainder of this section only results for $r = 5$. We first examine whether *cross-correlation between the idiosyncratic components* could be a problem. We follow Boivin and Ng (2006) and estimate, for each variable, the correlation of its idiosyncratic component with each other variable's idiosyncratic component. Distributions of the maximum absolute correlations are shown in Figure 1, panel (b). We also show below the sums over all absolute correlations which should be bounded. There are relatively more variables with highly cross-correlated errors in the

[12] At first sight, it seems surprising that the correlation between the target variable and PLS factors does not decline monotonically with lower-ranked factors. The reason is twofold. First, we look at correlations here and not at covariances which are considered in the iterations. Second, the algorithm maximises the covariance between the target variable and a linear combination of the predictors for the first PLS factor, but between the "cleaned" target variable (i.e. the residual from a regression of the target variable and the previous PLS factor(s)) and a linear combination of the "cleaned" predictors for the second, third etc. PLS factors.

international and combined datasets than in the national dataset. Moreover, WPCBN clearly helps reducing the amount of cross-correlation in the errors.

We next assess the *commonality*. Average commonality ratios of our international data are at 32 percent (based on 3 factors) and 42 percent (based on 5 factors) and are, thus, comparable with those reported in other international business cycle papers (e.g. Eickmeier 2008, Marcellino *et al.* 2000). Figure 1, panel (c) shows box plots of commonalities of all individual variables. There are relatively more variables with low commonality in the international dataset than in the national dataset. However, the dispersion of the commonality ratios is also larger for the national than the international dataset, and Boivin and Ng (2008) have shown that this also worsens the factor estimates. Interestingly, WPCBN does not help to increase the commonality in our setup. The reason may be the hierarchical way WPCBN proceeds: it considers only variables with highly cross-correlated errors for elimination. If these variables are not characterised by an unimportant common component, WPCBN will not help increasing the average commonality.

Overall it appears that most potential difficulties seem to be relievable with pre-selection except for the low and dispersed commonality of many variables in our dataset which could pose a problem. This, as noted, leads us to consider WPCSW as well.

## 5. Forecasting results

In this section we first summarise our forecasting results for New Zealand GDP growth based on data-rich methods exploiting only national data (also to allow us for broad comparison with Matheson's (2006) results using a very similar national dataset) and then present our forecast results based on international and both national and international data.

### 5.1. Only national data

Table 4 provides RMSEs based on data-rich methods applied to the national dataset and, relative to these, RMSEs obtained of univariate benchmark models for different horizons. Relative RMSEs below 1 indicate that the univariate benchmark model is outperformed.

We find that univariate models deliver quite good forecasts relative to many data-rich methods. Their average absolute forecasts errors lie between 0.2 and 0.6 percentage points.[13]

[13] In the tables, we provide RMSEs averaged over horizons. It is therefore not surprising that forecast errors decline with the horizon.

No data-rich method outperforms the univariate model at horizon 8. The quality of forecasts based on data-rich methods for the other horizons differs widely. Of the factor models, simple PC performs well for short horizons (1 and 2). Preselection only helps in some cases: TPH yields lower RMSEs than simple PC for horizon 4 while TPS with 100 or 200 selected variables deliver relatively small forecast errors at shorter horizons. It is noticeable that in our setup (and unlike in the US setup of Bai and Ng 2008) more than 30 are necessary for TPS to outperform simple PC and even the univariate benchmark model. WPCSW is very useful, and WPCSW with 3 and 5 factors even yields the best model for the one-year horizon. PLS also performs well at shorter horizons, and yields the best model at horizon 2 with a reduction of the RMSE by 21 percent.

Of the shrinkage methods, RR with greater shrinkages parameters ($\boldsymbol{N}$, $5\boldsymbol{N}$ and $10\boldsymbol{N}$) is successful at forecasting New Zealand growth at horizons 1, 2 and 4. By contrast, applying RR and setting the shrinkage parameter to only 0.25 tends to worsen forecasts relative to the benchmark. EN with 30 variables (and $\boldsymbol{\nu}_2 = 0.25$) generally outperforms RR with $\boldsymbol{\nu} = 0.25$, but relative RMSEs still exceed 1 still in most cases. EN with a larger number of variables having non-zero coefficients does worse.

These results are roughly consistent with Matheson (2006). He compares his PC results not with a univariate benchmark model, but instead with RBNZ's official forecasts for GDP and finds that these are outperformed by PC only at longer horizons (4 quarters and more). Moreover, he compares PC with variable selection prior to PC which is similar to TPH. In line with our results, he finds that forecasts of New Zealand GDP can be improved with variable selection compared to PC applied to the entire dataset for a few longer horizons, but that, in the short run, it seems better to exploit all variables.

## 5.2. International (and national) data

Table 4 also shows results with *only international data*, i.e. the trade-weighted averages or the large international dataset, considered. Of the models with trade-weighted aggregates, the models based on RBNZ's current practice (which uses only exports for GDP12, only imports for CPI5, 80:20 for interest rates) perform worst. Models based on time-varying exports plus imports weights (as in the GVAR) perform best and outperform models based on constant exports plus imports weights, although the gains are very small. All three models outperform the univariate benchmark only at horizon 1.

Of the factor methods, only PLS outperforms the benchmark in most cases while relative RMSEs based on the other factor methods generally exceed unity. The shrinkage methods tend to outperform factor methods and the benchmark model. In two out of four cases, PLS and in the other two cases, the RR yield the best models. These methods also tend to outperform the models with trade-weighted averages.[14]

That table also shows forecast results based on both *national and international data*. We obtain similar results as for purely international data with shrinkage methods and PLS performing very well. For small horizons simple PC and TPH work fairly well, and for horizons up to 4 quarters TPS with 200 variables yields low forecast errors.

## 6. What type of international information is most relevant for New Zealand GDP forecasting?

In this section, we partition the international dataset into subsets, for two purposes. First, we assess the marginal predictive content of each subset in terms of forecasting accuracy, and second we look at the contributions of each subset to the forecasts over time when all data are included to produce the forecast. This allows us to interpret the forecasts in terms of estimated foreign influences by region and by economic type of data and transmission mechanisms for our sample. The first partitioning scheme is geographical – data are split by North America, Europe, Asia and Australia. The second partitioning scheme is economic – data are split into subsets comprising New Zealand's trading partners' real, price and monetary/financial data, and world commodity prices.

In the forecasting accuracy exercise we replace the international datasets with each of these subsets in turn. Since the subsets are much smaller than the full datasets, we consider for EN and TPS only 30, not 100 or 200 variables. In subsection 6.1. we report the best five models for each horizon. In subsection 6.2., we summarise the results – national versus international data, trade-weighted aggregates versus data-rich methods, forecasts based on subsets of international data. In Table A.2. of the appendix we report all results associated with subsets of the international dataset in detail.

[14] We have also looked at which were among the 10 most frequently selected variables by EN with 30 variables applied to our international dataset. Results are quite stable over recursions with the 10 most frequently selected variables appearing between 24 and 32 out of 32 forecasts. Euro-area and Japanese (real) variables are relatively frequently selected. World commodity prices are among the most frequently selected variables for horizons 1, 4 and 8 and seem to be good leading indicators for New Zealand growth. We also find that 3 and 2 out of 10 most frequently selected variables are typical foreign leading indicators for horizons 4 (Japanese business confidence, British consumer confidence, and the Australian leading indicator) and 8 (again Japanese business confidence and the euro-area term spread), respectively. Some outcome is also unexpected and hard to interpret: for example Malaysian M1 and the euro-area labour force are among the most frequently selected variables. This finding is unsatisfying but consistent with De Mol *et al*. (2008) who apply LASSO and possibly due to the fact that we give the (stabilising) ridge penalisation term a relatively small weight.

In subsection 6.3., we show historical contributions of the international data by subset, and discuss whether the apparent international influences and transmission mechanisms are plausible. In subsection 6.4., we ask to what extent national survey data reflect international information and whether this may affect our results.

## 6.1. Best forecasting models

Table 5 reports the best five forecasting models for each horizon out of all models we have considered. Consistent with previous results, PLS performs very well: 7 out of the 20 best models were estimated with PLS. TPS is also quite successful (4 out of the 20 best models were estimated with TPS), followed by RR, WPCSW and WPCBN.

The set of best five models never includes a model using only national data, unsurprisingly. Indeed, 9 out of the 20 best models exploit only international, but no national data, and 14 out of the 20 best models exploit only subsets of international data (possibly along with national data). The finding that forecast models using only regional subsets of the international data often dominate in forecasting performance, at least for factor models, is a little surprising at first sight. One might have supposed that more data would be better, or at least no worse, in terms of forecasting accuracy.

## 6.2. Summary of forecast results

Figure 2 compares *models exploiting international information with models exploiting only national information*. The comparison is made for the best models of each category, the average over the best five models and the average over all models for each horizon. The two upper bar graphs show that it can be worth exploiting international information. The gains are particularly large for long horizons, suggesting that spillovers take some time to materialize, possibly also due to third market effects. In some cases, it is better to exploit international information only, in some cases international information along with national information. The lower graph shows that on average over all models, results are less clear-cut, reflecting, again, that forecast errors differ widely over models. In the following, we mainly concentrate on the best and the average over the best five models.

It is also clear from Figure 3 that *forecasts based on data-rich methods* (where, for comparability, we consider models which use only international data) outperform *forecasts based on trade-weighted aggregates*.

Based on Figure 4, we can assess which type of international data yields the greatest improvements over only national forecast models. Gains can be obtained by including North

American, European, Asian, and Australian data (panel (a)). Gains from including Australian data are particularly large at the one-quarter horizon reflecting strong direct linkages between New Zealand and Australia. The quite big improvements which can be achieved by including European data (relative to data from other regions) seems surprising. Europe is a less important trading partner by pure value of trade with New Zealand, and the conventional wisdom is that fast-moving international financial and confidence shocks typically emanate from the US rather than Europe. One interpretation would be that national survey data already contain international, and here North American, information given the dominance of the US for financial markets and its role as a locomotive for the world economy. We will test this explanation in subsection 6.4. Contributions from world commodity prices are generally smaller and positive only for horizons 4 and 8.

Turning to the split by economic type of data (panel (b)), gains can be achieved by including international real, price and monetary and financial data (except for horizon 2 where only real data seem to help reducing the forecast error). Gains are surprisingly large from including international price data (at horizons 1, 4 and 8) relative to real and monetary and financial data. Again, given the focus of financial markets in the press, information on financial developments might be captured by proxy (or indeed with a better signal/noise ratio) in national survey data.

## 6.3. Historical contributions to New Zealand GDP growth forecasts

Figure 5 shows the 1-quarter ahead forecasts of New Zealand GDP growth (as deviations from the mean) and the contributions of national and international variables.[15] These contributions allow us to explain the forecasts in terms of the economic and regional drivers over the forecast period. This is particularly important for policy making institutions: the forecasts need to be communicated clearly and credibly, both to decision makers and to the public. We show results for RR(*N*) which has been shown to perform very well.[16] By construction, RR yields biased estimates of the parameters, and hence, we can, of course, decompose our forecast in the contributing variable groups, but should not overdo interpretation in terms of international spillovers. To check robustness with respect to the data-rich methods, we have also computed historical forecast contributions for PC (which also yields biased parameters estimates). Results are similar to those based on RR, and we do not present them here, but make them available upon request.

[15] We are grateful to Troy Matheson for suggesting this exercise to us.
[16] The optimal model contains 1 lag of the target variable. We take into account that lags of the target variable are explained by past predictors of the large dataset. Differences between the forecasts (black line in figures 5 and 7) and the sums of the contributions (bars in figures 5 and 7) are the sums of GDP growth in the first period and the contribution of the forecast equation's residuals.

Looking first at the results for the contribution of national versus international variables, the results seem quite plausible. As documented in RBNZ (2007), New Zealand came out of the developed world recession of 2000-2001 quite early relative to other countries, so the relative contribution of national variables (dark blue bars) is quite substantial over the early period from 2001 to 2004. In particular, New Zealand experienced a large influx of migrants and returning New Zealanders over this period, partly reflecting terrorism and personal security fears leading expatriate New Zealanders to return home. At the same time, coincidentally, there was a large influx of foreign students of English language. These two factors combined to add strength to the local housing market upturn, which began one or two years earlier than in other countries. Later in the sample, around 2005-2006, the New Zealand economy began slowing and this was most evident in the housing market cooling down somewhat. This was probably the dominant factor explaining (negative) deviations of growth from mean, along with very dry weather conditions that adversely affected farm output (weather is captured by the Southern Oscillation Index in our model). Our results are in line with Bloor and Matheson (2008) who have analyzed the effect of migration and climate shocks on New Zealand GDP, finding some evidence that climate and migration matter.

When we partition the international variables by region (panel (a)) and look at the contributions, we see that the international contribution to growth forecasts early in the sample (2000-2003) is dominated by Asia, which had the effect of holding up New Zealand growth and buffering it to some degree against the downdrafts coming from North America at this time – possibly reflecting the growing strength in China's economy and the relatively large trade linkages of New Zealand with Asia. From mid-2003 to 2005 as the US housing market boom acquired momentum, its positive influence comes through in the contribution of the North American data. The influence of Australia is surprisingly small given its large trade weight. Interestingly also (and consistent with the previous subsection), through the whole period, the influence of Europe is not trivial. World commodity price fluctuations are, on average, slightly less important than contributions from Australia, North American or Europe. Interestingly, contributions are positive in 2005/2006, possibly due to other than oil price commodity increases and the cyclical expansion of the world economy, and negative thereafter which could be due to oil price increase.

The partitioning of the data by economic type (panel (b)) also produces interesting results. The pattern of real contributions is not surprising, being negative early in the sample when the world was in recession, then turning positive as the world recovered, then negative again as world growth slowed. The prices pattern resembles the cycle in oil prices on international markets – with the extraordinary run-up in oil prices, and consequent pressure on inflation around the world, from 2004 having a substantial downward impact on GDP growth – as well as the world cycle. The financial variables' contributions are strongly positive through the entire sample – a striking result, lending credence to the view that the worldwide "savings glut" and/or "liquidity glut" that emerged after the recovery from the 2000-2001 world recession played a large role in sustaining world growth until the subprime crisis in 2007 (too late in the sample to appear).

## 6.4. Do national surveys reflect international information?

The use of business survey data in the national datasets bears more discussion. As noted, business survey data may capture international information and therefore bias the estimated contribution of national data upwards at the expense of international information. It is surprising, for example, that European and price data make large contributions whereas North American and Australian data make relatively small contributions. One interpretation may be that survey respondents attribute a relatively large weight to US or Australian rather than European data, given the dominant role of the US in financial markets and in the press and its role as a "locomotive" for other countries' business cycles, and the close proximity to Australia, and to real and monetary and financial rather than to price data.

To assess to what extent national survey data contain international information, we extract $\boldsymbol{r} = 5$ PCs from datasets which contain only national survey data, only other national data, all international data and subsets of international data as defined above. (Results are again very similar for 3 factors, and we do not report them here.) We compute the trace $R^2$ which is a multivariate correlation measure and which is defined as the sum of the variances of the projections of the survey factors on the national and the international factors, respectively, divided by the sum of the variances of the latter factors (Stock and Watson 1998). A value of one indicates that the two sets of factors are identical. We find that national survey data indeed seem to reflect international rather than national information. The correlation between survey and international factors (the trace $R^2$ statistic is at 0.47) exceed those between survey and national factors (0.43). National survey data is most highly correlated with North American factors (0.48) followed by Asian (0.47), European (0.42) and Australian factors (0.36). Moreover, they seems to capture international monetary and financial information (0.49) rather than foreign real economic activity (0.41) and prices (0.40). This supports our conjecture and may explain the somewhat counterintuitive small influence of North American and monetary and financial indicators we found in the previous subsections. We finally also find a low correlation of the national survey factors with world commodity price factors (0.08).

We also test more indirectly to what extent survey data mask international influences on New Zealand forecasts by removing survey data from the national dataset and re-estimating the marginal forecast contributions and the historical forecast contributions of subgroups of international data based on this reduced dataset.

North American influences come through stronger than before (when national surveys were included in the national dataset) at horizon 8 as can be seen by figure 6. From that figure, it is also apparent that for longer horizons (4 and 8) on average over all models international data, and in particular not only North American but also Asian data, improve the forecasts over models using only national data by more than before. The marginal contribution of world commodity prices are generally not much altered which is consistent with the low correlation between world commodity price factors and national survey factors we found before.

Figure 7 which shows historical contributions after exclusion of national survey data also suggests that surveys indeed seem to have captured international information. The recession in 2001 is now entirely attributed to real (and North American) data. The composition of the effects of international information seems to have shifted as well. A greater role is now attributed to real data and a smaller role to price relative to other international data. This, again, supports our conjecture that national survey data capture international information.

## 7. Conclusion and outlook

To recap, this paper's main contributions were to systematically assess the value of exploiting international information in producing national forecasts, specifically of New Zealand GDP growth. We looked at various data-rich approaches. Most of them have only been applied to US data, and not very often. We have found it useful to apply them in a wider range of applications and suggest that further work along these lines would be useful to improve understanding on whether these lesser-known techniques are useful in practice. Finally, we showed how the data-rich techniques could fill a gap between structural econometric work and forecasting settings by investigating the marginal predictive content of subsets of international data, and by looking at historical contributions to GDP growth forecasts. This is important for policy makers, among others, who usually want to explain the forecasts in a consistent and accessible way as a basis for policy decisions.

Our main findings were as follows. International information drawn from large international datasets can substantially improve forecasts of New Zealand GDP growth and outperform forecasts based on trade weights. The gains are particularly large at long forecast horizons. However, the performance of data-rich methods differs widely with shrinkage methods, i.e. RR and EN, and partial least squares performing particularly well in our international context. Historical decompositions based on the data-rich techniques seem to be consistent with the economic experience over the sample as documented by policymakers. National business survey data seem to capture a substantial proportion of the relevant information in developments offshore.

This work could be extended in two dimensions. Using inflation as a forecast target variable would test how well the data-rich approaches generalise to other forecasting settings. Moreover, a structural analysis would capture the relevant international shocks and assess their transmission in a more formal (identified) setting.

## References


Bai, J. (2003), "Inferential theory for factor models of large dimensions", *Econometrica*, 71, 135-171.

Bai J. (2004), "Estimating cross-section common stochastic trends in nonstationary panel data", *Journal of Econometrics*, 122, 137-183.

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

Bai J, S. Ng (2004), "A PANIC attack on unit roots and cointegration", *Econometrica*, 72(4), 1127-1177.

Bai, J., S. Ng (2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics*, 146 (2008), 304-317.

Banerjee, A., M. Marcellino, I. Masten (2006), "Forecasting macroeconomic variables for the new member states" in: Artis, M. J., A. Banerjee, M. Marcellino (eds.), The central and eastern European countries and the European Union, Cambridge University Press, Cambridge, Chapter 4, 108-134.

Bernanke B, J. Boivin, P. Eliasz (2005), "Measuring the effects of monetary policy: a Factor-Augmented Vector Autoregressive (FAVAR) approach", *The Quarterly Journal of Economics*, 120(1), 387-422.

Bloor, C., T. Matheson (2008), "Analyzing shock transmission in a data-rich environment: a large BVAR for New Zealand", RBNZ Discussion Paper 2008/09.

Boivin J, S. Ng (2006), "Are more data always better for factor analysis", *Journal of Econometrics*, 132, 169-194.

Brisson, M., B. Campbell, J. W. Galbraith (2003), "Forecasting some low-predictability time series using diffusion indices", *Journal of Forecasting*, 22, 515-531.

Buckle, R. A., K. Kim, H. Kirkham, N. McLellan, J. Sharma (2007), "A structural VAR business cycle model for a volatile small open economy", *Economic Modelling*, 24, 990-1017.

Chamberlain G, M. Rothschild (1983), "Arbitrage factor structure and mean-variance analysis in large asset markets", *Econometrica* 51, 1305-1324.

Chen, Y. K. Rogoff (2003), "Commodity currencies", *Journal of International Economics*, 60(1), 133-160.

Cheung, C., F. Demers (2007), "Evaluating Forecasts from Factor Models for Canadian GDP Growth and Core Inflation", Bank of Canada Working Paper 8.

Dées, S., F. di Mauro, H. Pesaran, V. Smith (2007), "Exploring the international linkages of the euro area: a Global VAR analysis", *Journal of Applied Econometrics*, 22(1), 1-38.

Del Negro M, C. Otrok (2008), "Dynamic factor models with time-varying parameters: measuring changes in international business cycles", Federal Reserve Bank of New York Staff Report 326 (2005).

De Mol, C., D., Giannone, L. Reichlin (2008), "Forecasting using a large number of predictors: is Bayesian regression a valid alternative to principal components?", *Journal of Econometrics*, forthcoming.

Drew, A., M. Frith (1998), "Forecasting at the Reserve Bank of New Zealand", RBNZ Bulletin, 61(4), 317-327.

Dungey, M., R. Fry (2008), "The identification of fiscal and monetary policy in a structural VAR", CAMA Working Paper 29/2007.

Efron, B., T., Hastie, I. Johnstone, R. Tibshirani (2004), "Least angle regression", *Annals of Statistics*, 32, 407-499.

Eickmeier, S. (2008), "Comovements and heterogeneity analyzed in a non-stationary dynamic factor model", *Journal of Applied Econometrics*, forthcoming.

Eickmeier, S., C. Ziegler (2008), "How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach", *Journal of Forecasting*, 27(3), 237-265.

Forni M., L. Reichlin (1998), "Let's get real: a factor analytic approach to disaggregated business cycle dynamics", *Review of Economic Studies*, 65, 453-474.

Forni, M., M. Hallin, M. Lippi, L. Reichlin (2000), "The generalized dynamic factor model: identification and estimation", *The Review of Economic and Statistics*, 82, 540-554.

Forni, M., D. Giannone, M.Lippi, L. Reichlin (2008), "Opening the black box: structural factor models versus structural VARs", *Econometric Theory*, forthcoming.

Groen, J. J., G. Kapetanios (2008), "Revisiting useful approaches to data-rich macroeconomic forecasting", Federal Reserve Bank of New York Staff Report 327, May.

Gosselin, M.-A., G. Tkacz (2001), "Evaluating factor models: an application to forecasting inflation in Canada", Bank of Canada Working Paper 18.

Imbs, J. (2004), "Trade, finance, specialization and synchronization", Review of Economics and Statistics, 86(3); 723-734.

Kose, A., C. Otrok, E. S. Prasad (2008), "Global business cycles: convergence and decoupling?", IMF WP/08/143.

Kose, M. A., E. S. Prasad, M. E. Terrones (2003), "How does globalization affect the synchronization of business cycles?" American Economic Association Papers and Proceedings, 93(2) (May), 57-62.

Lin, J., R. S. Tsay (2006), "Comparisons of forecasting methods with many predictors", mimeo.

Marcellino M., J. H. Stock, M. W. Watson (2000), "A dynamic factor analysis of the EMU", mimeo. http://www.igier.uni-bocconi.it/whos.php?vedi=1139&tbn=albero &id_doc=177.

Matheson, T., (2006), "Factor model forecasts for New Zealand", International Journal fo Central Banking, June.

Onatski, A. (2007), "Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise", mimeo, Columbia University.

Pesaran M. H., T. Schuermann, S. M. Weiner (2004), „Modelling regional interdependencies using a global error-correcting macroeconometric model", *Journal of Business and Economic Statistics*, 22, 129-162.

RBNZ (2007), "A review of economic developments and monetary policy since 2000", paper A2 in *Submission to Finance and Expenditure Select Committee's inquiry into the future monetary policy framework*. Mimeo at www.rbnz.govt.nz

Robertson, J. C. (2000), "Central bank forecasting: an international comparison", *Federal Reserve Bank of Atlanta Economic Review*, Second Quarter 2000, 21-32.

Schumacher, C. (2007), “Forecasting German GDP using alternative factor models based on large datasets”, *Journal of Forecasting*, 26(4), 271-302.

Schumacher, C. (2009), “Factor forecasting using international targeted predictors: the case of German GDP”, mimeo.

Stock, J., M. Watson (1998), “Diffusion indexes”, NBER Working Paper 6702.

Stock, J., M. Watson (2002a), “Macroeconomic forecasting using diffusion indexes”, *Journal of Business and Economic Statistics*, 20,147-162.

Stock, J., M. Watson (2002b), “Forecasting using principal components from a large number of predictors”, *Journal of the American Statistical Association*, 97, 1167-1179.

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

Stock, J., M., Watson (2005), “Implications of dynamic factor models for VAR analysis”, NBER Working Paper 11467.

Wold (1982), „Soft modeling. The basic design and some extensions“, in: K.-G. Jöreskog and H. Wold (eds.), *Systems under indirect observation*, 2, North-Holland, Amsterdam.

Zou, H., T. Hastie (2005), „Regularization and variable selection via the elastic net”, *Journal of Royal Statistical Society*, Series B, 67(2), 301-320.

## Table 1: Overview of international data

| | North Amer. | Europe | Asia | Australia | World | Sum |
|---|---|---|---|---|---|---|
| Real | 50 | 41 | 95 | 20 | - | 206 |
| Prices | 28 | 28 | 41 | 11 | - | 108 |
| Monetary + financial | 23 | 34 | 62 | 11 | - | 130 |
| Commodity prices | - | - | - | - | 38 | 38 |
| Sum | 101 | 103 | 198 | 42 | 38 | 482 |

Notes: North America comprises the US and Canada, Europe the UK and the euro area, Asia Japan, Korea, China, Hong Kong, Malaysia, Singapore, Taiwan.

## Table 2: Overview of specifications

| | | |
|---|---|---|
| *trade1* | Trade-weighted averages | RBNZ practice |
| *trade2* | | constant weights based on exports + imports |
| *trade3* | | 2-year moving average weights based on exports + imports |
| *PC* | Principal components | $r$ = 1, 2, 3, 5 |
| *WPCBN* | Weighted PC | Rule 1c in Bai and Ng (2008), $r$ = 1, 2, 3, 5 |
| *WPCSW* | | Rule SWa in Bai and Ng (2008), $r$ = 1, 2, 3, 5 |
| *TPH* | Targeted predictors | Hard thresholding, r = 1, 2, 3, 5 |
| *TPS* | | Soft thresholding, 30, 100, 200 variables, $v_2$ = 0.25, $r$ = 1, 2, 3, 5 |
| *PLS* | Partial least squares | $k$ = 1, 2, 3, 5 |
| *RR* | Ridge regression | $v$ = 0.25, N, 5N, 10N |
| *EN* | Elastic net | 30, 100, 200 variables, $v_2$ = 0.25 |

## Table 3: Correlations between factors and the target variable

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|---|---|---|---|---|---|
| **National data** | | | | | |
| PC | 0.53 | 0.04 | 0.15 | 0.02 | 0.06 |
| TPH | 0.54 | 0.05 | 0.14 | 0.16 | 0.22 |
| TPS (30 variables) | 0.73 | 0.31 | 0.14 | 0.19 | 0.22 |
| TPS (100 variables) | 0.54 | 0.03 | 0.52 | 0.01 | 0.06 |
| TPS (200 variables) | 0.53 | 0.04 | 0.31 | 0.02 | 0.28 |
| PLS | 0.59 | 0.60 | 0.25 | 0.29 | 0.28 |
| **International data** | | | | | |
| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
| PC | 0.28 | 0.03 | 0.32 | 0.27 | 0.06 |
| TPH | 0.45 | 0.36 | 0.09 | 0.00 | 0.26 |
| TPS (30 variables) | 0.59 | 0.10 | 0.60 | 0.21 | 0.00 |
| TPS (100 variables) | 0.21 | 0.30 | 0.34 | 0.15 | 0.54 |
| TPS (200 variables) | 0.31 | 0.10 | 0.40 | 0.20 | 0.23 |
| PLS | 0.62 | 0.52 | 0.47 | 0.24 | 0.19 |
| **National and international data** | | | | | |
| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
| PC | 0.46 | 0.28 | 0.08 | 0.19 | 0.09 |
| TPH | 0.56 | 0.12 | 0.13 | 0.20 | 0.15 |
| TPS (30 variables) | 0.86 | 0.09 | 0.09 | 0.27 | 0.12 |
| TPS (100 variables) | 0.55 | 0.58 | 0.15 | 0.16 | 0.11 |
| TPS (200 variables) | 0.51 | 0.51 | 0.08 | 0.23 | 0.12 |
| PLS | 0.64 | 0.53 | 0.40 | 0.30 | 0.17 |

Notes: Correlation between $y_{t+1,t}$ and factors estimated (in sample) from $X_t$ or a reduced version of $X_t$. The factors were normalized to be positively correlated with the target variable.

## Table 4: Relative RMSEs

| | National | | | | International | | | | National + international | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.0061 | 0.0045 | 0.0034 | 0.0024 | 0.0061 | 0.0045 | 0.0034 | 0.0024 | 0.0061 | 0.0045 | 0.0034 | 0.0024 |
| trade1 | - | - | - | - | 0.984 | 1.247 | 1.657 | 2.500 | - | - | - | - |
| trade2 | - | - | - | - | 0.951 | 1.034 | 0.970 | 1.247 | - | - | - | - |
| trade3 | - | - | - | - | 0.934 | 1.022 | 0.933 | 1.242 | - | - | - | - |
| PC (r = 1) | 0.951 | 0.865 | 0.985 | 1.113 | 1.033 | 1.090 | 1.060 | 0.948 | 0.984 | 1.247 | 1.657 | 2.500 |
| PC (r = 2) | 0.951 | 0.910 | 1.119 | 1.387 | 1.016 | 1.056 | 1.149 | 1.247 | 0.951 | 1.034 | 0.970 | 1.247 |
| PC (r = 3) | **0.902** | 0.899 | 1.127 | 1.381 | 1.082 | 1.034 | 1.134 | 1.361 | 0.934 | 1.022 | 0.933 | 1.242 |
| PC (r = 5) | 0.951 | 0.876 | 1.351 | 1.361 | 1.098 | 1.258 | 1.493 | 1.289 | 1.016 | 1.067 | 1.187 | 1.242 |
| WPCBN (r = 1) | 0.951 | 0.899 | 0.993 | 1.124 | 1.033 | 1.079 | 1.045 | 1.031 | 1.082 | 1.180 | 1.194 | 1.268 |
| WPCBN (r = 2) | 0.951 | 0.854 | 1.082 | 1.392 | 1.016 | 1.079 | 1.149 | 1.227 | 1.082 | 1.180 | 1.351 | 1.284 |
| WPCBN (r = 3) | 0.984 | 0.888 | 1.187 | 1.232 | 1.000 | 1.011 | 1.425 | 1.412 | 1.131 | 1.202 | 1.582 | 1.799 |
| WPCBN (r = 5) | 0.967 | 0.955 | 1.410 | 1.485 | 1.033 | 1.348 | 1.373 | 1.407 | 1.033 | 1.090 | 1.216 | 1.247 |
| WPCSW (r = 1) | 1.033 | 1.056 | 1.037 | 1.026 | 1.016 | 1.022 | 1.000 | **0.851** | 1.131 | 1.225 | 1.291 | 1.232 |
| WPCSW (r = 2) | 0.951 | 0.854 | 0.948 | 1.170 | 0.967 | 0.910 | 0.851 | 1.093 | 1.131 | 1.225 | 1.381 | 1.201 |
| WPCSW (r = 3) | **0.902** | 0.843 | **0.896** | 1.093 | 1.049 | 0.955 | 0.918 | 1.139 | 1.098 | 1.169 | 1.306 | 1.778 |
| WPCSW (r = 5) | 0.984 | 0.921 | **0.896** | 1.026 | 1.066 | 1.225 | 1.224 | 1.139 | 0.934 | 0.933 | 1.045 | 1.113 |
| TPH (r = 1) | 0.951 | 0.966 | 0.963 | 1.021 | 1.082 | 1.101 | 1.127 | 1.103 | 0.984 | 1.045 | 1.194 | 1.191 |
| TPH (r = 2) | 0.951 | 1.022 | 0.948 | 1.325 | 1.049 | 1.112 | 1.194 | 1.320 | 0.984 | 1.135 | 1.246 | 1.196 |
| TPH (r = 3) | 1.000 | 1.056 | 0.933 | 1.206 | 1.197 | 1.157 | 1.179 | 1.521 | 0.934 | 0.989 | 1.194 | 1.387 |
| TPH (r = 5) | 1.180 | 1.112 | 1.261 | 1.237 | 1.180 | 1.146 | 1.373 | 1.665 | 0.984 | 1.034 | 1.269 | 1.562 |
| TPS (30 variables, r = 1) | 1.016 | 1.213 | 1.194 | 1.211 | 1.098 | 1.101 | 1.216 | 1.206 | 1.131 | 1.022 | 1.022 | 1.052 |
| TPS (30 variables, r = 2) | 1.033 | 1.303 | 1.149 | 1.289 | 1.246 | 1.213 | 1.373 | 1.201 | 1.033 | 1.022 | 1.015 | 1.046 |
| TPS (30 variables, r = 3) | 1.033 | 1.225 | 1.224 | 1.273 | 1.311 | 1.202 | 1.261 | 1.263 | 1.049 | 1.045 | 1.015 | 1.062 |
| TPS (30 variables, r = 5) | 1.000 | 1.247 | 1.194 | 1.289 | 1.328 | 1.180 | 1.306 | 1.253 | 1.049 | 1.056 | 0.993 | 0.964 |
| TPS (100 variables, r = 1) | 0.934 | 0.831 | 1.157 | 1.139 | 1.082 | 0.955 | 0.963 | 1.356 | 1.049 | 1.191 | 1.037 | 1.082 |
| TPS (100 variables, r = 2) | 1.098 | 0.854 | 1.261 | 1.160 | 1.164 | 1.000 | 1.112 | 1.273 | 1.033 | 1.157 | 0.985 | 1.175 |
| TPS (100 variables, r = 3) | 1.066 | 0.843 | 1.276 | 1.206 | 1.197 | 1.079 | 1.060 | 1.340 | 1.016 | 1.101 | 1.231 | 1.247 |
| TPS (100 variables, r = 5) | 1.115 | 0.899 | 1.164 | 1.196 | 1.082 | 1.169 | 1.134 | 1.448 | 1.049 | 0.989 | 1.254 | 1.237 |
| TPS (200 variables, r = 1) | 0.934 | **0.820** | 1.067 | 1.186 | 1.066 | 1.056 | 1.022 | 1.129 | 0.934 | 0.820 | 0.955 | 1.144 |
| TPS (200 variables, r = 2) | 0.984 | 0.899 | 1.119 | 1.505 | 1.066 | 1.112 | 1.082 | 1.258 | 0.934 | 0.865 | 0.978 | 1.165 |
| TPS (200 variables, r = 3) | 0.951 | 0.944 | 1.209 | 1.649 | 1.000 | 0.933 | 0.925 | 1.418 | 0.951 | 0.820 | 0.970 | 1.196 |
| TPS (200 variables, r = 5) | 1.000 | 0.933 | 1.284 | 1.691 | 1.131 | 1.180 | 1.179 | 1.521 | 0.934 | 0.809 | 1.045 | 1.129 |
| PLS (k = 1) | 0.951 | 0.843 | 0.940 | 1.108 | 1.000 | 1.067 | 0.784 | 1.098 | 0.951 | 0.865 | 0.925 | 1.057 |
| PLS (k = 2) | 1.066 | 1.056 | 1.336 | 1.351 | 0.869 | 0.944 | 0.761 | 1.196 | 1.000 | 0.854 | 0.806 | 1.021 |
| PLS (k = 3) | 0.934 | **0.787** | 1.007 | 1.412 | 0.869 | 1.045 | 0.836 | 1.124 | **0.852** | 0.831 | 0.731 | 0.948 |
| PLS (k = 5) | 0.951 | 1.034 | 1.209 | 1.361 | **0.787** | 0.955 | 0.761 | 0.923 | **0.852** | 0.944 | 0.784 | 0.804 |
| RR ($v_2$ = 0.25) | 1.049 | 0.966 | 1.239 | 1.371 | 0.803 | 1.011 | 0.903 | 1.216 | 0.885 | 0.865 | 0.873 | 0.969 |
| RR ($v_2$ = N) | 0.918 | 0.820 | 0.955 | 1.186 | 0.836 | **0.899** | **0.716** | 1.057 | **0.852** | 0.775 | **0.701** | 1.000 |
| RR ($v_2$ = 5N) | 0.934 | 0.876 | 0.933 | 1.046 | 0.902 | 0.910 | 0.806 | 1.010 | 0.902 | 0.854 | 0.813 | 0.995 |
| RR ($v_2$ = 10N) | 0.934 | 0.899 | 0.940 | 1.010 | 0.934 | 0.933 | 0.866 | 1.005 | 0.934 | 0.888 | 0.866 | 0.995 |
| EN (30 variables, $v_2$ = 0.25) | 1.016 | 0.910 | 1.239 | 1.077 | 0.951 | 1.000 | 0.993 | 1.021 | 0.934 | **0.764** | 0.873 | **0.716** |
| EN (100 variables, $v_2$ = 0.25) | 1.082 | 1.101 | 1.500 | 1.227 | 0.918 | 0.921 | 0.963 | 1.119 | 0.918 | 0.876 | 1.037 | 0.778 |
| EN (200 variables, $v_2$ = 0.25) | 1.082 | 1.011 | 1.373 | 1.351 | 0.836 | 0.955 | 0.940 | 1.206 | 0.885 | 0.843 | 0.963 | 0.938 |

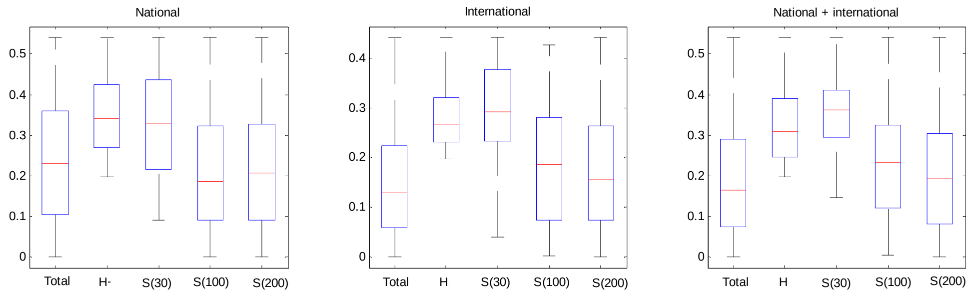
Notes: RMSE(model)/RMSE(univariate). Minima are in bold, and relative RMSEs < 1 are in gray. For more details on the forecasting setup, see the text.

## Table 5: Best five models

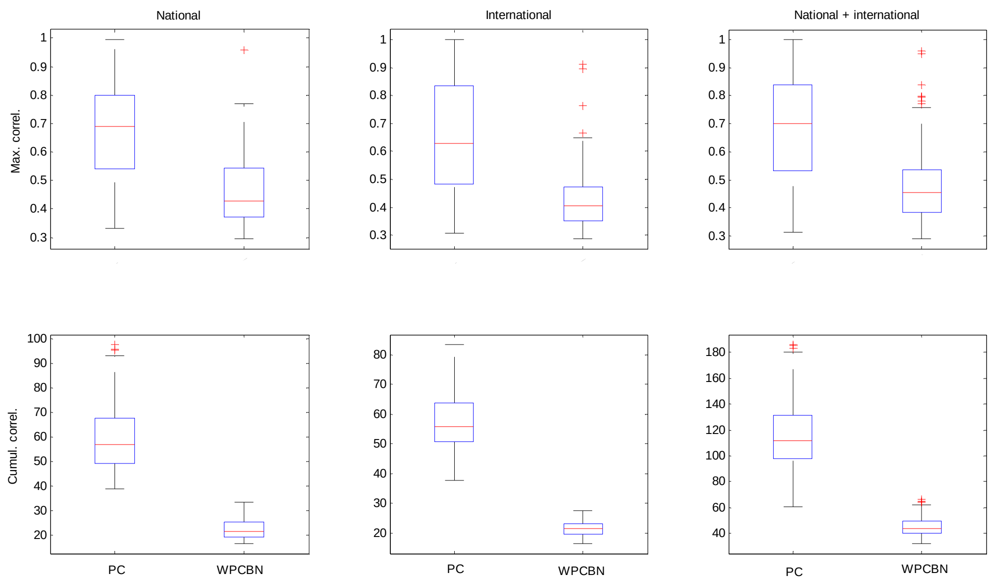
| | h = 1 | h = 2 |
|---|---|---|
| 1 | nat + int, all, WPCSW (*r* = 1) | int only, real data, TPS (30 variables, *r* = 1) |
| 2 | nat + int, Australian data, WPCBN (*r* = 5) | int only, Asian data, EN (30 variables, $v_2$ = 0.25) |
| 3 | nat + int, Australian data, PLS (*k* = 1) | int only, European data, PC (*r* = 3) |
| 4 | nat + int, all, WPCSW (*r* = 2) | int only, all, RR ($v_2 = N$) |
| 5 | nat + int, real data, TPS (30 variables, *r* = 2) | int only, all, PLS (*k* = 3) |
| | **h = 4** | **h = 8** |
| 1 | nat + int, European data, TPS (30 variables, *r* = 5) | int only, real data, PLS (*k* = 3) |
| 2 | nat + int, European, TPH (*r* = 5) | int only, Asian data, WPCBN (*r* = 2) |
| 3 | nat + int, price data, TPS (30 variables, *r* = 5) | nat + int, price data, PLS (*k* = 2) |
| 4 | nat + int, European data, PLS (*k* = 1) | int only, all, RR ($v_2 = N$) |
| 5 | int only, all, PLS (*k* = 3) | nat + int, Asian data, PLS (*k* = 1) |

## Figure 1: Characteristics of the large national and international datasets – box plots

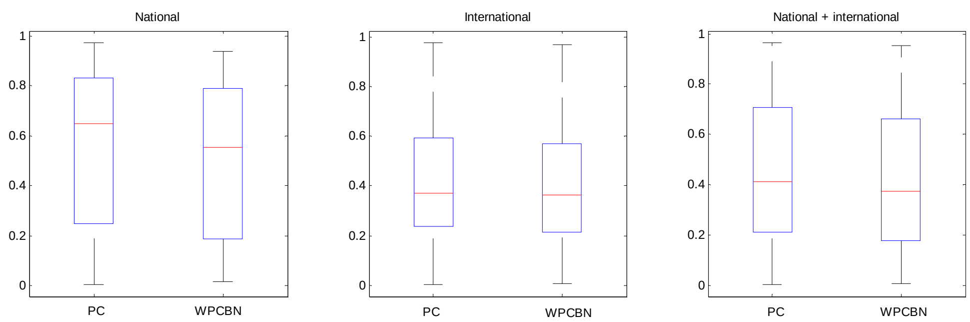
### (a) Absolute correlations between predictors and target variable




### (b) Absolute cross-correlation of idiosyncratic components

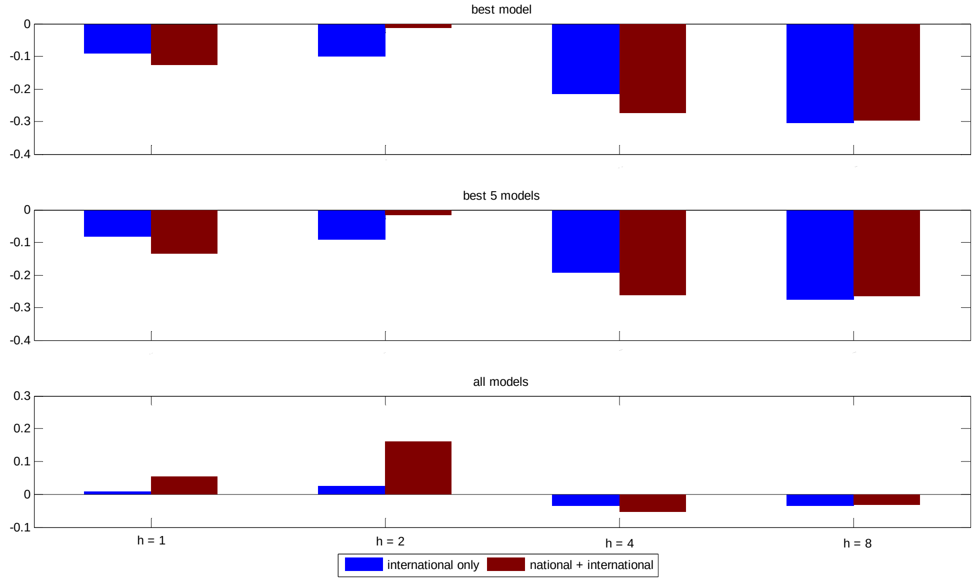



### (c) Commonality



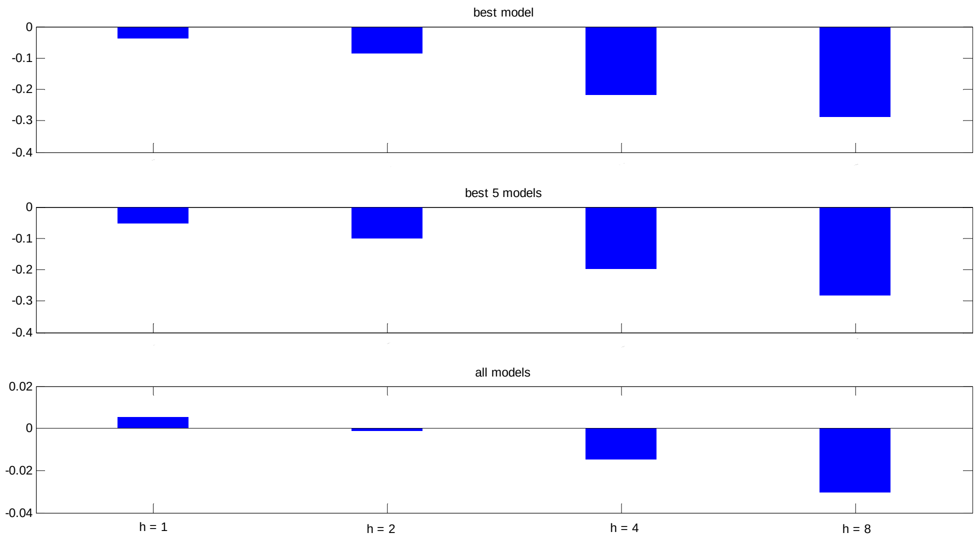

Notes: The red lines are the medians, the box indicate the lower and upper quartile values. Whiskers extend to the most extreme values of the data. Red crosses indicate outliers. Result shown in panels (b) and (c) are based on $r$ = 5. H indicates variable selection with hard thresholding. S(30), S(100), S(200) indicate variable selection with soft thresholding with 30, 100 and 200 variables.

## Figure 2: Forecast results - national versus international information




Notes: RMSE(model based on international and both national and international information)/RMSE(model based on national information)-1; for more details on the forecasting setup, see the text.

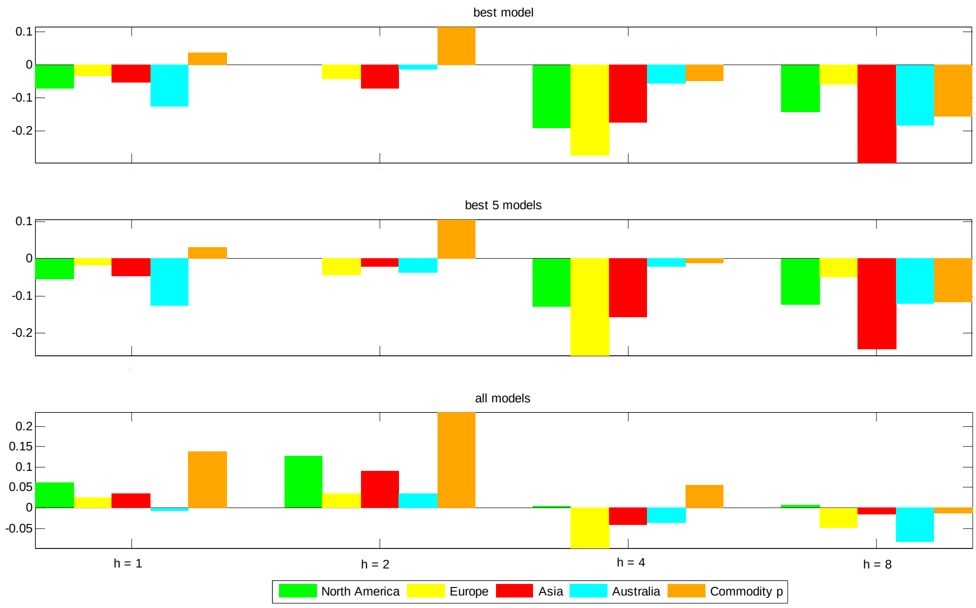
## Figure 3: Forecast results - data-rich methods versus trade-weights



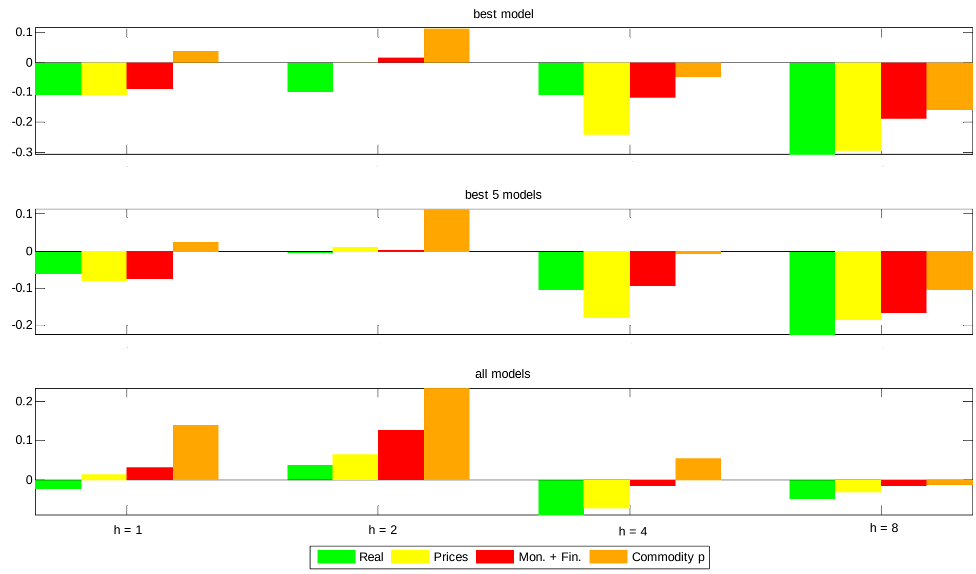

Notes: RMSE(model based on data-rich methods using only international data)/RMSE(model based on trade-weights)-1; for more details on the forecasting setup, see the text.

**Figure 4: Forecast results – national versus international information (regional and economic breakdown)**

**(a) Regional breakdown**



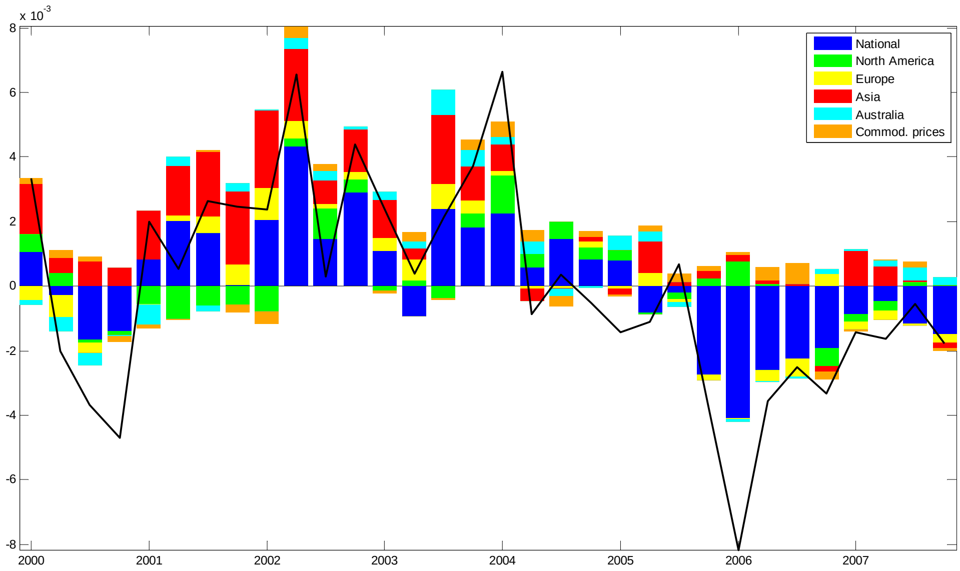

**(b) Economic breakdown**



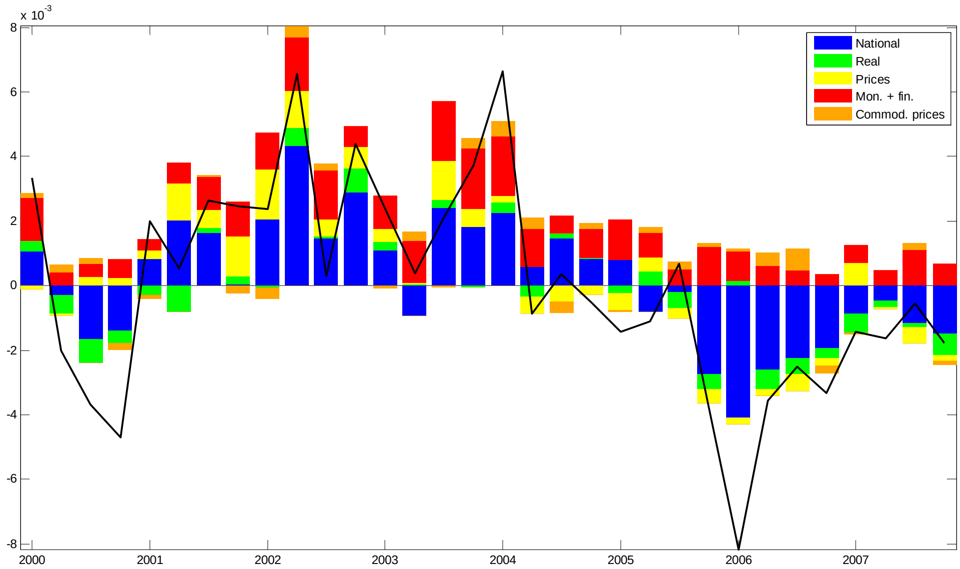

Notes: RMSE(model based on subsets of international data and both national and international data)/RMSE(model based on national data)-1; for more details on the forecasting setup, see the text.

**Figure 5: Historical contribution to New Zealand's GDP growth forecasts – RR(N), national and international information**

**(a) Regional breakdown**



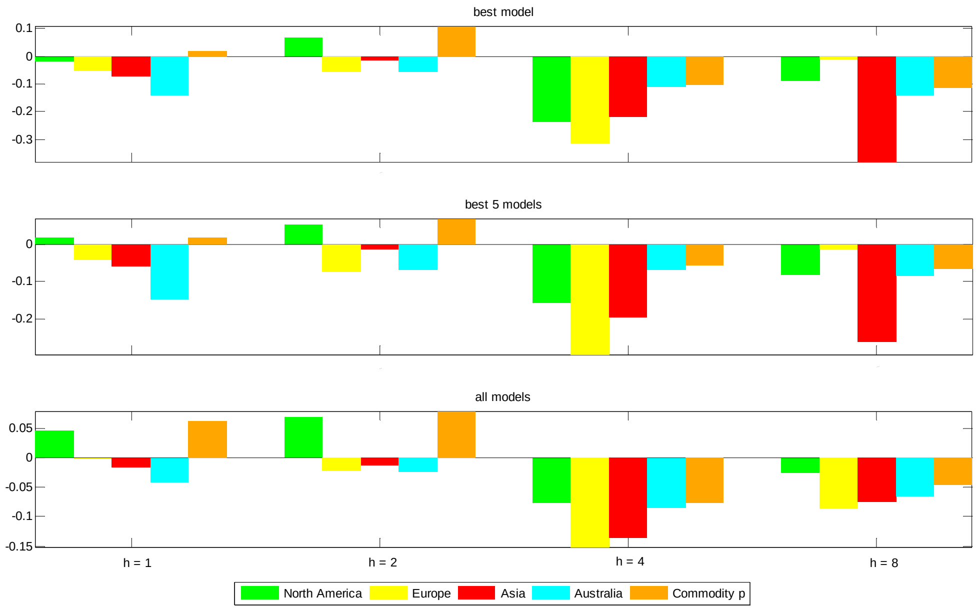

**(b) Economic breakdown**



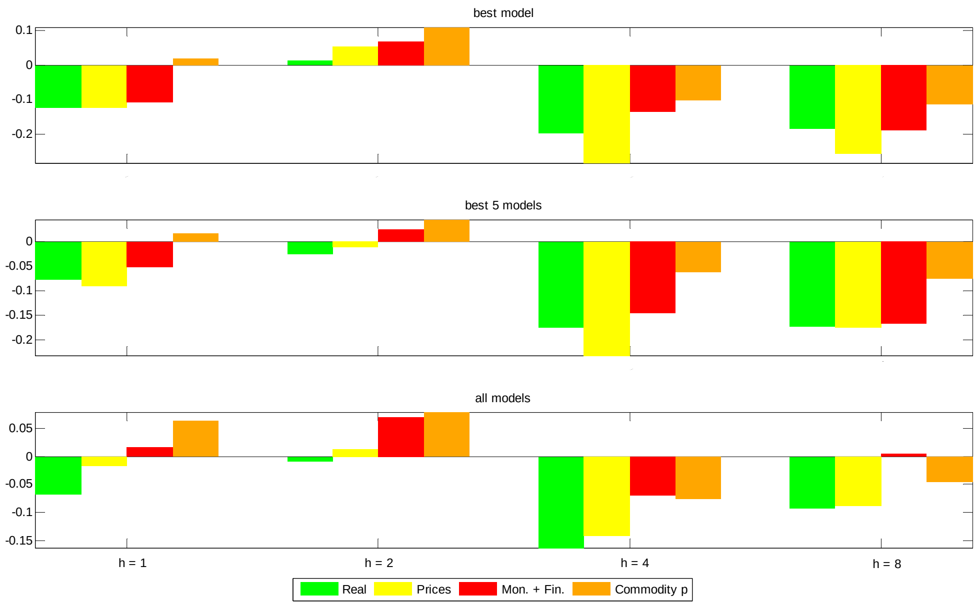

Notes: Based on a model using national and international data, RR(N), $h$ = 1; $p$ is chosen optimally; the black line is the forecast; deviations from the mean.

## Figure 6: Forecast results – national (without surveys) versus international information (regional and economic breakdown)

### (a) Regional breakdown



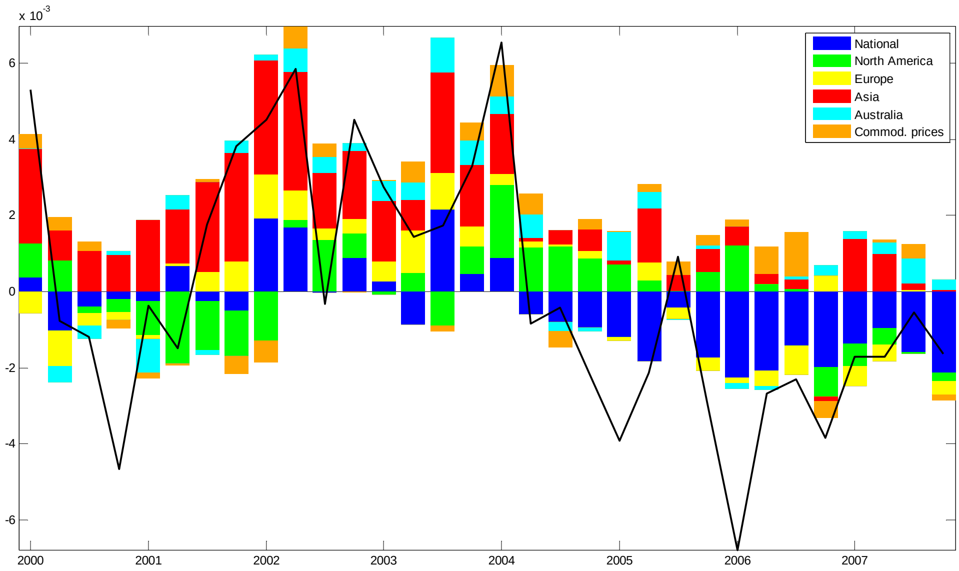

### (b) Economic breakdown



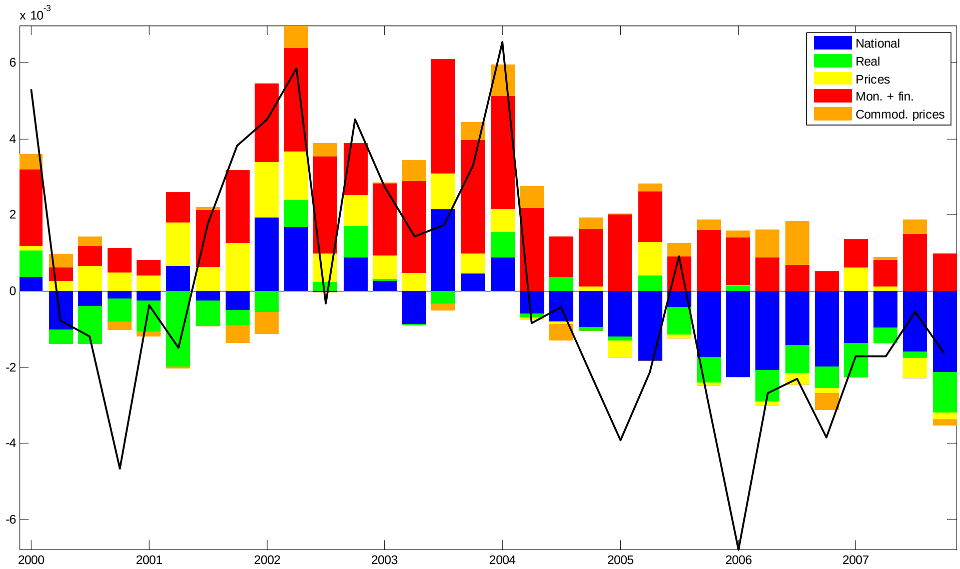

Notes: RMSE(model based on subsets of international data and both national and international data)/RMSE(model based on national data)-1; national survey data were removed from the national dataset; for more details on the forecasting setup, see the text.

**Figure 7: Historical contribution to New Zealand's GDP growth forecasts – RR(N), national (without surveys) and international information**

**(a) Regional breakdown**




**(b) Economic breakdown**




Notes: Based on a model using national and international data, RR(N), $h$ = 1; $p$ is chosen optimally; the black line is the forecast; deviations from the mean.

# Table A.1: International dataset

| # | Country | Variable | Treatment | Source |
|---|---|---|---|---|
| 1 | US | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 2 | | Private final consumption, real | 4 | OECD, Main Economic Indicators |
| 3 | | Gross fixed capital formation, real | 4 | OECD, Main Economic Indicators |
| 4 | | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 5 | | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 6 | | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 7 | | Employment | 4 | OECD, Main Economic Indicators |
| 8 | | Personal consumption, ex food and energy | 4 | OECD, Main Economic Indicators |
| 9 | | Capacity utilization | 3 | OECD, Main Economic Indicators |
| 10 | | Industrial production | 4 | OECD, Main Economic Indicators |
| 11 | | Income, nominal | 4 | OECD, Main Economic Indicators |
| 12 | | Orders of manufactured goods | 4 | OECD, Main Economic Indicators |
| 13 | | Labour force | 4 | OECD, Main Economic Indicators |
| 14 | | Total retail trade | 4 | OECD, Main Economic Indicators |
| 15 | | Dwelling productions | 4 | OECD, Main Economic Indicators |
| 16 | | Weekly hours worked | 4 | OECD, Main Economic Indicators |
| 17 | | Machinery production | 4 | OECD, Main Economic Indicators |
| 18 | | Intermediate goods production | 4 | OECD, Main Economic Indicators |
| 19 | | Non-durable goods production | 4 | OECD, Main Economic Indicators |
| 20 | | Total manufacturing production | 4 | OECD, Main Economic Indicators |
| 21 | | Total wholesale trade | 4 | OECD, Main Economic Indicators |
| 22 | | Ratio of inventories to shipments | 2 | OECD, Main Economic Indicators |
| 23 | | Unemployment rate, male | 4 | OECD, Main Economic Indicators |
| 24 | | Unemployment rate | 1 | OECD, Main Economic Indicators |
| 25 | | Consumer confidence index | 3 | OECD, Main Economic Indicators |
| 26 | | Purchasing managers index | 3 | OECD, Main Economic Indicators |
| 27 | | Consumer price index | 4 | OECD, Main Economic Indicators |
| 28 | | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 29 | | GDP deflator | 4 | OECD, Main Economic Indicators |
| 30 | | Personal consumption expenditures deflator | 4 | Bureau of Economic Analysis |
| 31 | | Durable goods deflator | 4 | Bureau of Economic Analysis |
| 32 | | Nondurable goods deflator | 4 | Bureau of Economic Analysis |
| 33 | | Gross private domestic investment deflator | 4 | Bureau of Economic Analysis |
| 34 | | Fixed investment deflator | 4 | Bureau of Economic Analysis |
| 35 | | Nonresidential investment deflator | 4 | Bureau of Economic Analysis |
| 36 | | Residential investment deflator | 4 | Bureau of Economic Analysis |
| 37 | | Government consumption expenditures and gross investment deflator | 4 | Bureau of Economic Analysis |
| 38 | | Producer price index | 4 | OECD, Main Economic Indicators |
| 39 | | Total labour costs | 4 | OECD, Main Economic Indicators |
| 40 | | Unit labour costs | 4 | OECD, Main Economic Indicators |
| 41 | | Export price index | 4 | OECD, Main Economic Indicators |
| 42 | | Import price index | 4 | OECD, Main Economic Indicators |
| 43 | | Benchmarked real output, total | 4 | OECD, Main Economic Indicators |
| 44 | | Hourly earnings, manufacturing | 4 | OECD, Main Economic Indicators |
| 45 | | CPI housing | 4 | OECD, Main Economic Indicators |
| 46 | | 3-month Certificates of deposit rate | 1 | OECD, Main Economic Indicators |
| 47 | | 10-year federal government securities (composite) rate | 1 | OECD, Main Economic Indicators |
| 48 | | 3-month euro-dollar deposit rate | 1 | OECD, Main Economic Indicators |
| 49 | | 10-year federal government bond yield | 1 | OECD, Main Economic Indicators |
| 50 | | Federal funds rate | 1 | OECD, Main Economic Indicators |
| 51 | | Prime rate | 1 | OECD, Main Economic Indicators |
| 52 | | M3 | 4 | OECD, Main Economic Indicators |
| 53 | | M2 | 4 | OECD, Main Economic Indicators |
| 54 | | M1 | 4 | OECD, Main Economic Indicators |
| 55 | | Real effective exchange rate | 4 | OECD, Main Economic Indicators |
| 56 | | Share price | 4 | IMF, International Financial Statistics |
| 57 | Canada | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 58 | | Private final consumption, real | 4 | OECD, Main Economic Indicators |
| 59 | | Gross fixed capital formation, real | 4 | OECD, Main Economic Indicators |
| 60 | | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 61 | | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 62 | | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 63 | | Change in inventories | 1 | OECD, Main Economic Indicators |
| 64 | | Employment | 4 | OECD, Main Economic Indicators |
| 65 | | Labour force | 4 | OECD, Main Economic Indicators |
| 66 | | Capacity utilisation | 3 | OECD, Main Economic Indicators |
| 67 | | Total production | 4 | OECD, Main Economic Indicators |
| 68 | | CLI total industry, ex construction | 4 | OECD, Main Economic Indicators |
| 69 | | Unemployment level | 4 | OECD, Main Economic Indicators |
| 70 | | Unemployment rate | 1 | OECD, Main Economic Indicators |
| 71 | | Total retail trade | 4 | OECD, Main Economic Indicators |
| 72 | | Manufacturing, finished goods stocks level | 2 | OECD, Main Economic Indicators |
| 73 | | Manufacturing, order books level | 2 | OECD, Main Economic Indicators |
| 74 | | Manufacturing, orders, inflow/demand tendency | 1 | OECD, Main Economic Indicators |
| 75 | | Manufacturing, production, future tendency | 1 | OECD, Main Economic Indicators |
| 76 | | Employees, manufacturing | 4 | OECD, Main Economic Indicators |
| 77 | | Total employees | 4 | OECD, Main Economic Indicators |
| 78 | | Permits issued for buildings | 4 | OECD, Main Economic Indicators |
| 79 | | Permits issued for dwellings | 4 | OECD, Main Economic Indicators |
| 80 | | Orders for total manufactured goods | 4 | OECD, Main Economic Indicators |
| 81 | | Consumer price index | 4 | OECD, Main Economic Indicators |
| 82 | | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 83 | | GDP deflator | 4 | OECD, Main Economic Indicators |
| 84 | | Total labour costs | 4 | OECD, Main Economic Indicators |
| 85 | | Producer price index | 4 | OECD, Main Economic Indicators |
| 86 | | Consumer price index, ex housing | 4 | OECD, Main Economic Indicators |
| 87 | | CPI services less housing | 4 | OECD, Main Economic Indicators |
| 88 | | Hourly earnings, manufacturing | 4 | OECD, Main Economic Indicators |
| 89 | | Wages and salaries, manufacturing | 4 | OECD, Main Economic Indicators |
| 90 | | 3-month prime corporate paper rate | 1 | OECD, Main Economic Indicators |
| 91 | | 10-year government bonds yield | 1 | OECD, Main Economic Indicators |
| 92 | | Central bank interest rate, < 24 hours | 1 | OECD, Main Economic Indicators |
| 93 | | Overnight money market financing rate | 1 | OECD, Main Economic Indicators |
| 94 | | M2+, gross | 4 | OECD, Main Economic Indicators |
| 95 | | M3, gross | 4 | OECD, Main Economic Indicators |
| 96 | | M1+, gross | 4 | OECD, Main Economic Indicators |
| 97 | | Narrow money M1 | 4 | OECD, Main Economic Indicators |
| 98 | | Real effective exchange rate, CPI based | 4 | OECD, Main Economic Indicators |
| 99 | | Real effective exchange rate, ULC based | 4 | OECD, Main Economic Indicators |
| 100 | | CAD/USD exchange rate, monthly average | 4 | OECD, Main Economic Indicators |

| | | | |
|---|---|---|---|
| 101 | Share price index | 4 | OECD, Main Economic Indicators |
| 102 Euro area | Gross domestic product, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 103 | Total demand, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 104 | Consumption, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 105 | Gross investment, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 106 | Government consumption, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 107 | Net public investment, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 108 | Labour force | 4 | AWM database, see Fagan *et al.* (2007) |
| 109 | Total employment | 4 | AWM database, see Fagan *et al.* (2007) |
| 110 | Labour productivity, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 111 | Exports, volume index | 4 | AWM database, see Fagan *et al.* (2007) |
| 112 | Imports, volume index | 4 | AWM database, see Fagan *et al.* (2007) |
| 113 | Household disposable income, real | 4 | AWM database, see Fagan *et al.* (2007) |
| 114 | Wealth | 4 | AWM database, see Fagan *et al.* (2007) |
| 115 | Employees, level | 4 | AWM database, see Fagan *et al.* (2007) |
| 116 | Retail sales | 4 | AWM database, see Fagan *et al.* (2007) |
| 117 | Industrial production, manufacturing | 4 | AWM database, see Fagan *et al.* (2007) |
| 118 | Industrial production, ex construction | 4 | AWM database, see Fagan *et al.* (2007) |
| 119 | Unemployment, level | 4 | AWM database, see Fagan *et al.* (2007) |
| 120 | Industrial confidence | 1 | European Commission |
| 121 | Consumer price index | 4 | AWM database, see Fagan *et al.* (2007) |
| 122 | Consumer price index, energy | 4 | AWM database, see Fagan *et al.* (2007) |
| 123 | Consumer price index, ex energy | 4 | AWM database, see Fagan *et al.* (2007) |
| 124 | Growth of consumption deflator | 1 | AWM database, see Fagan *et al.* (2007) |
| 125 | Import of goods and services deflator | 4 | AWM database, see Fagan *et al.* (2007) |
| 126 | Consumption deflator | 4 | AWM database, see Fagan *et al.* (2007) |
| 127 | Unit labour costs | 4 | AWM database, see Fagan *et al.* (2007) |
| 128 | Exports of goods and services deflator | 4 | AWM database, see Fagan *et al.* (2007) |
| 129 | Compensation to employees | 4 | AWM database, see Fagan *et al.* (2007) |
| 130 | Producer price index, manufacturing | 4 | AWM database, see Fagan *et al.* (2007) |
| 131 | Producer price index | 4 | AWM database, see Fagan *et al.* (2007) |
| 132 | Wage rate | 2 | AWM database, see Fagan *et al.* (2007) |
| 133 | GDP deflator | 4 | AWM database, see Fagan *et al.* (2007) |
| 134 | Short-term nominal interest rate | 1 | AWM database, see Fagan *et al.* (2007) |
| 135 | Long-term interest rate | 1 | AWM database, see Fagan *et al.* (2007) |
| 136 | 10-year interest rate | 1 | AWM database, see Fagan *et al.* (2007) |
| 137 | 5-year interest rate | 1 | AWM database, see Fagan *et al.* (2007) |
| 138 | 2 -year interest rate | 1 | AWM database, see Fagan *et al.* (2007) |
| 139 | 2-year yield spreads | 1 | AWM database, see Fagan *et al.* (2007) |
| 140 | 5-year yield spreads | 1 | AWM database, see Fagan *et al.* (2007) |
| 141 | 10-year yield spreads | 1 | AWM database, see Fagan *et al.* (2007) |
| 142 | Household's savings ratio | 4 | AWM database, see Fagan *et al.* (2007) |
| 143 | Consumer loans | 4 | AWM database, see Fagan *et al.* (2007) |
| 144 | Loans to NFCs | 4 | AWM database, see Fagan *et al.* (2007) |
| 145 | M1 | 4 | AWM database, see Fagan *et al.* (2007) |
| 146 | M2 | 4 | AWM database, see Fagan *et al.* (2007) |
| 147 | M2-M1 | 4 | AWM database, see Fagan *et al.* (2007) |
| 148 | M3 | 4 | AWM database, see Fagan *et al.* (2007) |
| 149 | M3-M2 | 4 | AWM database, see Fagan *et al.* (2007) |
| 150 | M3-M1 | 4 | AWM database, see Fagan *et al.* (2007) |
| 151 | Nominal effective exchange rate | 4 | AWM database, see Fagan *et al.* (2007) |
| 152 | Real effective exchange rate | 4 | AWM database, see Fagan *et al.* (2007) |
| 153 | Stock price | 4 | AWM database, see Fagan *et al.* (2007) |
| 154 UK | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 155 | Private final consumption, real | 4 | OECD, Main Economic Indicators |
| 156 | Gross fixed capital formation, real | 4 | OECD, Main Economic Indicators |
| 157 | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 158 | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 159 | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 160 | Industrial production | 4 | OECD, Main Economic Indicators |
| 161 | Total retail trade | 4 | OECD, Main Economic Indicators |
| 162 | Household disposable income | 4 | Office for National Statistics |
| 163 | Consumer spending | 4 | Office for National Statistics |
| 164 | Workforce jobs | 4 | Office for National Statistics |
| 165 | Productivity, whole economy | 4 | Office for National Statistics |
| 166 | Industrial production index, manufacturing | 4 | Office for National Statistics |
| 167 | Consumer confidence | 1 | Office for National Statistics |
| 168 | Unemployment level | 4 | OECD, Main Economic Indicators |
| 169 | Unemployment rate | 1 | OECD, Main Economic Indicators |
| 170 | Unemployment rate, survey-based | 1 | OECD, Main Economic Indicators |
| 171 | Passenger car registrations | 4 | OECD, Main Economic Indicators |
| 172 | Work started for dwellings | 4 | OECD, Main Economic Indicators |
| 173 | Composite leading indicator | 4 | OECD, Main Economic Indicators |
| 174 | Consumer confidence | 1 | Office for National Statistics |
| 175 | Business optimism | 1 | Confederation of British Industry |
| 176 | Consumer price index | 4 | OECD, Main Economic Indicators |
| 177 | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 178 | Consumer price index, HICP all items | 4 | OECD, Main Economic Indicators |
| 179 | Consumer price index, services less housing | 4 | OECD, Main Economic Indicators |
| 180 | GDP deflator | 4 | OECD, Main Economic Indicators |
| 181 | Producer price index, manufacturing output | 4 | OECD, Main Economic Indicators |
| 182 | Total labour costs | ` | OECD, Main Economic Indicators |
| 183 | Unit labour costs | 4 | OECD, Main Economic Indicators |
| 184 | Average earning index | 4 | Office for National Statistics |
| 185 | Weekly earnings, manufacturing | 4 | OECD, Main Economic Indicators |
| 186 | Weekly earnings, private sector | 4 | OECD, Main Economic Indicators |
| 187 | Export price index | 4 | Office for National Statistics |
| 188 | Retail price index | 4 | OECD, Main Economic Indicators |
| 189 | Retail price index, less mortgage interest rates | 4 | OECD, Main Economic Indicators |
| 190 | CPI housing | 4 | OECD, Main Economic Indicators |
| 191 | 3-month mean, LIBID/LIBOR rate | 1 | OECD, Main Economic Indicators |
| 192 | 3-month treasury bills discount rate | 1 | OECD, Main Economic Indicators |
| 193 | 10-year central government securities yield | 1 | OECD, Main Economic Indicators |
| 194 | Sterling overnight interbank rate | 1 | OECD, Main Economic Indicators |
| 195 | Bank base rate, 4 UK banks | 1 | OECD, Main Economic Indicators |
| 196 | 3-month treasury bills discount rate, sterling | 1 | Bank of England |
| 197 | Sterling overnight interbank rate, mean LIBID/LIBOR | 1 | OECD, Main Economic Indicators |
| 198 | Prime lending rate, major banks | 1 | Bank of England |
| 199 | M2 | 4 | OECD, Main Economic Indicators |
| 200 | M4 | 4 | OECD, Main Economic Indicators |

| | | | | |
|---|---|---|---|---|
| 201 | | Real effective exchange rate, CPI based | 4 | OECD, Main Economic Indicators |
| 202 | | Real effective exchange rate, ULC based | 4 | OECD, Main Economic Indicators |
| 203 | | GBP/ USD exchange rate, monthly average | 4 | OECD, Main Economic Indicators |
| 204 | | Share price index, FTSE 100 | 4 | OECD, Main Economic Indicators |
| 205 | Australia | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 206 | | Private final consumption, real | 4 | OECD, Main Economic Indicators |
| 207 | | Gross fixed capital formation, real | 4 | OECD, Main Economic Indicators |
| 208 | | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 209 | | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 210 | | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 211 | | Total employment | 4 | OECD, Main Economic Indicators |
| 212 | | Employees,household survey | 4 | OECD, Main Economic Indicators |
| 213 | | Labor force | 4 | OECD, Main Economic Indicators |
| 214 | | Manufacturers and wholesale trade sales | 4 | Australian Bureau of Statistics |
| 215 | | Gross national income | 4 | Australian Bureau of Statistics |
| 216 | | Productivity, GDP per employed person | 4 | Australian Bureau of Statistics |
| 217 | | Production in total manufacturing, level | 4 | OECD, Main Economic Indicators |
| 218 | | Production in total manufacturing, index | 4 | OECD, Main Economic Indicators |
| 219 | | Production in total service sector | 4 | OECD, Main Economic Indicators |
| 220 | | Total retail trade | 4 | OECD, Main Economic Indicators |
| 221 | | Unemployment level | 4 | OECD, Main Economic Indicators |
| 222 | | Unemployment rate | 2 | OECD, Main Economic Indicators |
| 223 | | Melbourne Westpac consumer sentiment index | 3 | Westpac Melbourne Institute |
| 224 | | 10-year central government securities yield | 4 | Westpac Melbourne Institute |
| 225 | | Consumer price index | 4 | OECD, Main Economic Indicators |
| 226 | | GDP deflator | 4 | OECD, Main Economic Indicators |
| 227 | | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 228 | | Export price index | 4 | Australian Bureau of Statistics |
| 229 | | Import price index | 4 | Australian Bureau of Statistics |
| 230 | | Producer price index, food, beverage and tobacco | 4 | OECD, Main Economic Indicators |
| 231 | | Producer price index, manufacturing | 4 | OECD, Main Economic Indicators |
| 232 | | Total labour costs | 4 | OECD, Main Economic Indicators |
| 233 | | Unit labour costs | 4 | OECD, Main Economic Indicators |
| 234 | | Real output, total | 4 | OECD, Main Economic Indicators |
| 235 | | CPI housing | 4 | OECD, Main Economic Indicators |
| 236 | | 90-day bank accepted bills yield | 1 | OECD, Main Economic Indicators |
| 237 | | 10-year government bonds yield | 1 | OECD, Main Economic Indicators |
| 238 | | M3 | 4 | OECD, Main Economic Indicators |
| 239 | | Broad money | 4 | OECD, Main Economic Indicators |
| 240 | | M1 | 4 | OECD, Main Economic Indicators |
| 241 | | Narrow money | 4 | OECD, Main Economic Indicators |
| 242 | | Real effective exchange rates, CPI based | 4 | OECD, Main Economic Indicators |
| 243 | | Real effective exchange rates, ULC based | 4 | OECD, Main Economic Indicators |
| 244 | | AUD/USD exchange rate, monthly average | 4 | OECD, Main Economic Indicators |
| 245 | | Share price index, S&P ASX 200 | 4 | OECD, Main Economic Indicators |
| 246 | | Share price index, S&P ASE industrials | 4 | OECD, Main Economic Indicators |
| 247 | Japan | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 248 | | Private final consumption, real | 4 | OECD, Main Economic Indicators |
| 249 | | Private residential investment, real | 4 | OECD, Main Economic Indicators |
| 250 | | Private non-residential investment, real | 4 | OECD, Main Economic Indicators |
| 251 | | Private inventory, real | 2 | OECD, Main Economic Indicators |
| 252 | | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 253 | | Public investment, real | 4 | OECD, Main Economic Indicators |
| 254 | | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 255 | | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 256 | | Consumer confidence | 3 | OECD, Main Economic Indicators |
| 257 | | Total employment | 4 | OECD, Main Economic Indicators |
| 258 | | Labour force | 4 | OECD, Main Economic Indicators |
| 259 | | Capacity utilisation | 3 | OECD, Main Economic Indicators |
| 260 | | Total production | 4 | OECD, Main Economic Indicators |
| 261 | | Retail trade | 4 | OECD, Main Economic Indicators |
| 262 | | Unemployment level | 4 | OECD, Main Economic Indicators |
| 263 | | Unemployment rate | 2 | OECD, Main Economic Indicators |
| 264 | | Manufactured goods orders | 4 | OECD, Main Economic Indicators |
| 265 | | Business condition survey | 1 | Ministry of Finance, Japan |
| 266 | | Consumer price index | 4 | OECD, Main Economic Indicators |
| 267 | | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 268 | | Unit labor costs | 4 | OECD, Main Economic Indicators |
| 269 | | Producer price index, manufacturing | 4 | OECD, Main Economic Indicators |
| 270 | | CPI housing | 4 | OECD, Main Economic Indicators |
| 271 | | GDP deflator | 4 | OECD, Main Economic Indicators |
| 272 | | Export unit price | 4 | Bank of Japan |
| 273 | | Import unit price | 4 | Bank of Japan |
| 274 | | 3-month interest rate, certificates of deposit | 1 | OECD, Main Economic Indicators |
| 275 | | 10-year interest bearing government bonds yield | 1 | OECD, Main Economic Indicators |
| 276 | | Central bank discount rate | 1 | OECD, Main Economic Indicators |
| 277 | | Overnight call rate, uncollateralized | 1 | OECD, Main Economic Indicators |
| 278 | | Broad money, M4 | 4 | OECD, Main Economic Indicators |
| 279 | | M2 | 4 | OECD, Main Economic Indicators |
| 280 | | M2 + Certificates of deposit | 4 | OECD, Main Economic Indicators |
| 281 | | Broadly defined liquidity | 4 | OECD, Main Economic Indicators |
| 282 | | M1 | 4 | OECD, Main Economic Indicators |
| 283 | | Real effective exchange rate, CPI based | 4 | OECD, Main Economic Indicators |
| 284 | | Real effective exchange rate, ULC based | 4 | OECD, Main Economic Indicators |
| 285 | | JPY/USD exchange rate, monthly average | 4 | OECD, Main Economic Indicators |
| 286 | | Share price index, TSE TOPIX | 4 | OECD, Main Economic Indicators |
| 287 | Korea | Gross domestic product, real | 4 | OECD, Main Economic Indicators |
| 288 | | Private final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 289 | | Gross fixed capital formation, real | 4 | OECD, Main Economic Indicators |
| 290 | | Government final consumption expenditure, real | 4 | OECD, Main Economic Indicators |
| 291 | | Changes in inventories | 1 | OECD, Main Economic Indicators |
| 292 | | Exports, volume index | 4 | OECD, Main Economic Indicators |
| 293 | | Imports, volume index | 4 | OECD, Main Economic Indicators |
| 294 | | Total employment | 4 | OECD, Main Economic Indicators |
| 295 | | Composite leading indicators, excluding constructions | 4 | OECD, Main Economic Indicators |
| 296 | | Machinery orders received | 4 | National Statistical Office |
| 297 | | Labour productivity index | 4 | National Statistical Office |
| 298 | | Industrial production | 4 | Bank of Korea |
| 299 | | Manufacturing production index | 4 | National Statistical Office |
| 300 | | Retail sales | 4 | Bank of Korea |

| | | | | |
|---|---|---|---|---|
| 301 | | Labour force | 4 | OECD, Main Economic Indicators |
| 302 | | Capacity utilisation | 3 | OECD, Main Economic Indicators |
| 303 | | Total retail trade | 4 | OECD, Main Economic Indicators |
| 304 | | Unemployment level | 4 | OECD, Main Economic Indicators |
| 305 | | Unemployment rate | 1 | OECD, Main Economic Indicators |
| 306 | | Leading composite index | 4 | National Statistical Office |
| 307 | | Consumer price index | 4 | OECD, Main Economic Indicators |
| 308 | | Housing purchase price index | 4 | Bank of Korea |
| 309 | | Export price index | 4 | Bank of Korea |
| 310 | | Import price index | 4 | Bank of Korea |
| 311 | | Consumer price index, ex food and energy | 4 | OECD, Main Economic Indicators |
| 312 | | GDP deflator | 4 | IMF, International Financial Statistics |
| 313 | | Producer price index | 4 | OECD, Main Economic Indicators |
| 314 | | Unit labour costs | 4 | OECD, Main Economic Indicators |
| 315 | | Benchmarked real output, total | 4 | OECD, Main Economic Indicators |
| 316 | | CPI housing | 4 | OECD, Main Economic Indicators |
| 317 | | 10-year government bond yields | 1 | OECD, Main Economic Indicators |
| 318 | | M2 | 4 | OECD, Main Economic Indicators |
| 319 | | M3 | 4 | OECD, Main Economic Indicators |
| 320 | | M1 | 4 | OECD, Main Economic Indicators |
| 321 | | Real effective exchange rate, ULC based | 4 | OECD, Main Economic Indicators |
| 322 | | KRW/USD exchange rate, monthly average | 4 | OECD, Main Economic Indicators |
| 323 | | Share price index, Kse KOSPI index | 4 | OECD, Main Economic Indicators |
| 324 | China | Gross domestic product, real | 4 | National Bureau of Statistics, China |
| 325 | | Exports, volume index | 4 | National Bureau of Statistics, China |
| 326 | | Imports, volume index | 4 | National Bureau of Statistics, China |
| 327 | | Employment | 4 | National Bureau of Statistics, China |
| 328 | | Saving rate | 2 | National Bureau of Statistics, China |
| 329 | | Unemployment rate | 2 | National Bureau of Statistics, China |
| 330 | | Unemployment level | 4 | National Bureau of Statistics, China |
| 331 | | Disposable income | 4 | National Bureau of Statistics, China |
| 332 | | Industrial production | 4 | National Bureau of Statistics, China |
| 333 | | Production of cement | 4 | OECD, Main Economic Indicators |
| 334 | | Composite leading indicator | 4 | National Bureau of Statistics, China |
| 335 | | Consumer price index (%YOY) | 1 | National Bureau of Statistics, China |
| 336 | | Average wage | 4 | National Bureau of Statistics, China |
| 337 | | Overnight interest rate | 1 | OECD, Main Economic Indicators |
| 338 | | Prime lending rate | 1 | People's Bank of China |
| 339 | | Discount rate | 1 | People's Bank of China |
| 340 | | M2 | 4 | OECD, Main Economic Indicators |
| 341 | | M1 | 4 | OECD, Main Economic Indicators |
| 342 | | SDR reserve assets | 4 | OECD, Main Economic Indicators |
| 343 | | M0, Currency in circulation | 4 | OECD, Main Economic Indicators |
| 344 | | CNY/USD exchange rate | 4 | OECD, Main Economic Indicators |
| 345 | Hong Kong | Gross domestic product, real | 4 | Census and Statistics Department of Hong Kong |
| 346 | | Private consumption expenditure, real | 4 | Census and Statistics Department of Hong Kong |
| 347 | | Gross fixed capital formation, real | 4 | IMF, International Financial Statistics |
| 348 | | Government consumption expenditure, real | 4 | Census and Statistics Department of Hong Kong |
| 349 | | Exports, volume index | 4 | IMF, International Financial Statistics |
| 350 | | Imports, volume index | 4 | IMF, International Financial Statistics |
| 351 | | Retail sales | 4 | Census and Statistics Department of Hong Kong |
| 352 | | Labour force | 4 | Census and Statistics Department of Hong Kong |
| 353 | | Unemployment level | 4 | Census and Statistics Department of Hong Kong |
| 354 | | Unemployment rate | 1 | Census and Statistics Department of Hong Kong |
| 355 | | Merchandise trade balance | 2 | Census and Statistics Department of Hong Kong |
| 356 | | Industrial production, manufacturing | 4 | Census and Statistics Department of Hong Kong |
| 357 | | Consumer price index | 4 | IMF, International Financial Statistics |
| 358 | | GDP deflator | 4 | IMF, International Financial Statistics |
| 359 | | Export unit value index | 4 | Census and Statistics Department of Hong Kong |
| 360 | | Import unit value index | 4 | Census and Statistics Department of Hong Kong |
| 361 | | Producer price index, manufacturing industries | 4 | Census and Statistics Department of Hong Kong |
| 362 | | 3-month interbank offered rate | 1 | Hong Kong Monetary Authority |
| 363 | | Prime lending rate | 1 | Hong Kong Monetary Authority |
| 364 | | M0 | 4 | Hong Kong Monetary Authority |
| 365 | | M1 | 4 | Hong Kong Monetary Authority |
| 366 | | HKD/USD exchange rate | 4 | IMF, International Financial Statistics |
| 367 | | Effective exchange rate indices, trade-weighted | 4 | Census and Statistics Department of Hong Kong |
| 368 | | Share price index, Hang Seng | 4 | Census and Statistics Department of Hong Kong |
| 369 | Malaysia | Gross domestic product, real | 4 | IMF, International Financial Statistics |
| 370 | | Industrial production | 4 | IMF, International Financial Statistics |
| 371 | | Exports, volume index | 4 | IMF, International Financial Statistics |
| 372 | | Imports, volume index | 4 | IMF, International Financial Statistics |
| 373 | | Industrial production, manufacturing | 4 | Department of Statistics, Malaysia |
| 374 | | Consumer price index | 4 | IMF, International Financial Statistics |
| 375 | | Producer price index | 4 | IMF, International Financial Statistics |
| 376 | | Manufacturing sector salaries and wages paid | 4 | Department of Statistics, Malaysia |
| 377 | | 3-month fixed deposit rate | 1 | IMF, International Financial Statistics |
| 378 | | 3-month treasury bills rate | 1 | IMF, International Financial Statistics |
| 379 | | Overnight interbank rate | 1 | IMF, International Financial Statistics |
| 380 | | Reserve money | 4 | IMF, International Financial Statistics |
| 381 | | M0 | 4 | Bank of Malaysia |
| 382 | | M1 | 4 | Bank of Malaysia |
| 383 | | M2 | 4 | Bank of Malaysia |
| 384 | | M3 | 4 | Bank of Malaysia |
| 385 | | MYR/USD official exchange rate | 4 | IMF, International Financial Statistics |
| 386 | | Share price index | 4 | IMF, International Financial Statistics |
| 387 | Singapore | Gross domestic product, real | 4 | Department of Statistics, Singapore |
| 388 | | Private consumption expenditure, real | 4 | Department of Statistics, Singapore |
| 389 | | Gross fixed capital formation, real | 4 | Department of Statistics, Singapore |
| 390 | | Government consumption expenditure, real | 4 | Department of Statistics, Singapore |
| 391 | | Composite leading indicator | 4 | Department of Statistics, Singapore |
| 392 | | Exports, volume index | 4 | Department of Statistics, Singapore |
| 393 | | Imports, volume index | 4 | Department of Statistics, Singapore |
| 394 | | Unemployment rate | 4 | Department of Statistics, Singapore |
| 395 | | Retail sales index | 4 | Department of Statistics, Singapore |
| 396 | | Industrial productions, ex rubber processing | 4 | Department of Statistics, Singapore |
| 397 | | Business expectations, manufacturing next 6 months | 1 | Department of Statistics, Singapore |
| 398 | | Composite leading indicators | 4 | Department of Statistics, Singapore |
| 399 | | Consumer price index | 4 | IMF, International Financial Statistics |
| 400 | | Producer price index | 4 | IMF, International Financial Statistics |

| | | | | |
|---|---|---|---|---|
| 401 | | Export unit value | 4 | IMF, International Financial Statistics |
| 402 | | Import unit value | 4 | IMF, International Financial Statistics |
| 403 | | Domestic supply price index | 4 | Department of Statistics, Singapore |
| 404 | | Unit labour costs | 4 | Department of Statistics, Singapore |
| 405 | | 3-month interbank rate | 1 | IMF, International Financial Statistics |
| 406 | | 3-month deposit rate | 1 | IMF, International Financial Statistics |
| 407 | | 3-month treasury bill rate | 1 | Monetary Authority of Singapore |
| 408 | | Prime lending rate | 1 | Monetary Authority of Singapore |
| 409 | | M0 | 4 | Monetary Authority of Singapore |
| 410 | | M1 | 4 | Monetary Authority of Singapore |
| 411 | | M2 | 4 | Monetary Authority of Singapore |
| 412 | | M3 | 4 | Monetary Authority of Singapore |
| 413 | | SGD/USD exchange rate | 4 | Monetary Authority of Singapore |
| 414 | | Share price index | 4 | IMF, International Financial Statistics |
| 415 | Taiwan | Gross domestic product, real | 4 | Directorate General of Budget, Accounting and Statistics |
| 416 | | Private consumption, real | 4 | Directorate General of Budget, Accounting and Statistics |
| 417 | | Gross domestic fixed capital formation, real | 4 | Directorate General of Budget, Accounting and Statistics |
| 418 | | Gross national production | 4 | Directorate General of Budget, Accounting and Statistics |
| 419 | | Government consumption, real | 4 | Directorate General of Budget, Accounting and Statistics |
| 420 | | Exports, volume index | 4 | Directorate General of Budget, Accounting and Statistics |
| 421 | | Imports, volume index | 4 | Directorate General of Budget, Accounting and Statistics |
| 422 | | Industrial production index | 4 | Ministry of Economic Affairs |
| 423 | | Industrial production index, manufacturing | 4 | Ministry of Economic Affairs |
| 424 | | Export orders received | 4 | Directorate General of Budget, Accounting and Statistics |
| 425 | | Unemployment rate | 1 | Directorate General of Budget, Accounting and Statistics |
| 426 | | Labour force | 4 | Directorate General of Budget, Accounting and Statistics |
| 427 | | Employment | 4 | Directorate General of Budget, Accounting and Statistics |
| 428 | | Unemployment level | 4 | Directorate General of Budget, Accounting and Statistics |
| 429 | | Labour productivity index | 4 | Directorate General of Budget, Accounting and Statistics |
| 430 | | Composite leading indicators | 4 | Directorate General of Budget, Accounting and Statistics |
| 431 | | Consumer price index | 4 | Directorate General of Budget, Accounting and Statistics |
| 432 | | Consumer price index, ex food and energy | 4 | Directorate General of Budget, Accounting and Statistics |
| 433 | | Average monthly earnings, manufacturing | 4 | Directorate General of Budget, Accounting and Statistics |
| 434 | | Unit labour costs | 4 | Directorate General of Budget, Accounting and Statistics |
| 435 | | Export price index | 4 | Directorate General of Budget, Accounting and Statistics |
| 436 | | Import price index | 4 | Directorate General of Budget, Accounting and Statistics |
| 437 | | GDP deflator | 4 | Directorate General of Budget, Accounting and Statistics |
| 438 | | Discount rate | 1 | Central Bank of China |
| 439 | | 90-day money market rate | 1 | Central Bank of China |
| 440 | | Prime lending rate, 5 major banks | 1 | Central Bank of China |
| 441 | | M0 | 4 | Central Bank of China |
| 442 | | M1 | 4 | Central Bank of China |
| 443 | | M2 | 4 | Central Bank of China |
| 444 | | TWD/USD exchange rate | 4 | Central Bank of China |
| 445 | Commodity | Aluminum | 4 | London Metal Exchange |
| 446 | | Cooper | 4 | London Metal Exchange |
| 447 | | Lead | 4 | London Metal Exchange |
| 448 | | Nickel | 4 | London Metal Exchange |
| 449 | | Tin | 4 | London Metal Exchange |
| 450 | | Zinc | 4 | London Metal Exchange |
| 451 | | Cotton | 4 | US Department of Agriculture |
| 452 | | Nylon yarn | 4 | Taiwan Economic Journal |
| 453 | | Polyester | 2 | Taiwan Economic Journal |
| 454 | | Wool | 4 | Australian Wool Council |
| 455 | | Fuel oil | 4 | Dow Jones energy services |
| 456 | | Gas oil | 4 | ICIS Pricing |
| 457 | | Gasoline | 4 | Dow Jones energy services |
| 458 | | Jet kerosene-cargos | 4 | ICIS Pricing |
| 459 | | Crude oil | 4 | ICIS Pricing |
| 460 | | Palm oil | 4 | Public Ledger |
| 461 | | Sunflower oil | 4 | Public Ledger |
| 462 | | Soya oil | 4 | Public Ledger |
| 463 | | Peanut oil | 4 | Public Ledger |
| 464 | | Linseed oil | 4 | Public Ledger |
| 465 | | Raw sugar | 4 | Public Ledger |
| 466 | | Rubber | 4 | Malaysian Rubber Exchange |
| 467 | | Coffee | 4 | International Coffee Organization |
| 468 | | Cocoa | 4 | International Cocoa Organization |
| 469 | | Soybeans | 4 | US Department of Agriculture |
| 470 | | Rapeseed | 4 | Public Ledger |
| 471 | | Wheat (hard Kansas, Cts./Bu) | 4 | US Department of Agriculture |
| 472 | | Wheat 2 (soft) | 4 | US Department of Agriculture |
| 473 | | Rice | 4 | Public Ledger |
| 474 | | Platinum | 4 | Metal Bulletin |
| 475 | | Silver | 4 | London Metal Exchange |
| 476 | | Palladium | 4 | Metal Bulletin |
| 477 | | Corn | 4 | US Department of Agriculture |
| 478 | | Gold | 4 | London Metal Exchange |
| 479 | | Pork bellies | 4 | US Department of Agriculture |
| 480 | | Eggs | 4 | US Department of Agriculture |
| 481 | | Oats | 4 | US Department of Agriculture |
| 482 | | Live hogs | 4 | Standard & Poor's |

Notes: Series are treated as follows. 1: no transformation, 2: first difference, 3: log, 4: log difference.

# Table A.2: Forecast results for subsets of international data

## (a) Only North American data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.8525 | 1.0225 | 0.8881 | 1.4948 | 0.9836 | 0.8989 | 1.0672 | 1.3866 |
| trade1 | 1.115 | 1.416 | 1.963 | 2.665 | - | - | - | - |
| trade2 | 1.115 | 1.416 | 1.955 | 2.655 | - | - | - | - |
| trade3 | 1.115 | 1.416 | 1.955 | 2.660 | - | - | - | - |
| PC (r = 1) | 1.033 | 1.045 | 0.836 | **0.825** | 0.984 | 0.933 | 1.030 | 1.170 |
| PC (r = 2) | 1.049 | 1.056 | 0.948 | 1.077 | 1.016 | 0.978 | 1.142 | 1.160 |
| PC (r = 3) | 1.082 | 1.090 | 0.978 | 1.119 | 1.016 | 1.000 | 1.149 | 1.180 |
| PC (r = 5) | 1.148 | 1.382 | 1.418 | 1.948 | 1.033 | 0.876 | 1.119 | 0.928 |
| WPCBN (r = 1) | 1.049 | 1.079 | **0.828** | 0.871 | 0.984 | 0.989 | 1.030 | 0.959 |
| WPCBN (r = 2) | 1.082 | 1.101 | 0.940 | 1.139 | 1.000 | 0.944 | 1.045 | 0.938 |
| WPCBN (r = 3) | 1.000 | 1.067 | 0.925 | 1.392 | **0.869** | 1.000 | 0.933 | **0.820** |
| WPCBN (r = 5) | 1.049 | 1.236 | 1.127 | 1.464 | 0.852 | 0.955 | 1.425 | 1.165 |
| WPCSW (r = 1) | 1.016 | 1.067 | 0.925 | 0.871 | 1.066 | 1.202 | 1.187 | 1.134 |
| WPCSW (r = 2) | 1.000 | 1.000 | 0.993 | 1.170 | 0.967 | 1.022 | 1.037 | 0.948 |
| WPCSW (r = 3) | 1.000 | 1.045 | 0.978 | 1.093 | 1.049 | 1.000 | 1.157 | 1.314 |
| WPCSW (r = 5) | **0.984** | 1.056 | 0.851 | 1.134 | 1.016 | 1.011 | 1.194 | 1.031 |
| TPH (r = 1) | 1.066 | 1.236 | 1.209 | 1.206 | 0.918 | 0.910 | 1.134 | 1.175 |
| TPH (r = 2) | 1.131 | 1.337 | 1.336 | 1.521 | 0.885 | 0.899 | 1.134 | 1.155 |
| TPH (r = 3) | 1.164 | 1.404 | 1.313 | 1.515 | 0.902 | 0.978 | 1.119 | 1.191 |
| TPH (r = 5) | 1.344 | 1.449 | 1.396 | 1.701 | 0.918 | 1.101 | 1.164 | 1.340 |
| TPS (30 variables, r = 1) | 1.016 | 1.056 | 0.963 | 1.072 | 1.131 | 1.157 | 1.134 | 1.005 |
| TPS (30 variables, r = 2) | 1.131 | 1.056 | 1.388 | 1.402 | 1.066 | 1.135 | 1.224 | 1.242 |
| TPS (30 variables, r = 3) | 1.098 | 1.135 | 1.396 | 1.454 | 0.984 | 1.079 | 1.164 | 1.258 |
| TPS (30 variables, r = 5) | 1.131 | 1.236 | 1.500 | 1.675 | 1.016 | 1.011 | 1.104 | 1.206 |
| PLS (k = 1) | 1.098 | 1.180 | 0.873 | 1.160 | 0.967 | 0.876 | 0.978 | 1.129 |
| PLS (k = 2) | 1.328 | 1.416 | 1.194 | 1.608 | 1.115 | 1.011 | 1.052 | 1.186 |
| PLS (k = 3) | 1.475 | 1.697 | 1.545 | 1.876 | 0.967 | **0.787** | **0.813** | 1.160 |
| PLS (k = 5) | 1.410 | 1.607 | 1.560 | 1.670 | 0.984 | 1.090 | 1.052 | 1.098 |
| RR ($v_2$ = 0.25) | 1.836 | 1.910 | 2.142 | 2.670 | 1.033 | 1.000 | 1.104 | 1.165 |
| RR ($v_2$ = N) | 1.098 | 1.225 | 1.127 | 1.330 | 0.918 | 0.820 | 0.851 | 1.082 |
| RR ($v_2$ = 5N) | 1.033 | 1.056 | 0.948 | 1.067 | 0.934 | 0.888 | 0.896 | 1.015 |
| RR ($v_2$ = 10N) | 1.016 | 1.022 | 0.955 | 1.031 | 0.951 | 0.910 | 0.925 | 1.000 |
| EN (30 variables, $v_2$ = 0.25) | 1.311 | 1.584 | 1.970 | 1.794 | 0.984 | 0.933 | 1.224 | 1.005 |

## (b) Only European data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 1.0656 | 1.1573 | 1.1269 | 1.2526 | 1.0492 | 0.9326 | 1.0149 | 1.2526 |
| trade1 | 1.033 | 1.112 | 1.052 | 1.057 | - | - | - | - |
| trade2 | 1.049 | 1.079 | 1.127 | 1.113 | - | - | - | - |
| trade3 | 1.049 | 1.067 | 1.127 | 1.119 | - | - | - | - |
| PC (r = 1) | 1.000 | 1.045 | 1.060 | 1.021 | 0.951 | 0.865 | 1.007 | 1.129 |
| PC (r = 2) | 1.000 | 0.955 | 0.836 | 1.124 | 0.951 | 0.843 | 1.067 | 1.134 |
| PC (r = 3) | 1.016 | 0.944 | 0.873 | 1.139 | 0.951 | 0.888 | 1.075 | 1.299 |
| PC (r = 5) | 1.000 | 0.876 | 1.045 | 1.309 | 0.984 | 0.955 | 1.239 | 1.572 |
| WPCBN (r = 1) | 1.000 | 1.011 | 1.015 | **0.887** | 0.951 | 0.978 | 0.948 | **0.866** |
| WPCBN (r = 2) | 1.000 | 0.910 | 0.724 | 0.938 | 0.934 | 1.056 | 0.970 | 1.144 |
| WPCBN (r = 3) | 0.934 | 0.865 | 0.769 | 1.119 | 0.951 | 1.079 | 1.239 | 1.624 |
| WPCBN (r = 5) | 1.066 | 0.944 | 1.075 | 1.077 | 1.115 | 1.000 | 1.403 | 1.804 |
| WPCSW (r = 1) | 1.033 | 1.067 | 1.037 | 1.021 | 1.066 | 1.135 | 1.082 | 1.082 |
| WPCSW (r = 2) | 0.934 | 0.910 | 0.955 | 1.155 | 0.984 | 0.865 | 1.007 | 0.923 |
| WPCSW (r = 3) | 1.000 | 1.034 | 1.030 | 1.186 | 0.984 | 0.899 | 1.030 | 1.222 |
| WPCSW (r = 5) | 0.967 | 0.955 | 1.022 | 1.170 | 0.934 | 0.820 | 1.007 | 1.232 |
| TPH (r = 1) | 0.967 | 1.000 | 0.948 | 0.995 | 0.885 | 0.944 | 1.112 | 1.211 |
| TPH (r = 2) | 0.951 | 0.989 | 1.030 | 1.196 | 0.852 | 0.921 | 1.224 | 1.134 |
| TPH (r = 3) | 0.951 | 1.000 | 1.187 | 1.072 | **0.836** | 0.933 | 1.194 | 1.232 |
| TPH (r = 5) | 0.951 | 1.011 | 1.291 | 1.180 | 0.967 | 1.000 | 1.119 | 1.232 |
| TPS (30 variables, r = 1) | 1.000 | 1.045 | 1.007 | 1.175 | 1.246 | 1.146 | 1.201 | 1.144 |
| TPS (30 variables, r = 2) | 1.000 | 1.169 | 1.060 | 1.299 | 1.213 | 1.135 | 1.209 | 1.119 |
| TPS (30 variables, r = 3) | 1.033 | 1.090 | 1.201 | 1.284 | 1.230 | 1.146 | 1.261 | 1.134 |
| TPS (30 variables, r = 5) | 1.016 | 1.079 | 1.224 | 1.381 | 1.246 | 1.191 | 1.284 | 1.113 |
| PLS (k = 1) | 0.918 | 0.854 | 0.701 | 1.031 | 0.934 | 0.831 | 0.910 | 1.072 |
| PLS (k = 2) | **0.869** | **0.775** | 0.672 | 1.175 | 1.000 | 0.910 | 1.030 | 1.253 |
| PLS (k = 3) | 1.033 | 0.899 | 0.799 | 1.088 | 0.918 | **0.753** | **0.716** | 1.242 |
| PLS (k = 5) | 1.279 | 1.079 | 0.828 | 0.969 | 0.967 | 1.034 | 0.925 | 1.216 |
| RR ($v_2$ = 0.25) | 1.770 | 1.461 | 1.164 | 1.304 | 1.000 | 0.910 | 0.925 | 1.278 |
| RR ($v_2$ = N) | 0.902 | 0.809 | **0.649** | 1.041 | 0.902 | 0.787 | 0.784 | 1.119 |
| RR ($v_2$ = 5N) | 0.902 | 0.843 | 0.694 | 0.995 | 0.918 | 0.854 | 0.851 | 1.021 |
| RR ($v_2$ = 10N) | 0.918 | 0.876 | 0.776 | 0.985 | 0.934 | 0.876 | 0.896 | 1.000 |
| EN (30 variables, $v_2$ = 0.25) | 1.016 | 1.022 | 0.903 | 0.948 | 1.016 | 0.831 | 1.045 | 1.031 |

## (c) Only Asian data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 1.3279 | 1.4157 | 1.1940 | 1.6082 | 1.1148 | 1.0112 | 1.0522 | 1.1856 |
| trade1 | 1.066 | 1.124 | 0.955 | 1.191 | - | - | - | - |
| trade2 | 1.115 | 1.213 | 0.978 | 1.294 | - | - | - | - |
| trade3 | 1.115 | 1.213 | 0.978 | 1.294 | - | - | - | - |
| PC (r = 1) | 0.984 | 0.944 | 0.851 | 1.000 | 0.967 | 0.933 | 1.075 | 1.165 |
| PC (r = 2) | 1.115 | 1.258 | 0.910 | 1.216 | 0.984 | 0.876 | 1.052 | 1.376 |
| PC (r = 3) | 1.033 | 1.101 | 0.828 | 1.129 | 0.951 | 0.876 | 0.918 | 1.309 |
| PC (r = 5) | 1.098 | **0.899** | 0.873 | 1.423 | 0.984 | **0.809** | **0.828** | 1.742 |
| WPCBN (r = 1) | 0.984 | 0.955 | 0.881 | 0.969 | 0.984 | 1.011 | 1.119 | 1.196 |
| WPCBN (r = 2) | 1.115 | 1.202 | 0.896 | 1.186 | 0.984 | 0.933 | 1.157 | 1.490 |
| WPCBN (r = 3) | 1.000 | 1.169 | 1.067 | 1.268 | 0.967 | 0.899 | 1.127 | 1.289 |
| WPCBN (r = 5) | 0.951 | 1.056 | 1.000 | 1.345 | 0.934 | 0.888 | 1.216 | 1.443 |
| WPCSW (r = 1) | 1.016 | 0.944 | **0.739** | 0.923 | 1.000 | 1.067 | 1.045 | 1.067 |
| WPCSW (r = 2) | 1.066 | 1.157 | 0.806 | 1.005 | 0.934 | 0.820 | 0.910 | 1.211 |
| WPCSW (r = 3) | 1.066 | 1.180 | 0.873 | 1.124 | 0.967 | 0.843 | 1.082 | 1.284 |
| WPCSW (r = 5) | 1.049 | 1.146 | 0.993 | 1.356 | 1.000 | 0.944 | 1.179 | 1.531 |
| TPH (r = 1) | 1.180 | 1.191 | 1.187 | 1.959 | 0.934 | 0.843 | 1.030 | 1.129 |
| TPH (r = 2) | 1.213 | 1.213 | 1.396 | 1.985 | 0.934 | 0.854 | 1.045 | 1.237 |
| TPH (r = 3) | 1.213 | 1.213 | 1.418 | 1.979 | 0.967 | 0.989 | 1.179 | 1.227 |
| TPH (r = 5) | 1.279 | 1.315 | 1.634 | 2.041 | 0.984 | 0.989 | 1.194 | 1.572 |
| TPS (30 variables, r = 1) | 0.951 | 1.022 | 1.254 | 0.923 | 1.066 | 1.124 | 1.381 | 1.144 |
| TPS (30 variables, r = 2) | 1.115 | 1.202 | 1.455 | 1.412 | 1.098 | 1.247 | 1.433 | 1.247 |
| TPS (30 variables, r = 3) | 1.377 | 1.292 | 1.440 | 1.387 | 1.098 | 1.258 | 1.463 | 1.258 |
| TPS (30 variables, r = 5) | 1.377 | 1.270 | 1.478 | 1.392 | 1.180 | 1.236 | 1.410 | 1.191 |
| PLS (k = 1) | 1.082 | 1.213 | 1.060 | 1.082 | 0.951 | 0.843 | 0.940 | 1.052 |
| PLS (k = 2) | **0.852** | 0.966 | 0.799 | 1.093 | 0.984 | 0.854 | 0.948 | 1.242 |
| PLS (k = 3) | 0.885 | 0.933 | 0.761 | 0.881 | **0.852** | 0.865 | 0.896 | 1.046 |
| PLS (k = 5) | 0.984 | 1.270 | 0.940 | **0.722** | 0.967 | 1.112 | 0.918 | 0.907 |
| RR ($v_2$ = 0.25) | 1.115 | 1.225 | 1.254 | 0.969 | 0.967 | 0.933 | 1.015 | 1.062 |
| RR ($v_2$ = N) | 0.902 | 1.011 | 0.799 | 0.923 | 0.885 | **0.809** | **0.828** | 1.046 |
| RR ($v_2$ = 5N) | 0.918 | 0.933 | 0.813 | 0.943 | 0.918 | 0.865 | 0.851 | 1.000 |
| RR ($v_2$ = 10N) | 0.934 | 0.944 | 0.851 | 0.974 | 0.934 | 0.888 | 0.881 | 0.995 |
| EN (30 variables, $v_2$ = 0.25) | 0.918 | 1.067 | 1.060 | 1.186 | 0.951 | 0.933 | 1.045 | **0.711** |

## (d) Only Australian data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.8689 | 0.7753 | 0.6716 | 1.1753 | 1.0000 | 0.9101 | 1.0299 | 1.2526 |
| trade1 | 0.934 | 1.022 | 1.142 | 0.990 | - | - | - | - |
| trade2 | 0.984 | 1.011 | 0.955 | 0.928 | - | - | - | - |
| trade3 | 0.984 | 1.011 | 0.955 | 0.928 | - | - | - | - |
| PC (r = 1) | 0.934 | 1.000 | 1.007 | 1.005 | 1.000 | 1.000 | 1.201 | 1.325 |
| PC (r = 2) | 0.902 | 1.011 | 1.060 | 1.031 | 1.000 | 1.034 | 1.321 | 1.655 |
| PC (r = 3) | 0.902 | 0.955 | 1.067 | 1.046 | 0.984 | 1.056 | 1.381 | 1.552 |
| PC (r = 5) | 0.820 | 0.989 | 1.134 | 1.113 | 0.918 | 1.056 | 1.552 | 1.531 |
| WPCBN (r = 1) | 0.918 | 1.034 | 1.007 | 0.979 | 0.934 | 0.865 | 1.000 | 1.139 |
| WPCBN (r = 2) | 0.902 | 0.978 | 1.067 | 1.077 | 0.934 | 0.876 | 1.067 | 1.361 |
| WPCBN (r = 3) | 0.885 | 0.888 | 1.075 | 1.093 | 0.934 | 0.865 | 1.104 | 1.067 |
| WPCBN (r = 5) | **0.787** | 1.034 | 0.985 | 1.170 | 1.000 | 0.933 | 1.164 | 1.474 |
| WPCSW (r = 1) | 0.852 | 0.854 | 1.030 | 1.026 | 1.246 | 1.315 | 1.276 | 1.082 |
| WPCSW (r = 2) | 0.918 | 0.933 | 1.060 | 1.021 | 1.098 | 1.079 | 1.022 | 1.376 |
| WPCSW (r = 3) | 1.033 | 1.022 | 1.045 | 1.077 | 1.098 | 1.112 | 1.269 | 1.608 |
| WPCSW (r = 5) | 0.984 | 1.011 | 1.075 | 1.010 | 1.016 | 0.978 | 1.067 | 1.335 |
| TPH (r = 1) | 0.951 | 0.989 | 0.993 | 1.088 | 0.951 | 0.955 | 0.963 | 1.000 |
| TPH (r = 2) | 1.000 | 1.045 | 0.948 | 0.964 | 0.951 | 0.989 | 1.007 | 1.304 |
| TPH (r = 3) | 0.984 | 1.056 | 0.940 | 1.505 | 0.934 | 1.011 | 1.052 | 1.335 |
| TPH (r = 5) | 1.016 | 1.146 | 1.336 | 1.758 | 1.016 | 1.124 | 1.313 | 1.376 |
| TPS (30 variables, r = 1) | 1.016 | 1.067 | 1.052 | 0.985 | 1.016 | 1.056 | 1.284 | 1.258 |
| TPS (30 variables, r = 2) | 1.016 | 1.045 | 1.037 | 0.990 | 1.148 | 1.135 | 1.284 | 1.392 |
| TPS (30 variables, r = 3) | 1.049 | 1.056 | 1.060 | 1.041 | 1.295 | 1.180 | 1.231 | 1.448 |
| TPS (30 variables, r = 5) | 0.984 | 1.090 | 0.955 | 0.985 | 1.295 | 1.315 | 1.216 | 1.418 |
| PLS (k = 1) | **0.787** | 0.831 | 1.104 | 1.072 | 0.934 | 0.831 | 0.940 | 1.103 |
| PLS (k = 2) | 0.869 | 0.899 | 0.925 | 1.005 | 1.016 | 1.000 | 1.306 | 1.345 |
| PLS (k = 3) | 0.918 | 0.888 | 0.903 | 0.959 | 0.902 | **0.730** | 0.970 | 1.402 |
| PLS (k = 5) | 1.066 | 1.124 | 0.918 | 0.990 | 0.902 | 0.944 | 1.119 | 1.284 |
| RR ($v_2$ = 0.25) | 1.803 | 1.831 | 1.776 | 1.005 | 1.016 | 0.933 | 1.157 | 1.366 |
| RR ($v_2$ = N) | 0.803 | **0.775** | **0.843** | **0.825** | **0.869** | 0.775 | 0.933 | 1.180 |
| RR ($v_2$ = 5N) | 0.836 | 0.854 | 0.955 | 0.912 | 0.918 | 0.865 | **0.925** | 1.036 |
| RR ($v_2$ = 10N) | 0.869 | 0.888 | 0.985 | 0.948 | 0.934 | 0.888 | 0.940 | 1.005 |
| EN (30 variables, $v_2$ = 0.25) | 1.246 | 1.281 | 1.261 | 1.026 | 1.000 | 0.820 | 1.246 | 1.134 |

## (e) Only international real data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.0061 | 0.0045 | 0.0034 | 0.0024 | 0.0061 | 0.0045 | 0.0034 | 0.0024 |
| trade1 | 1.049 | 1.090 | 0.948 | 0.923 | - | - | - | - |
| trade2 | 1.049 | 1.090 | 0.948 | 0.933 | - | - | - | - |
| trade3 | 1.049 | 1.079 | 0.940 | 0.923 | - | - | - | - |
| PC (r = 1) | 1.066 | 1.056 | 0.970 | 0.990 | 0.951 | 0.888 | 1.037 | 1.149 |
| PC (r = 2) | 1.066 | 1.146 | 0.985 | **0.912** | 0.967 | 0.921 | 1.194 | 1.387 |
| PC (r = 3) | 1.049 | 1.045 | 0.866 | 1.412 | 0.984 | 0.966 | 1.231 | 1.443 |
| PC (r = 5) | 1.148 | 1.146 | 1.187 | 1.608 | 1.000 | 1.000 | 1.254 | 1.464 |
| WPCBN (r = 1) | 1.049 | 1.034 | 0.970 | 1.026 | 0.934 | 0.899 | 1.030 | 1.170 |
| WPCBN (r = 2) | 1.066 | 1.112 | 0.970 | 0.938 | 0.951 | 0.933 | 1.142 | 1.077 |
| WPCBN (r = 3) | 1.082 | 1.034 | 0.963 | 1.309 | 0.967 | 0.910 | 1.119 | 1.160 |
| WPCBN (r = 5) | 1.131 | 1.124 | 1.030 | 1.325 | 0.934 | 0.933 | 1.358 | 1.320 |
| WPCSW (r = 1) | 1.049 | 1.101 | 1.045 | 0.979 | 1.033 | 1.056 | 1.134 | 1.103 |
| WPCSW (r = 2) | 1.066 | 1.124 | 1.030 | 1.005 | 0.967 | 0.933 | 1.172 | 1.289 |
| WPCSW (r = 3) | 1.033 | 1.112 | 1.075 | 1.418 | 0.984 | 0.966 | 1.284 | 1.603 |
| WPCSW (r = 5) | 1.016 | 1.180 | 0.933 | 1.186 | 0.951 | 0.910 | 1.104 | 1.402 |
| TPH (r = 1) | 1.066 | 1.079 | 1.097 | 1.124 | 0.984 | 0.933 | 1.149 | 1.253 |
| TPH (r = 2) | 1.066 | 1.045 | 1.299 | 1.423 | 0.984 | 0.910 | 1.164 | 1.402 |
| TPH (r = 3) | 1.180 | 1.112 | 1.254 | 1.314 | 1.016 | 0.978 | 1.134 | 1.423 |
| TPH (r = 5) | 1.180 | 1.157 | 1.254 | 1.351 | 1.098 | 1.101 | 1.246 | 1.381 |
| TPS (30 variables, r = 1) | 1.016 | 1.022 | 0.948 | 1.345 | 1.213 | 1.180 | 1.037 | 1.103 |
| TPS (30 variables, r = 2) | 1.033 | 1.079 | 1.112 | 1.273 | 1.197 | 1.348 | 1.075 | 1.031 |
| TPS (30 variables, r = 3) | 1.098 | 1.079 | 1.149 | 1.459 | 1.180 | 1.337 | 1.000 | 1.021 |
| TPS (30 variables, r = 5) | 1.262 | 1.404 | 1.291 | 1.536 | 1.148 | 1.371 | 0.963 | 1.062 |
| PLS (k = 1) | 1.066 | 1.045 | **0.851** | 1.242 | 0.967 | 0.843 | 0.963 | 1.124 |
| PLS (k = 2) | 0.967 | 0.966 | 0.940 | 1.330 | 1.049 | 0.933 | 0.978 | 1.222 |
| PLS (k = 3) | **0.885** | **0.933** | 0.881 | 1.299 | **0.902** | **0.708** | 0.821 | 1.046 |
| PLS (k = 5) | 0.967 | 1.034 | **0.851** | 1.139 | 0.918 | 0.955 | 0.903 | 1.031 |
| RR ($v_2$ = 0.25) | 0.934 | 1.067 | 1.000 | 1.381 | 0.951 | 0.865 | 0.993 | 1.103 |
| RR ($v_2$ = N) | 0.934 | 1.067 | 0.993 | 1.381 | 0.902 | 0.798 | **0.799** | 1.046 |
| RR ($v_2$ = 5N) | 0.918 | 1.045 | 0.978 | 1.361 | 0.934 | 0.888 | 0.866 | 0.995 |
| RR ($v_2$ = 10N) | 0.918 | 1.034 | 0.963 | 1.340 | 0.951 | 0.910 | 0.910 | 0.990 |
| EN (30 variables, $v_2$ = 0.25) | 1.082 | 0.978 | 0.896 | 1.278 | 0.984 | 0.865 | 0.918 | **0.701** |

## (f) Only international price data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.9180 | 0.8202 | 0.9552 | 1.1856 | 0.8361 | 0.8989 | 0.7164 | 1.0567 |
| trade1 | 0.902 | 0.955 | 1.007 | 1.010 | - | - | - | - |
| trade2 | 0.885 | 0.933 | 0.918 | 0.985 | - | - | - | - |
| trade3 | 0.885 | 0.921 | 0.925 | 0.979 | - | - | - | - |
| PC (r = 1) | 0.902 | **0.888** | 0.821 | 0.866 | 0.951 | 0.888 | 1.015 | 1.134 |
| PC (r = 2) | 0.951 | 0.966 | 0.828 | 0.809 | 0.984 | 0.933 | 1.127 | 1.098 |
| PC (r = 3) | 0.869 | 0.955 | 0.940 | 1.155 | 0.967 | 0.978 | 1.127 | 1.268 |
| PC (r = 5) | **0.803** | 1.101 | 1.373 | 1.459 | 0.934 | 0.921 | 1.090 | 1.500 |
| WPCBN (r = 1) | 0.918 | 0.910 | 0.858 | 0.897 | 0.951 | 0.921 | 0.896 | 1.129 |
| WPCBN (r = 2) | 0.918 | 1.079 | 1.082 | **0.784** | 0.967 | 0.910 | 1.015 | 1.211 |
| WPCBN (r = 3) | 0.967 | 1.056 | 1.276 | 1.696 | 0.918 | 0.966 | 1.157 | 1.412 |
| WPCBN (r = 5) | 1.098 | 1.056 | 1.231 | 1.892 | 1.016 | 1.045 | 1.463 | 1.562 |
| WPCSW (r = 1) | 0.984 | 0.989 | 0.963 | 0.933 | 1.049 | 1.191 | 1.149 | 1.093 |
| WPCSW (r = 2) | 0.951 | 0.978 | 0.888 | 0.825 | 0.934 | 0.876 | 0.955 | **0.985** |
| WPCSW (r = 3) | 0.934 | 1.011 | 1.022 | 0.964 | 1.000 | 0.888 | 0.925 | 1.170 |
| WPCSW (r = 5) | 0.934 | 1.067 | 1.090 | 0.979 | 0.951 | 0.876 | **0.836** | 1.103 |
| TPH (r = 1) | 0.869 | 1.022 | 0.940 | 1.000 | 0.984 | 0.854 | 0.970 | 1.072 |
| TPH (r = 2) | 0.885 | 1.079 | 1.082 | 1.144 | 0.984 | 0.854 | 0.970 | 1.196 |
| TPH (r = 3) | 0.902 | 1.202 | 0.933 | 1.474 | 1.000 | 0.888 | 0.985 | 1.072 |
| TPH (r = 5) | 0.984 | 1.337 | 1.187 | 1.680 | 1.049 | 0.888 | 1.321 | 1.273 |
| TPS (30 variables, r = 1) | 0.967 | 1.067 | 0.978 | 0.959 | 1.131 | 1.157 | 1.269 | 1.227 |
| TPS (30 variables, r = 2) | 0.967 | 1.090 | 1.291 | 1.340 | 1.164 | 1.090 | 1.269 | 1.242 |
| TPS (30 variables, r = 3) | 1.016 | 1.202 | 1.433 | 1.536 | 1.230 | 1.090 | 1.299 | 1.278 |
| TPS (30 variables, r = 5) | 0.984 | 1.326 | 1.500 | 1.588 | 1.230 | 1.157 | 1.321 | 1.206 |
| PLS (k = 1) | 0.852 | 0.910 | 0.799 | 0.923 | 0.934 | 0.843 | 0.940 | 1.062 |
| PLS (k = 2) | 0.852 | 1.022 | 0.888 | 1.495 | 0.984 | 0.899 | 1.067 | 1.387 |
| PLS (k = 3) | 0.885 | 0.989 | 0.828 | 1.356 | 0.934 | 0.843 | 0.851 | 1.206 |
| PLS (k = 5) | 0.951 | 1.090 | 0.851 | 0.990 | 0.967 | 1.056 | 0.955 | 1.124 |
| RR ($v_2$ = 0.25) | 1.098 | 1.247 | 1.007 | 1.330 | 0.984 | 0.933 | 1.007 | 1.268 |
| RR ($v_2$ = N) | **0.803** | 0.921 | **0.746** | 1.129 | **0.869** | **0.787** | 0.866 | 1.149 |
| RR ($v_2$ = 5N) | 0.852 | **0.888** | 0.813 | 0.990 | 0.902 | 0.843 | 0.873 | 1.031 |
| RR ($v_2$ = 10N) | 0.902 | 0.910 | 0.858 | 0.979 | 0.918 | 0.876 | 0.903 | 1.010 |
| EN (30 variables, $v_2$ = 0.25) | 0.967 | 1.056 | 1.022 | 1.345 | 0.951 | 0.876 | 1.022 | 1.175 |

## (g) Only international monetary and financial data

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.9672 | 0.9663 | 0.9403 | 1.3299 | 1.0492 | 0.9326 | 0.9776 | 1.2216 |
| trade1 | 1.066 | 1.169 | 1.590 | 2.232 | - | - | - | - |
| trade2 | 1.000 | 1.011 | 1.052 | 1.196 | - | - | - | - |
| trade3 | 1.000 | 1.011 | 1.045 | 1.191 | - | - | - | - |
| PC (r = 1) | 1.082 | 1.112 | 0.933 | 0.742 | 1.000 | 1.000 | 1.119 | 1.211 |
| PC (r = 2) | 1.016 | 0.978 | 0.739 | 0.979 | 0.967 | 0.899 | 1.127 | 1.474 |
| PC (r = 3) | 1.066 | 1.000 | **0.679** | 1.015 | 0.984 | 0.921 | 1.149 | 1.603 |
| PC (r = 5) | 1.148 | 1.079 | 0.873 | 1.222 | 1.082 | 0.978 | 1.388 | 1.742 |
| WPCBN (r = 1) | 1.033 | 1.022 | 0.873 | **0.711** | 1.000 | 1.034 | 1.157 | 1.186 |
| WPCBN (r = 2) | 1.049 | 1.112 | 1.194 | 0.871 | 0.918 | 0.910 | 1.134 | 1.031 |
| WPCBN (r = 3) | 1.049 | 1.169 | 1.082 | 1.273 | 0.967 | 0.910 | 1.164 | 1.381 |
| WPCBN (r = 5) | 1.115 | 1.135 | 1.246 | 1.222 | 1.033 | 1.090 | 1.052 | 1.562 |
| WPCSW (r = 1) | 1.066 | 1.101 | 1.000 | 0.923 | 1.000 | 1.067 | 1.045 | 1.077 |
| WPCSW (r = 2) | 1.082 | 1.169 | 1.037 | 1.242 | 0.951 | 0.831 | 0.903 | 1.299 |
| WPCSW (r = 3) | 1.197 | 1.315 | 1.276 | 1.335 | 0.951 | 0.921 | 0.970 | 1.227 |
| WPCSW (r = 5) | 1.180 | 1.270 | 1.209 | 1.057 | 0.984 | 0.888 | 1.082 | 1.325 |
| TPH (r = 1) | 1.115 | 1.191 | 1.194 | 1.247 | 0.951 | 0.865 | 1.104 | 1.227 |
| TPH (r = 2) | 1.197 | 1.258 | 1.127 | 1.345 | 0.984 | 0.809 | 1.149 | 1.289 |
| TPH (r = 3) | 1.180 | 1.258 | 1.030 | 1.366 | 0.984 | 0.843 | 1.067 | 1.139 |
| TPH (r = 5) | 1.098 | 1.494 | 1.224 | 1.351 | **0.820** | 0.831 | 1.015 | 1.273 |
| TPS (30 variables, r = 1) | 1.000 | 1.022 | 0.910 | 1.165 | 0.934 | 1.213 | 1.030 | 1.088 |
| TPS (30 variables, r = 2) | 0.902 | 1.079 | 0.918 | 1.247 | 1.033 | 1.146 | 1.045 | 1.093 |
| TPS (30 variables, r = 3) | **0.820** | 1.169 | 1.194 | 1.330 | 1.016 | 1.270 | 1.164 | 0.943 |
| TPS (30 variables, r = 5) | 1.033 | 1.315 | 1.194 | 1.325 | 1.016 | 1.191 | 1.097 | **0.897** |
| PLS (k = 1) | 1.180 | 1.315 | 1.351 | 1.186 | 0.951 | 0.854 | 0.940 | 1.077 |
| PLS (k = 2) | 1.066 | 1.157 | 1.127 | 1.253 | 1.049 | 0.933 | 1.015 | 1.253 |
| PLS (k = 3) | 1.049 | 1.247 | 1.216 | 1.103 | 0.885 | 0.843 | **0.791** | 1.134 |
| PLS (k = 5) | 1.148 | 1.303 | 1.187 | 1.304 | 0.934 | 1.056 | 0.940 | 1.057 |
| RR ($v_2$ = 0.25) | 1.689 | 1.494 | 1.470 | 1.876 | 0.967 | 0.910 | 0.963 | 1.149 |
| RR ($v_2$ = N) | 0.984 | 1.056 | 1.067 | 1.088 | 0.902 | **0.798** | 0.821 | 1.082 |
| RR ($v_2$ = 5N) | 0.984 | 1.000 | 0.970 | 1.010 | 0.934 | 0.865 | 0.881 | 1.021 |
| RR ($v_2$ = 10N) | 0.967 | **0.966** | 0.933 | 1.010 | 0.934 | 0.888 | 0.896 | 1.005 |
| EN (30 variables, $v_2$ = 0.25) | 1.033 | 1.247 | 1.358 | 1.206 | 1.049 | 0.888 | 0.993 | 1.113 |

## (h) Only world commodity prices

| | International | | | | National + International | | | |
|---|---|---|---|---|---|---|---|---|
| | h = 1 | h = 2 | h = 4 | h = 8 | h = 1 | h = 2 | h = 4 | h = 8 |
| Univar. benchm. (RMSE/h) | 0.8525 | 0.9663 | 0.7985 | 1.0928 | 0.9836 | 0.8539 | 0.9478 | 1.2423 |
| PC (r = 1) | 0.984 | 1.034 | 1.037 | 1.031 | 0.951 | 0.865 | 0.993 | 1.119 |
| PC (r = 2) | 1.082 | 1.067 | 1.075 | 1.041 | 0.984 | 0.910 | 1.097 | 1.284 |
| PC (r = 3) | 1.082 | 1.079 | 1.060 | 1.026 | 1.000 | 0.910 | 1.276 | 1.325 |
| PC (r = 5) | 1.049 | 1.090 | 1.104 | 1.144 | 1.016 | 0.989 | 1.366 | 1.686 |
| WPCBN (r = 1) | 0.984 | 1.022 | **0.985** | 0.943 | 0.951 | 0.910 | 1.007 | 1.134 |
| WPCBN (r = 2) | 1.033 | 1.067 | 1.015 | 0.948 | 0.934 | 0.831 | 1.194 | 1.144 |
| WPCBN (r = 3) | 1.066 | 1.079 | 1.015 | 0.974 | 1.016 | 0.854 | 1.142 | 1.134 |
| WPCBN (r = 5) | 1.000 | 1.146 | 1.097 | 1.010 | 1.033 | 0.921 | 1.224 | 1.232 |
| WPCSW (r = 1) | 1.033 | 1.090 | 1.075 | 0.995 | 1.066 | 1.180 | 1.134 | 1.077 |
| WPCSW (r = 2) | 1.016 | 1.056 | 1.015 | 0.954 | 0.967 | 0.876 | 1.030 | 1.134 |
| WPCSW (r = 3) | 1.033 | 1.079 | 1.067 | 0.954 | 0.934 | 0.888 | 1.022 | 1.325 |
| WPCSW (r = 5) | 1.049 | 1.101 | 1.075 | 1.067 | 1.016 | 0.899 | 1.112 | 1.294 |
| TPH (r = 1) | 1.049 | 1.180 | 1.231 | 1.242 | 0.951 | 0.978 | 0.970 | 1.015 |
| TPH (r = 2) | **0.951** | 1.225 | 1.246 | 1.057 | 0.967 | 0.955 | 0.955 | 1.320 |
| TPH (r = 3) | 1.180 | 1.213 | 1.090 | 1.000 | 0.984 | 1.034 | 0.978 | 1.273 |
| TPH (r = 5) | 1.164 | 1.079 | 1.149 | 1.175 | 1.131 | 1.000 | 1.470 | 1.402 |
| TPS (30 variables, r = 1) | 0.984 | 1.079 | 1.037 | **0.918** | 1.213 | 1.292 | 1.231 | 1.093 |
| TPS (30 variables, r = 2) | 0.984 | 1.067 | 1.015 | 0.995 | 1.164 | 1.281 | 1.254 | 1.196 |
| TPS (30 variables, r = 3) | 1.000 | 1.045 | 1.097 | 0.933 | 1.115 | 1.169 | 1.306 | 1.284 |
| TPS (30 variables, r = 5) | 0.967 | 1.101 | 1.142 | 1.031 | 1.148 | 1.236 | 1.239 | 1.314 |
| PLS (k = 1) | 1.049 | 1.236 | 1.119 | 1.196 | 0.951 | 0.854 | 0.940 | 1.098 |
| PLS (k = 2) | 1.295 | 1.618 | 1.328 | 1.479 | 1.049 | 1.045 | 1.299 | 1.356 |
| PLS (k = 3) | 1.492 | 1.596 | 1.478 | 1.536 | **0.902** | 0.787 | 0.970 | 1.351 |
| PLS (k = 5) | 1.770 | 1.966 | 1.784 | 1.701 | 0.934 | 1.034 | 1.194 | 1.273 |
| RR ($v_2$ = 0.25) | 2.951 | 3.382 | 3.396 | 3.227 | 1.016 | 0.989 | 1.194 | 1.335 |
| RR ($v_2$ = N) | 1.148 | 1.247 | 1.104 | 1.093 | **0.902** | **0.820** | 0.948 | 1.165 |
| RR ($v_2$ = 5N) | 1.016 | 1.067 | 1.000 | 0.995 | 0.934 | 0.876 | **0.925** | 1.036 |
| RR ($v_2$ = 10N) | 1.016 | 1.034 | 0.993 | 0.990 | 0.934 | 0.899 | 0.933 | 1.005 |
| EN (30 variables, $v_2$ = 0.25) | 1.934 | 2.112 | 1.746 | 2.325 | 1.049 | 0.933 | 1.299 | 1.057 |

Notes: RMSE(model)/RMSE(univariate). Minima are in bold, and relative RMSEs < 1 are in gray. For the trade-weighted aggregates, we consider the following information: real: GDPs of all 12 countries; prices: CPIs of 12 countries; monetary and financial: interest rates of 5 countries; North America: GDPs and CPIs of the US and Canada and US interest rates; Europe: GDPs, CPIs and interest rates of the euro area and the UK; Asia: GDPs and CPIs of all Asian countries and interest rates of Japan; Australia: Australian GDP, CPI and interest rates. For more details on the forecasting setup, see the text.