2018-14

# Forecasting with Many Predictors: How Useful are National and International Confidence Data?

Kevin Moran
Simplice Aimé Nonoy
Imad Rherrad

Août / August 2018

Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

www.crrep.ca

# Abstract

This paper assesses the contribution of Canadian and International (US) confidence data, drawn from consumer and business sentiment surveys, for forecasting Canadian GDP growth. The targeting approaches of Bai and Ng (2008) and Bai and Ng (2009) are employed to extract promising predictors from large databases each containing between several dozen and several hundred time series. The databases are categorised between those containing macroeconomic (Canadian and US) and confidence (Canadian and US) data, allowing us to assess the specific value added of international and confidence data. We find that forecasting ability is consistently improved by considering information from national confidence data; by contrast, their US counterparts appear to be helpful only when combined with national time-series. Overall, most relevant gains in forecasting performance are observed for short-term (up to threequarters-ahead) horizons, perhaps reflecting the timing advantage in the releases of sentiment data.

**Kevin Moran:** Department of Economics, Université Laval. Email: kevin.moran@ecn.ulaval.ca
**Simplice Aimé Nono:** Department of Economics, Université Laval. Email: simplice-aime.nono.1@ulaval.ca
**Imad Rherrad:** Ministère des Finances du Québec. Email: imad.rherrad@finances.gouv.qc.ca

# 1 Introduction

The use of survey data on business, consumer and investor confidence - or *sentiment*[1] - has become an important tool for policy makers worldwide. Surveys on sentiment are carried out on a timely monthly or quarterly manner to elicit early signals about future economic developments and are not subject to subsequent revisions. By contrast, information on the current state of the economy, while crucial to economic-policy-analysis and forecasting, is released after substantial delays and likely to be revised in future vintages of data releases. This timing advantage and the absence of revision represent key potential edges for sentiment data when forecasting future economic activity (Lahiri and Monokroussos, 2013; Lahiri et al., 2016).

The present paper assesses the extent to which this advantage is present when Canadian as well as International (ie. US) confidence data are employed to forecast Canadian GDP growth. To provide this assessment, we analyze the forecasting ability of four databases, where the first two contain standard macroeconomic and NIPA data from Canada and the US, respectively, while the latter two incorporate sentiment data from each of these two countries. A targeted approach adapted from Bai and Ng (2008) and Bai and Ng (2009) is employed to identify promising predictors for each of the databases: since these contain between several dozens and several hundred variables each, such an approach is necessary to reduce the dimension of the forecasting equation and efficiently use the information present in these data. We then test the forecasting ability of our different databases via a pseudo out-of-sample experiment based on a rolling window and formally compare the resulting forecasts. Finally, we progressively merge the four databases into larger ones arranged by theme (all Canadian data, all confidence data, etc) and repeat our analysis, until we have one large database using the complete set of around 1300 time series.

We find that forecasting ability is consistently improved by considering information from national confidence data; by contrast, their US counterparts have a less robust contribution, which may only appear when these data are combined with other time-series.

[1]In this paper, the terms sentiment and confidence are used interchangeably.

Overall, most gains in forecasting performance are observed for short-term (up to three-quarters-ahead) horizons, perhaps reflecting the time release advantage of sentiment data.

Our results contribute to two distinct research programs. First, our findings about the forecasting contribution from sentiment data extend the body of evidence documenting how such data can improve forecasts for real activity variables such as GDP growth or the likelihood of economic downturns. This literature includes contributions such as Matsusaka and Sbordone (1995), Santero and Westerlund (1996), Bodo et al. (2000), Hansson et al. (2005), or Taylor and McNabb (2007). It has recently adopted factor models as the benchmark of analysis, as in Chen et al. (2011), Lahiri and Monokroussos (2013), Christiansen et al. (2014), Martinsen et al. (2014), Lahiri et al. (2016) and Moran et al. (2018). Evidence specifically related to the Canadian case is more scarce, despite the existence of at least four surveys on sentiment in Canada. The forecasting frameworks presented in Binette and Chang (2013), Ferrara et al. (2015) and Chernis and Sekkel (2017) do include some confidence data but does not identify its specific contribution, while recent work by Pichette and Robitaille (2017) focuses solely on the *Business Outlook Survey* data produced by the Bank of Canada. The present paper argues that all available Canadian data on confidence should be used in conjunction with targeting approaches designed to use these data in the most efficient manner possible.

Second, our paper also adds to a literature that assesses the extent to which international variables are important to provide accurate forecasts for national variables of interest: Cheung and Demers (2007), Schumacher (2010), Eickmeier and Ng (2011) or Kopoin et al. (2013) are representative contributions to this literature. Interestingly, this research program has often reported that while international variables may help better forecast, this advantage is not present for all cases and at all forecasting horizons; sometimes, the efficient use of national variables might suffice to produce good forecasts (Kopoin et al., 2013). Such a finding accords well with those, documented in the present paper, whereby US confidence data might not be needed to provide the best possible forecasts for Canadian economic activity.

Throughout the paper, we apply targeting methods designed to efficiently extract information contained in large databases. These methods reflect the fact, documented in Boivin and Ng (2006), that applying factor modeling to larger databases does not invariably lead to better forecasting equations, but that instead identifying variables or factors likely to contain good information in advance of the factor extracting and forecasting exercise may produce superior results (Bai and Ng, 2008, 2009). More generally, the present paper contributes to the general literature assessing how data-rich datasets can be efficiently used to improve forecasts of a variety of economic variables. This literature originates from the seminal contributions of Stock and Watson (2002a,b) and Forni et al. (2005) and has become a standard part of the macroeconomic forecaster's toolkit; see Jurado et al. (2015) or Kotchoni and Leroux (2018), among many others, for recent applications. Moreover, its use is now facilitated by the emergence of publicly available and continuously updated large datasets (McCracken and Ng, 2016b; Fortin-Gagnon et al., 2018)

The rest of the paper is organized as follows. Section 2 describes our forecasting model and targeting approaches. Section 3 describes the data used and the rich variety of sentiment data we employ. Section 4 presents the empirical analysis. Section 5 reports our results, while Section 6 concludes by offering some suggestions for future research.

# 2 Framework

This section describes our econometric strategy. We first revisit the factor models employed in key papers of the forecasting literature. Next, we describe the principal component technique we employ to estimate these factors, and the variable and factor selection processes used to identify promising predictors. Finally we discuss the forecasting performance measure we use to compare altternative datasets.

## 2.1 Forecasting Models

Our goal is to forecast the time series $\{y_{\tau+h}\}_{\tau=T-h+1}^{T^*}$ conditional on $\mathcal{F}_\tau$, the information set available at time $\tau$, which contains the previous values of $y$ up to $\tau$ and a large number of potential predictors observed as of time $\tau$. The first step is to establish a benchmark model for comparison purposes, wherein the evolution of real economic activity only depend on its past realizations through an autoregressive process of order $p$:

$$y_{t+h} = \alpha^h + \sum_{d=1}^{p} \lambda_d^h y_{t-d} + \epsilon_{t+h}, \tag{1}$$

where $y_t$ is a measure of economic activity (here the growth rate of real GDP), $d$ the number of lags used, $t \leq T - h + 1$ the length of the estimation sample, $h$ the forecasting horizon and $\epsilon_t$ an $i.i.d$ $N(0, \sigma_\epsilon)$ error term. Given the information set, the $h$-step-ahead forecast of $y_\tau$ with $T - h + 1 \leq \tau \leq T^*$ is derived as

$$\widehat{y}_{\tau+h/\tau} = \widehat{\alpha}^h + \sum_{d=1}^{p} \widehat{\lambda}_d^h y_{\tau-d+1}.$$

where dynamic forecasts (ie. $\widehat{y}_{t+d+1}, \cdots, \widehat{y}_{t+h-1}$ serves to forecast $\widehat{y}_{t+h/\tau}$) are employed for the cases where $h > d$.

Next, suppose that $\mathbf{X}_t = (X_{1t}, X_{2t}, ..., X_{Nt}) = [X_{it}]_{i=1,...,N;t=1,...,T}$ represents a N-dimensional vector of time series of potential predictors for $y_{t+h}$. An extension of the baseline model is to account for current and lagged values of these variables as in:

$$\begin{cases} y_{t+h} = \alpha^h + \sum_{d=1}^{p} \lambda_d^h y_{t-d+1} + \sum_{d=1}^{q} \beta_{\mathbf{d}}^{\mathbf{h}} \mathbf{X_{t-d+1}} + \epsilon_{t+h}, \\ \widehat{y}_{\tau+h/\tau} = \widehat{\alpha}^h + \sum_{d=1}^{p} \widehat{\lambda}_d^h y_{\tau-d+1} + \sum_{d=1}^{q} \widehat{\beta}_{\mathbf{d}}^{\mathbf{h}} \mathbf{X}_{\tau-\mathbf{d}+\mathbf{1}}. \end{cases} \tag{2}$$

In a data-rich environment, several dozens time series may be available as potential predictors to include in the vector $\mathbf{X}_t$. In such an environment, the use of factor models can help reduce the dimension of the problem. These models combine the information content of many different variables into a few representative factors, which are then employed to forecast the variable of interest. Towards that goal, we thus assume that each potential predictor $X_{it}$ in (2) is represented by the factor structure

$$X_{it} = \alpha_i^{'} F_t + e_{it}, \quad i = 1, ..., N \quad , \quad t = 1, ..., T, \tag{3}$$

where $F_t$ is a $r \times 1$ vector of factors common to all $X_{it}$, $\alpha_i$ a $r \times 1$ vector of factor loadings collecting the influence of each factor on $X_{it}$, and $e_{it} \sim i.i.d.$ $N(0, 1)$ is an idiosyncratic component. The $r$ common factors ($r << N$) can be estimated via principal component decomposition on the normalized data vector $\mathbf{X}_t$ (Stock and Watson, 2006) and estimated factors $\widehat{F}_t$ can then be used in forecasting $y_{t+h}$. We then have

$$\begin{cases} y_{t+h} = \alpha^h + \sum_{d=1}^{p} \lambda_d^h y_{t-d} + \sum_{d=1}^{q} \gamma_d^h \widehat{F}_{\tau-d} + \epsilon_{t+h}, \\ \widehat{y}_{\tau+h/\tau} = \widehat{\alpha}^h + \sum_{d=1}^{p} \widehat{\lambda}_d^h y_{\tau-d} + \sum_{d=1}^{q} \widehat{\gamma}_d^h \widehat{F}_{\tau-d}, \end{cases} \tag{4}$$

where $\widehat{\alpha}$, $\widehat{\lambda}$, $\widehat{\gamma}$ are estimated coefficients conditional on the forecasting horizon $h$ and the estimated factors $\widehat{F}_\tau$.Various contributions have substantiated the usefulness of this strategy in forecasting and it has become standard in the literature. See for example Stock and Watson (2002a,b), Forni et al. (2005) and Bai and Ng (2002, 2006, 2008) for the building blocks of this literature, as well as Stock and Watson (2006) for on overview.

The seminal contributions of this literature originally considered every available variable $X_{it}$ as relevant when deriving the common factors used in the forecasting stage. However, Boivin and Ng (2006) show that additional variables may be noisy, less-informative or redundant, and therefore might not always be useful for deriving the factors; in fact including more variables may lead to decreases in model performance. Accordingly, Bai and Ng (2008) propose several methods designed to *preselect* promising, relevant variables before conducting the factor extracting process. In addition, Bai and Ng (2009) suggests that preselecting relevant *factors* as well before going on the forecasting stage may also be a valuable strategy. As discussed below, the present paper employs both strategies to boost the efficiency of our forecasting framework.

## 2.2 Factor Estimation

Factor model analysis posits that a small number of orthogonal variables –the factors– can explain a large proportion of the variability in one dataset; in practice estimating these factors is often accomplished via principal component analysis (PCA), following results in Stock and Watson (2006).

We denote our PCA decomposition following Johnson and Wichern (2007), Chapter 8:

let $\mathbf{X} = (X_1, ..., X_N)'$ be a vector of $N$ random variables, with the covariance matrix

$$\Sigma = var(\mathbf{X}) = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN}^2 \end{bmatrix}.$$

Next, form the following linear combinations of the columns of $\mathbf{X}$:

$$\begin{aligned} &F_1 = \alpha_{11}X_1 + \cdots + \alpha_{1N}X_N = \alpha_1^{'}\mathbf{X} \\ &\vdots \\ &F_r = \alpha_{rN}X_1 + \cdots + \alpha_{rN}X_N = \alpha_r^{'}\mathbf{X} \end{aligned}$$

where $\alpha_i$ is the coefficient of the regression of $F_i$ on $\mathbf{X}$ ($F_i$ is a random latent variable given that $X_i$ are random exogenous variables) and the covariance matrix for the factor is

$$\begin{aligned} Var(F_i) &= \textstyle\sum_{k=1}^{r}\sum_{l=1}^{r}\alpha_{ik}\alpha_{il}\sigma_{kl} = \alpha_i^{'}\Sigma\alpha_i, \\ Cov(F_i, F_j) &= \textstyle\sum_{k=1}^{r}\sum_{l=1}^{r}\alpha_{ik}\alpha_{jl}\sigma_{kl} = \alpha_i^{'}\Sigma\alpha_j. \end{aligned}$$

The PCA algorithm is as follows. The first principal component of $\mathbf{X}$ is computed as $F_1 = \alpha_1^{'}\mathbf{X}$ as the result of the following maximization problem

$$\underset{\alpha_1}{\mathbf{Max}}\, Var(F_1) \qquad s.t. \qquad \alpha_1^{'}\alpha_1 = \sum_{j=1}^{r}\alpha_{1j}^2 = 1, \tag{5}$$

where the constraint $\alpha_1^{'}\alpha_1 = 1$ ensures a unique solution exists.

The second PC is obtained as $F_2 = \alpha_2^{'}\mathbf{X}$ as the result of

$$\underset{\alpha_2}{\mathbf{Max}}\, Var(F_2) \qquad s.t. \qquad \begin{cases} \alpha_2^{'}\alpha_2 = 1 \\ Cov(F_1, F_2) = \alpha_1^{'}\Sigma\alpha_2 = 0 \end{cases}$$

where the constraint $Cov(F_1, F_2) = 0$ ensures there exists no correlation between the first two PCs. Continuing, for the $r^{th}$ PC we have $F_r = \alpha_r^{'}\mathbf{X}$ with

$$\underset{\alpha_r}{\mathbf{Max}}\, Var(F_r) \qquad s.t. \qquad \begin{cases} \alpha_r^{'}\alpha_r = 1 \\ Cov(F_1, F_r) = 0 \\ \vdots \\ Cov(F_{r-1}, F_r) = 0 \end{cases}$$

As a result, the extracted PCs provide a reflection of the common aspects in the complete set of confidence variables (in the survey data) or the macroeconomic variables but arrange this variation along orthogonal axis directions.

### 2.3 Factor and Predictor targeting

The two key practical issues of our analysis are as follows: first, determining which variables, and how many lags of them, to include in order to proceed with the factor estimation stage; second, which factors, and how many lags of them, to include in the forecast construction stage. Lets us analyze these two issues in reverse order.

In order to identify the common factors most useful for the forecasting process, we need to answer to the following question: which factor and which lags of these factors have predictive powers for the economic variable of interest? There is no a-priori reason for the first principal component to deliver a better forecast for $y_{t+h}$. We therefore need a procedure by which common factors are ordered according to their importance in the forecasting process of the variable of interest. To this end, suppose that we have obtained a set of $s$ factors ranked in decreasing order of importance as information content from the dataset $\mathbf{X}_t$: $\widehat{F}_1, ..., \widehat{F}_s$. A standard model selection may choose the best ones according to criteria such as AIC or BIC in a ranking of all the possible combinations when successively adding the factors one after another. However, this approach might miss a factor that is a better predictor of the specific variable of interest but has less importance in the overall factor ranking. To avoid this pitfall, we instead use a hard-thresholding method on the common factors, as in Bai and Ng (2009). This method assesses whether a candidate estimated factor shows good forecasting power before it is selected, by way of the single-predictor regression model

$$y_{t+h} = \alpha^h + \beta^h_{id}\widehat{F}_{i,t-d+1} + \epsilon_{t+h}, \qquad i = 1, ..., s \ , \qquad d = 1, ..., q$$

and keeping $\widehat{F}_{i,t-d+1}$ if $\widehat{\beta}^h_{id}$ is greater than a previously set threshold.

Note that the strategy considered so far treats all time series in $\mathbf{X}_t$ as equally promising in terms of forecasting, and thus includes them all when extracting the factors. However, as Boivin and Ng (2006) point out, it may occur that valuable information about the target variable to forecast is included in timeseries less relevant for explaining the overall variability in $\mathbf{X}_t$ and thus less important for factor estimation. In response, Bai and Ng (2008) propose selection methods whereby only variables with good potential as predictors are included in the factor estimation stage. One such method, hard-thresholding, investigates the predictive power of each variable individually and chooses those to keep via the single-predictor regression model

$$y_{t+h} = \alpha^h + \sum_{d=1}^{p} \lambda_d^h y_{t-d+1} + \sum_{d=1}^{q} \beta_{id}^h X_{i,t-d+1} + \epsilon_{t+h}, \qquad i = 1, ..., N\ , \qquad d = 1, ..., q$$

and keeping $X_{i,t}$ (which now denotes one single variable) for the factor extraction stage only if $\widehat{\beta}_{id}^h$ is statistically significant at some given threshold.

Soft-thresholding is another possible targeting method. It analyses all potential predictors simultaneously, within a multiple-predictor regression, to reduce the possibility that the predictors selected by hard-thresholding contain essentially the same information. The soft-thresholding approach thus considers

$$y_{t+h} = \alpha^h + \sum_{d=1}^{p} \lambda_d^h y_{t-d+1} + \sum_{i=1}^{N} \sum_{d=1}^{q} \beta_{id}^h \mathbf{X}_{i,t-d+1} + \epsilon_{t+h},$$

ie. includes all available potential predictors and then estimates $\beta^h$ as in Zou and Hastie (2005), by solving

$$min_{\beta^h} \left[ RSS + \kappa_1 \sum_{i,d} |\beta_{id}^h| + \kappa_2 \sum_{i,d} {\beta_{id}^h}^2 \right]$$

where $\kappa_1$ and $\kappa_2$ are parameters to be specified by the user. The calibration of the two parameters allows to shrink the number of coefficients to be estimated and to determine the regressors to consider in the factor derivation process.

## 2.4 Forecast Evaluation Measures

To evaluate the forecasting performance of each model, we first compute the mean squared forecast error (MSFE)

$$MSFE = \frac{1}{P}\sum_{t=T-h+1}^{T^*}\left(y_{t+h} - \widehat{y}_{t+h/T}\right)^2 \qquad and \quad T^* = T + P,$$

where $P$ is the number of forecasts. As our goal is to compare model performance across alternatives, we then compute the relative MSFE ($MSFE_{relative}$ henceforth) with respect to the benchmark, as in

$$MSFE_{relative} = \frac{MSFE_i}{MSFE_0},$$

where $MSFE_i$ is for the assessed model and $MSFE_0$ arises from the benchmark. Finally, to determine whether the predictive power of two models are statistically different, we consider the predictive accuracy test introduced by Diebold and Mariano (1995), but use the generalized version of this test proposed by Giacomini and White (2006) (GW henceforth); these tests are designed to compare a model $i$ to a benchmark, with the null hypothesis of equal performance being denoted as

$$H_0 : E\left(d_{i,t+h}\right) = 0, \qquad for \quad t = T - h + 1, \cdots, T*$$

where $d_{i,t+h} = g_t(e_{i,t+h}) - g_t(e_{0,t+h})$ is the differential loss between the model $i$ and the benchmark and $g_t(.)$ a general loss function defined on the forecasts $e_{i,t+h}$. In the context of a quadratic loss function, the GW statistic is as follows:

$$GW = \left(\frac{P + 1 - 2h + P^{-1}h(h-1)}{P}\right)^{1/2} \widehat{V}(\bar{d}_i)^{-1/2}\bar{d}_i, \tag{6}$$

where

$$d_{i,t} = \mathcal{L}_{it} - \mathcal{L}_{0t} = (y_t - \widehat{y}_{it})^2 - (y_t - \widehat{y}_{0t})^2,$$

$$\bar{d}_i = \frac{1}{P}\sum_{t=T+1}^{T^*} d_{i,t},$$

$h$ is the forecast horizon and $\widehat{V}(\bar{d}_i)$ the Newey-West estimated long-run variance of the series $d_{it}$. The GW is then compared to critical values from the t-student distribution with (P-1) degrees of freedom and is rejected if its value is outside the critical region. The advantage of the Diebold-Mariano test is its flexibility with regards to features of forecast errors such as non-zero means, non-normality or contemporaneous correlation. The Giacomini-White variant presents these properties but is developed under more general assumptions and estimations methods and corrects for small sample biases.

# 3 Data

We consider macroeconomic and financial time series, as well as confidence survey data at the national (Canada) and international (US) levels. At the national level, we make use of confidence survey data from the Bank of Canada and the Conference Board of Canada; for their part, macroeconomic and financial data originate from Statistics Canada. At the international (US) level, we use confidence survey data from the Institute of Supply Management (ISM) and the university of Michigan; macroeconomic and financial data are taken from FRED-MD, the database organized by McCracken and Ng (2016b). We consider a quarterly frequency for all the variables and the data span the period from 2002Q1 to 2015Q4.As is standard in the literature, data are pretreated following a three-step procedure: an adjustment for seasonality by performing a linear approximation to $X$-11, a screening for outliers by recording them as missing data and a test for the integration order using Augmented Dickey-Fuller, Phillips Perron and KPSS tests. Conditional on results from these tests, variables are thus subject to one of six possible transformations : No transformation, first-difference in level, logarithm, first-difference of logarithms second-difference of logarithms or difference in rates. After these transformations and a screening for outliers, all variables are then standardized to a zero-mean and unit-variance.[2] The complete set of data are then arranged in four separate subsets: Canadian macroeconomic and financial data ($CA$), Canadian confidence data ($CAc$), US macroeconomic and financial data ($US$) and US confidence data ($USc$).

[2] A list of all the variables with the applied transformations is available on request.

### 3.1 Confidence Variables

Throughout, we use the complete dataset for confidence data, including all sub-components and raw survey data with all answer, instead of relying on published consumer or business confidences indexes. Using all disaggregated confidence data allows us to account for all available information, some of which might have been lost by the construction of these indexes. As suggested in Curtin (2003) and congruent with results in Moran et al. (2018), keeping all raw sentiment data and aggregating it via the targeting and factor procedure might prove more useful for forecasting then relying on the very specific aggregating method used to produce the indices.

The Canadian confidence dataset is thus a panel of 88 time series that contain the raw data for all questions and all answering options to four confidence surveys managed by the Conference Board of Canada and the Bank of Canada. The Conference Board manages two of these surveys: the consumer and business confidence surveys. The Consumer Confidence survey was established in 1979 to sample 2000 Canadian households at a monthly frequency and query their views about their current and future economic conditions. The survey contains four questions and two response options (good, bad): respondents are asked to give their views about their current and expected financial positions, their short-term employment outlook and their opinions about choosing the current moment for a major purchase such as houses or cars. The data therefore define 8 variables, in terms of percentage of good or bad responses for each question.[3]

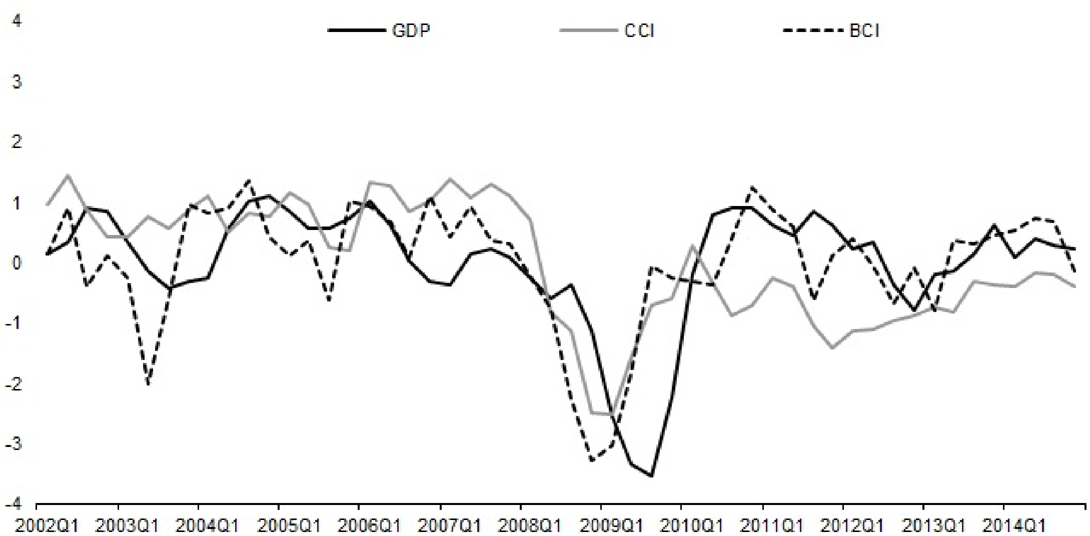
The second survey managed by the Conference Board is the quarterly Business Confidence survey, which exist since 1977. It interviews the Chief Executive Officers and Chief Financial Officers of major Canadian business organizations, in order to measure perceptions about current and future economic conditions and the investment intentions of businesses countrywide. The survey consists of ten descriptive questions, with eight of them focusing on the economic environment, financial conditions, inflation, profit, future

[3] Appendix A describes in details the questions and possible answers for the four surveys. Note that we transform the data to a quarterly frequency by simple averages.

investments, production level, capacity utilization, geographical expansion perspectives, employment level and speed of supplier deliveries. Questions may have multiple response options and data are provided in terms of the percentages of respondents answering each option, giving a total of 58 variables as potential predictors. These raw data are aggregated by the Conference Board of Canada to compute and publish at a quarterly frequency the *Business Confidence Index* ($BCI_t$).

As indicated above, the Conference Board of Canada constructs and publishes *Confidence indexes*, ie simple averages of the balance of opinions in the consumer and business confidence surveys. As such these represent very specific manners in which to aggregate the information contained in those surveys and the present papers instead argues that our PCA approach a more general and possibly more efficient aggregating method. Nevertheless, it is interesting to analyse the joint evolution of Canadian GDP growth and these two indexes.

Figure 1 – Confidence data and GDP growth: Conference Board Data




To this end, Figure 1 above illustrates the evolution of the Conference Board of Canada's confidence indexes and GDP growth for the sample of interest.[4] The two in-

[4]GDP growth is computed as $y_t = log\,(Y_t/Y_{t-4}))$, where $Y_t$ is quarterly Canadian real GDP.

dexes appear to have leading indicator properties. Since 2002, the indexes have decreased before each slowdown period, most notably in the months preceding the Great Recession of 2008-2009, and appear to have gone again nearly two quarters before growth resumed. To the best of our knowledge, the forecasting ability of the confidence data from the Conference Board of Canada has never been analysed systematically. As argued above, We assess this forecasting ability f these data through our general approach using all underlying confidence data instead of the aggregate indexes.

The Bank of Canada also conducts two confidence surveys: the Business Outlook Survey and the Senior Loan Officer Survey. The Bank of Canada began the Business Outlook Survey in 1997. The survey is conducted at a quarterly frequency by the Bank's regional offices and gathers information from firms about their sentiment on business developments and economic conditions. The senior management of 100 firms is interviewed on selected topics: survey respondents are asked to answer eleven attitudinal questions on topics regarding past and future sales growth, investment in machinery, ability to meet the demand, labour shortages, intensity of labour shortages, input and output price inflation, inflation expectations and credit conditions.[5] Most questions have three response options (greater, the same, lower) but the question about the expected inflation rate asks participants' views in more quantitative terms: (four options: above 3%, between 2 and 3%, between 1 and 2%, less than 1%). The responses to nine of the questions are available in the form of balance of opinions (the difference in percentages between the opposite options of each question), except for the responses on labour shortages (one variable), ability to meet the demand (arranged in two variables) and inflation expectations (arranged in four variables which are in percentages). The data represent all together a set of 15 variables.

Finally, the Senior Loan Officer (SLO) survey is a quarterly survey of the business-lending practices of the 11 major Canadian financial institutions and is managed by the Bank of Canada since 1999.[6] Survey participants are asked to provide their informed

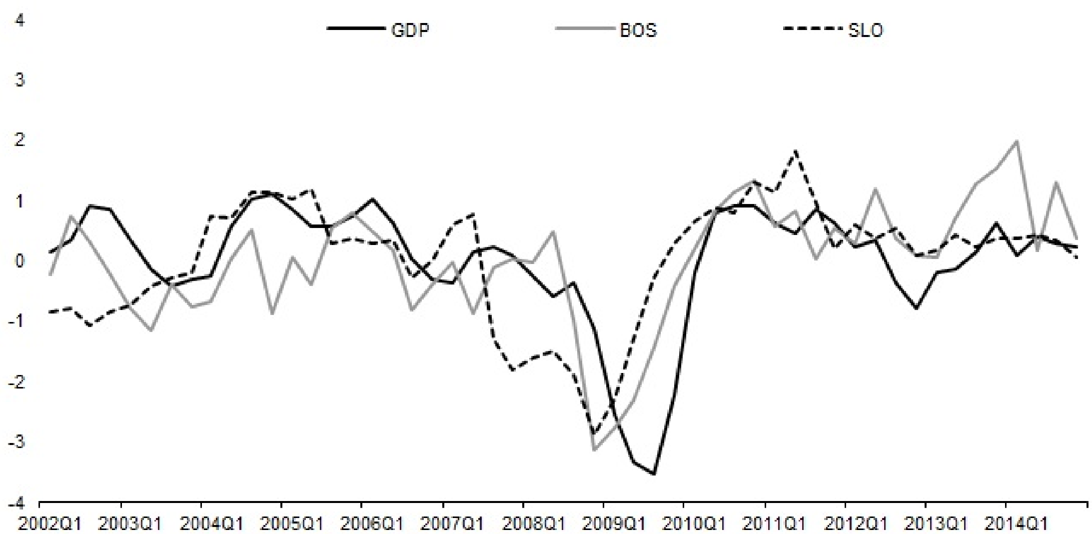
[5] The survey is described further in Martin (2004). All questions and answering options are listed in the Appendix.

[6] See Faruqui et al. (2008) for more details on this survey.

opinion on changes in both the price and non-price terms of business lending over the current quarter. Moreover, the respondents are surveyed about their views on how changing economic or financial conditions are affecting business lending. The survey is conducted during the two-weeks-period before the end of each quarter and summarized in percentages of surveyed financial institutions reporting tightened credit conditions and those reporting eased credit conditions. Data are provided in terms of balance of opinion of the respondents.

Once again, Figure 2 illustrates the relationship between Canadian GDP and simple-sum aggregates of the two surveys managed by the Bank of Canada, to obtain intuition about the likely forecasting ability of the data underlying these survey (the figure depicts negatives value for the index associated to the SLO, which fits better with the GDP growth dynamics. Again, the two indexes seem to display leading indicator properties, as the relevant changes in the GDP growth follow the ups and downs in the indicators

Figure 2 – Confidence data and GDP growth: Bank of Canada Data

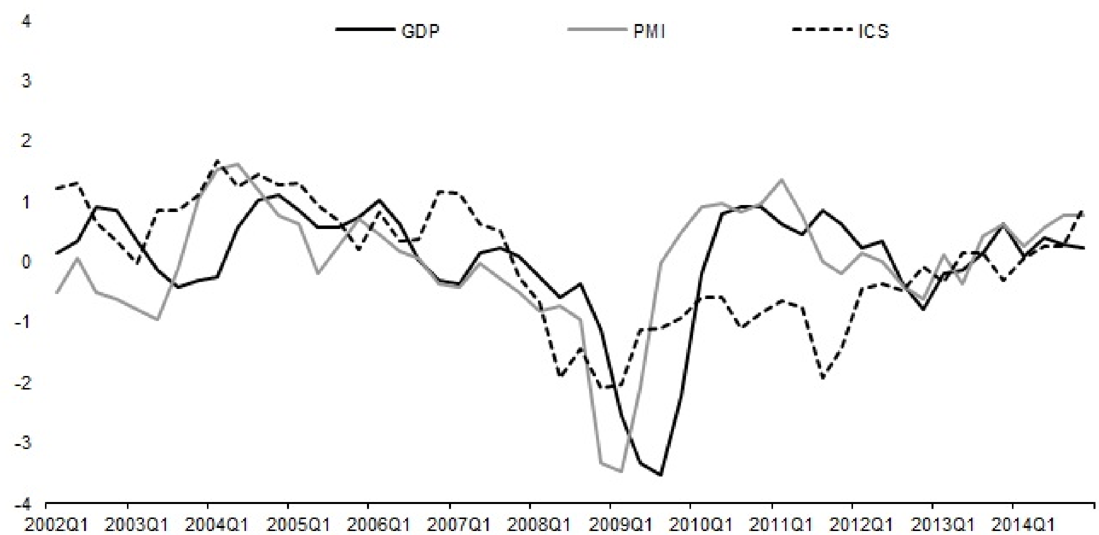



Let us now turn to US confidence data. Our dataset is a panel of 299 time series containing data from two confidence surveys managed by the University of Michigan and the Institute of Supply Management. First, the University of Michigan has been running the consumer confidence survey since the 1940s. Nowadays, the survey is conducted on a monthly basis but we use an (averaged) quarterly frequency. Each month, at least 500 US households are interviewed to gather information about their own current and expected financial situations; broader economic conditions in terms of unemployment and inflation; and buying and saving conditions. Specifically, Survey respondents answer 50 attitudinal questions pertaining to selected topics on current and expected developments of household-specific or country-wide economic items such as income, wealth, prices, interest rates. Each question has multiple response options and data are provided in terms of percentages for each response option, giving a total of 289 time series.[7] These data are used to compute a monthly *Index of Consumer Sentiment* ($ICS_t$) and its *Index of Consumer Expectation* ($ICE_t$), simple averages of answers to five selected questions. Finally, the Institute of Supply Management (ISM) is the oldest organization of its kind to conduct surveys of business confidence and confidence indexes. The data collected at a monthly frequency involve a sample of 400 US industrial companies. The respondents answer questions that compare their current level of activity with that of the previous month in order to measure their perception of current economic developments. The ten survey questions focus on production level, new orders, inventories, prices, employment level and speed of supplier deliveries. Each question has multiple response options and results are provided in terms of balance of opinions for each question and arranged in ten variables. The business confidence survey data are used by the ISM to provide another simple average of these data, the $PMI$ Index.

Figure 3 illustrates the joint evolution of the indexes from the US confidence surveys and GDP growth in Canada. At first look, the $ICS$ and $PMI$ appear to display leading indicator properties for Canadian GDP. In particular, $PMI$ may be seen as presenting stronger leading indicator properties for Canadian GDP whereas such properties appear less substantial for the $ICS$. Again, our strategy is to use the underlying raw data on sentiment rather than these published simple-sum averages, to allow these data to be used in the most efficient manner possible.

[7] Details about the time series and the survey questions are in the codebook available at *https://data.sca.isr.umich.edu/*.

Figure 3 – US confidence data and Canadian GDP growth




## 3.2 Macroeconomic and Financial variables

Our framework allows to compare how sentiment versus macroeconomic data help forecasting Canadian GDP growth either as substitutes or complements. For macroeconomic and financial variables, we have two datasets containing 764 and 128 time series for Canada and the US, respectively. The panels reflect the following broadly-defined categories of economic data: National income and product accounts, industrial production, investment and consumption, employment and unemployment, housing, inventories, orders and sales, prices, hours worked and earnings, interest rates, money and credit, exchange rates, stock markets.[8]

[8]Further details about these data are available in Cheung and Demers (2007) and Moran et al. (2018), for the Canadian database. See also Fortin-Gagnon et al. (2018) for a large database of Canadian macroeconomic and financial timeseries. The US data are from the well-known database from McCracken and Ng (2016b).

# 4 Empirical Analysis

## 4.1 Targeting procedure

We use the four datasets (*CA, CAc, US, USc*) and natural mergers of them (All Canadian data, All Confidence data, All data). These amalgamations provide qualitative selection criteria with which one can identify promising variables for forecasting. Table 1 shows how many candidate variables are in each subset. Next, we apply the Bai and Ng (2008) and Bai and Ng (2009) targeting methods to identify promising individual variables and factors in each dataset. Specifically, the targeting procedure consists of the following steps:

1. Following Bai and Ng (2008), we apply a hard thresholding method to select the predictors likely to be useful in the factor derivation process of our analysis. This method consists of preselecting the variables that shall remain in the pool from which factors are extracted by looking at their forecasting performance within a set of regressions run with only one predictor at time. The predictive power of each individual variable is evaluated by comparing the Student t-statistic with given thresholds $t^* = 0, 1.28, 1.65, 2.58,$, corresponding to no threshold, 10%, 5% and 1% critical values, respectively, in two-tailed t-tests.

2. As in Bai and Ng (2009) we compute the Principal Component decomposition of the matrix of selected variables in the pool and so derive the common factors. We then apply a hard thresholding method to select promising factors as indicated above for the individual variables.

3. In the spirit of Bai and Ng (2008), we experiment with the alternative soft-thresholding method to preselect useful predictors. This works by applying the least-angle regression with elastic net (LARS-EN) in equation (2), which accounts for all the predictors at the same time. This procedure requires to calibrate two main parameters $\kappa_1$ and $\kappa_2$. The first one is set by choosing the maximum number of variables for the factor derivation process $N_{\mathcal{A}}$ (we consider $N_{\mathcal{A}} = 15, 30, 75$) and the second one is set to 0.25, following Bai and Ng (2008).

Table 1 – Different subsets of national and international data

| | | | | |
|---|---|---|---|---|
| **Original Datasets** | CA | CAc | US | USc |
| **Number of series** | 764 | 88 | 128 | 299 |
| | | | | |
| **Amalgamated Datasets I** | - | CA+CAc | CA+US | CA+USc |
| **Number of series** | - | 852 | 892 | 1063 |
| | | | | |
| **Amalgamated Datasets II** | - | CA+CAc+US | CA+CAc+USc | CA+US+USc |
| **Number of series** | - | 980 | 1151 | 1191 |
| | | | | |
| **Amalgamated Datasets III** | - | - | - | CA+CAc+US+USc |
| **Number of series** | - | - | - | 1279 |

Notes: This table describes the number of series in each dataset used. National data include macroeconomic and confidence variables from Canada whereas international data consists of US macroeconomic and confidence indicators. $CA \equiv$ Canada economic data, $CAc \equiv$ Canada confidence data, $US \equiv$ US economic data, $USc \equiv$ US confidence data, $CA + CAc \equiv$ All Canadian data, $CA + US \equiv$ All macroeconomic data, $CA + USc \equiv$ Canadian macroeconomic data with US confidence data, $CA + CAc + US \equiv$ All Canadian data with US macroeconomic data, $CA + CAc + USc \equiv$ All Canadian data with US confidence data, etc.

## 4.2 Forecasting procedure

We aim to determine whether national and international confidence data contain additional information content relevant for forecasting national GDP over and above that already present in the macroeconomic and financial data. Towards this goal, we construct forecasts for the cumulative growth rate of Canadian GDP $y^h_{t+h} = log(GDP_{t+h}/GDP_t) = \sum_{i=1}^{h} \Delta log(GDP_{t+i})$. between periods $t$ and $t + h$, in the spirit of similar exercises in Cheung and Demers (2007), Schumacher (2010), Kopoin et al. (2013), McCracken and Ng (2016a) or Fortin-Gagnon et al. (2018). We derive the forecasts from one-quarter-ahead to eight-quarter-ahead ($h = 1, 2, \cdots, 8$). All the forecasts are based on a direct linear projections as specified in equation (4), using a rolling window.

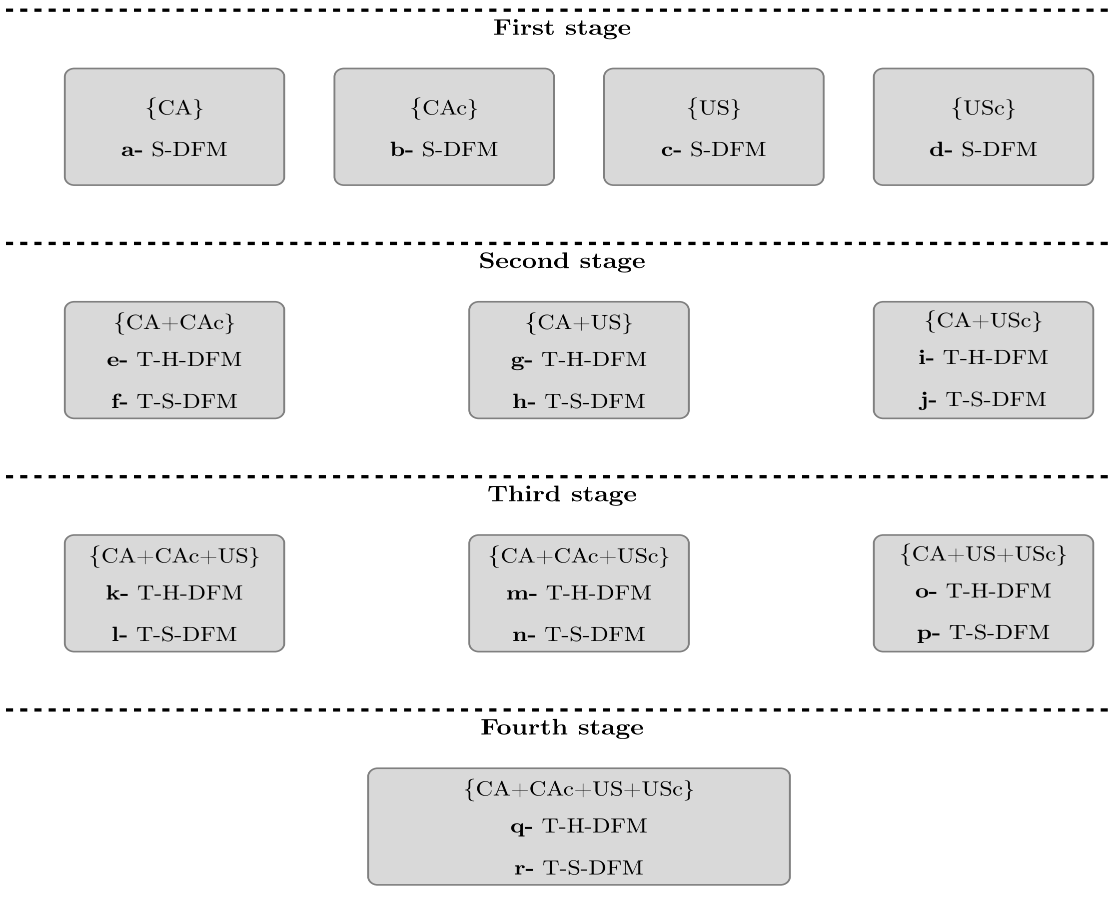
We thus estimate the models using data from 2002Q1 through 2010Q1, and use these to produce forecasts one to eight quarters ahead, ie for 2010Q2 to 2012Q1. This initial estimation and forecasts determine the width of the moving window $[2002Q1 - 2010Q1]$ for our rolling forecasts. Next, the window is therefore moved ahead one time period, that is $[2002Q2 - 2010Q2]$, and we re-estimate the models to produce another set of forecasts, for 2010Q3 through 2012Q2. the window, the estimates and the forecasts are updated in this manner until the end of the sample, at which point we have time series for one- to eight-quarter-ahead forecasts from 2010Q2 to 2015Q4. Table 2 summarizes the experiment.

Table 2 – The Forecasting Experiment (2010Q1 - 2015Q4)

| Estimate | Forecast $h$ periods ahead | | | | |
|---|---|---|---|---|---|
| | $h = 1$ | $h = 2$ | $h = 3$ | $\cdots$ | $h = 8$ |
| $2002Q1 \longrightarrow 2010Q1$ | $2010Q2$ | $2010Q3$ | $2010Q4$ | $\cdots$ | $2012Q1$ |
| $2002Q2 \longrightarrow 2010Q2$ | $2010Q3$ | $2010Q4$ | $2011Q1$ | $\cdots$ | $2012Q2$ |
| $2002Q3 \longrightarrow 2010Q3$ | $2010Q4$ | $2011Q1$ | $2011Q2$ | $\cdots$ | $2012Q3$ |
| $2002Q4 \longrightarrow 2010Q4$ | $2011Q1$ | $2011Q2$ | $2011Q3$ | $\cdots$ | $2012Q4$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $2005Q2 \longrightarrow 2013Q3$ | $2013Q4$ | $2014Q1$ | $2014Q2$ | $\cdots$ | $2015Q3$ |
| $2005Q3 \longrightarrow 2013Q4$ | $2014Q1$ | $2014Q2$ | $2014Q3$ | $\cdots$ | $2015Q4$ |

We first compute forecasts using our benchmarks, denoted by the data subset used to produce them: the simple autoregressive model (AR) or the model using only Canadian macroeconomic data (CA). Then, we estimate the targeted factor models for every combination and for each information sets. Given the four initial subsets (CA, CAc, US, USc), we proceed with our analysis over four different stages of amalgamation, eleven subsets of data and eighteen models (from (a) to (r)) estimated. Figure 4 summarizes the stages, the data subsets used and the type of the estimated models. In each comparison, we compute the $MSFE_{relative}$ to evaluate the information gain respect to the benchmark.

Figure 4 – Stages of the forecasting exercise




Notes: $S-DFM$, $T-H-DFM$ and $T-S-DFM$ denote the *Standard*, *Targeted with Hard-threshold* and *Targeted with Soft-threshold* Factor models, respectively. The acronym in the brackets {} refers to the data subset used to perform the forecasting exercise.

## 5 Results

Tables 3 to 7 present our results. The entries recorded in the tables report the $MSFE_{relative}$ for a specific case, derived as the ratio of the mean-squared forecast error obtained with the model considered to the one obtained with each of two benchmarks: the univariate $AR$ benchmark (Panel A of each table) or the factor model using only standard Canadian macroeconomic and financial variables as predictors (labelled *CA*, Panel B of each table). Recall that $MSFE_{relative} \leq 0$ suggests more informative forecasts relative to the benchmark. The Giacomini and White (2006) test is used assess this relative forecasting performance: rejection of the null of equal predictive accuracy is indicated by the symbols *, ** and *** (they represent 10%, 5% and 1% levels of significance, respectively).

## 5.1 Forecasting Performance with no Targeting

Table 3 presents results obtained when no pre-selection is conducted before extracting factors ($t^* = 0$). As indicated above, Panel A of the table report results relative to the *AR* benchmark, while Panel B compares them to the *CA* dataset.

Panel A of the Table reveals that when forecasting at a short-term horizon (a quarter, say) the *CA* dataset, on its own or used in conjunction with others (CAc, US, USc) performs better than the *AR* benchmark in a statistically significant manner. Interestingly, adding Canadian confidence variables (*CA + CAc*, $MSFE_{relative}$ of 0.834 rather than 0.863 for the benchmark) or general US variables (*CA + US*, $MSFE_{relative} = 0.841$) improves on the CA dataset alone but adding only US confidence variables (*CA + USc*) does not. Note as well that using all the four datasets together does not result in a better forecast than that arrived using Canadian variables only (contrasts results with *CA, CAc, US, USc* $MSFE_{relative}$ of 0.920 with those from *CA*). This under-performance is in line with findings in Boivin and Ng (2006) showing that having more data does not invariably lead to better forecasts when using a factor model framework. Finally it is interesting to note that forecasting with factor models may not deliver “transitive" sets of results: for example, the best forecasting performance is obtained by the combination *CA, CAc, USc* even though on its own, the database *USc* did not appear very informative. For longer forecasting horizons, Panel A of Table 3 shows that the advantage of our factor approach fades for forecasting horizons around the yearly mark: there, instances where forecasting performance is superior to the benchmark are much scarcer; interestingly however, the advantage of factor models appears to recover for longer horizons, around the eight-quarter mark.

Panel B of the table report results where the comparison benchmark is $CA$, the dataset with general Canadian macroeconomic variables; this allows a more direct assessment of the value added from using confidence and/or international variables on the performance of the forecasting model. Results from Panel A are confirmed but presented on a different scale: adding Canadian confidence variables, or general US variables, improves the framework's forecasting ability ($MSFE_{relative}$ of 0.966 and 0.975, respectively) while using US confidence variables does not $MSFE_{relative} = 1.141$). The inability of US confidence data to ameliorate, on its own, the performance of our forecasting model may appear puzzling, in light of results reported by Hudson and Green (2015), who show that when both US and UK investor sentiment are used in a regression predicting UK stock returns, US sentiment becomes highly significant with respect to its UK counterpart. Such a result suggests that UK investor sentiment is heavily influenced by that of the US and thus contains no independent information. Here, US sentiment is useful but only when used in conjunction with Canadian sentiment: the $MSFE_{relative}$ declines from 0.966 for $CA + CAc$ to a minimum of 0.941 for $CA + CAc + USc$, suggesting the Canadian and US sentiment contain complementary information that together can bring superior forecasts.

As per forecasting horizons, the general pattern mirrors the one from Panel A: adding variables appears not to help forecasting performance for horizons past the two or three-quarters ahead mark; this is in line with Schumacher (2007, 2010) and Kopoin et al. (2013)'s findings that additional (international) data may improve forecasting ability, but mostly for horizons shorter than one-year-ahead. Interestingly, as was the case in Panel A, results for longer-term horizons (eight quarters or two years ahead) suggest that the relative ability of larger databases with confidence or international variables improves again.

Overall, Table 3 shows that Canadian confidence and US macroeconomic variables improve the model's forecasting ability for Canadian GDP growth, particularly at shorter forecasting horizons, but that US confidence data on their own do not. In addition, the table also shows that although larger databases with more data are not a guarantee of better performance they may deliver this improvements for longer-term forecasting horizons.

### 5.2 Forecasting Performance with Predictor Hard Thresholding

Tables 4, 5 and 6 report our findings for the cases where the factors used in the forecasting equation are identified by first preselecting predictors using $t^* = 1.28, 1.65$ and 2.58 respectively. Recall that this method aims to extract factors from a data pool that includes only timeseries with the proven potential to help predicting the variable of interest.

Table 4 mostly confirms results from Table 3. Indeed, the table shows that when used alone, only the *CA* dataset can outperform the AR benchmark for short-term forecasting horizons; in addition, forecasting with targeting predictors may improve performance, both at the very short and at the longer end of the forecasting horizon (Panel A of the table). Looking at Panel B of the table reveals that when assessed in comparison with the *CA* dataset, Canadian confidence data as well as US macroeconomic data can both significantly improve performance, and that the value-added of US confidence data is at best indirect, when used in conjunction with other types of information. Finally, the table confirms the previous table's result about the consequences of using very large datasets that include all available information: such a strategy does not invariably lead to better performance, as suggested by Boivin and Ng (2006).

Comparing Table 4 and Table 3 also has important implications for our assessment of targeting methods. Table 4 shows that targeting predictors likely to be informative before extracting factors, as advocated by Bai and Ng (2008) and recently confirmed by work in Leroux et al. (2017), is promising for our forecasting framework. To see this compare, for example, results obtained when using the $CA + CAc + USc$ subset across the two tables: its $MSFE_{relative}$ declines from 0.941 in Panel B of Table 3, to 0.930 in Table 4. Such an improvement is also present for the $CA + CAc$ dataset (a decline in $MSFE_{relative}$ from 0.966 to 0.951) and, to a smaller extent, to the *CA + CAc + US + USc* dataset. More generally, the performances of the models with targeted predictors in Table 4 appear to improve over the complete one-quarter to four-quarter-ahead forecasting horizons, relative to what was the case in Table 3. This suggests that targeting variables before extracting factors is an efficient manner to process information in large databases, which then allows the efficient use of the largest datasets in our experiments.

However, studying Table 5 and 6 show that this targeting process can err on the "too strict" side. Recall that Tables 4, 5 and 6 reflect experiments whereby the targeting is increasingly stricter, so that in Table 6 individual variables have to show very significant predictive ability $t^* = 2.58$ on their own to remain in the pool from which factors are extracted. Setting such a high target leads to discard potentially informative variables and thus the resulting factors have a poorer forecasting ability. Overall, the tables show this decline in forecasting ability begins quickly in Table 5 and 6, so that the general forecasting performance peaks for a mild level of targeting ($t^* = 1.28$, Table 4).

### 5.3 Forecasting Performance with Predictor Soft Thresholding

Table 7 contains the results based on Bai and Ng (2008) LARS-EN variable pre-selection. Recall that this targeting method identifies the variables to keep in the pool used to extract factors in a manner that simultaneously assesses the forecasting ability of all other potential variables. Each panel of the table presents the $MSFE_{relative}$ with respect to the CA benchmark for a given maximum number ($N_{\mathcal{A}}$) of variables in the pool of variables used to derive the common factors. The results show strong evidence in favour of the relevance of national and international confidence data. The overall performance in coherent with the main results obtained in the previous subsection: the use of national and international confidence help to improve consistently the forecasting performance and the derived forecasts are most informative and statistically significant at the one-quarter-ahead horizon. Moreover, although a targeting approach with LARS-EN allow to broadly improve the forecasting ability, a closer look at the detailed results let observe that they are noticeably sensitive to the choose of $N_{\mathcal{A}}$. Selecting up to $N_{\mathcal{A}} = 75$ allow the results to vary over the forecast horizon, but for $N_{\mathcal{A}} \geq 75$ there are no other changes in the computed $MSFE_{relative}$, results become stable over the forecast horizon, consistent with the results in Schumacher (2010). The best forecast performance is obtained in panel B, for the combination of datasets $CA + CAc + USc$ when $N_{\mathcal{A}} = 30$. The advantages derived from the use of confidence data (both national and international) are more pronounced with the use of a soft thresholding approach (employing LARS-EN to preselect variables) which allow results in table 7 to outperform those depicted in tables 3 to 6. Results in line with the findings obtained by Bai and Ng (2008) that with a soft thresholding approach, the factor derivation process becomes more efficient.

### 5.4 Forecasting Performance with Factor Targeting

Finally, Table 8 presents results obtained with factor targeting Ãă la Bai and Ng (2009), a selection of factors derived with the use of the full datasets and no pre-selection of variables. Results are broadly similar to those in previous subsections using variable hard- and soft-thresholding. Indeed, accounting for both national and international confidence allows to importantly improve forecasting accuracy. Not surprisingly, forecasting ability does vary with the targeting threshold and results become more stable, according to the Giacomini and White (2006) test, over the horizon when the threshold increases. The best forecasting performance is obtained in Panel B of the table with the target $t^* = 1.65$ in the factor hard-thresholding process. In the this case, the same combination of dataset $CA + CAc + USc$ also displays the best predictive performance. This is in line with the findings in Bai and Ng (2009) as using a factor targeting contributes to boost the predictive ability of every data combination and allows the results to outperform those derived before with a variable hard-threhsolding method.

## 6 Conclusion

Over the last decade, confidence surveys have received increasing attention and diffusion from media and economic researchers alike. This paper provides evidence to support the view that national (Canada) and to a lesser extent international (US) confidence data are usefully contributors to the objective of forecasting Canadian economic activity. More specifically, we use data from various investor, business and consumer surveys in Canada and the US and investigate the marginal contribution they can have within large datasets that include several dozen macroeconomic variables. We find that Canadian confidence data contain especially valuable additional information, over and above information contained in benchmark macroeconomic variables. Our methodology employs the Bai and Ng (2008) and Bai and Ng (2009) targeting methods whereby variables are preselected before entering the pool of variables from which the factors used in the forecasting equation will be extracted. Doing so, we direct our attention to individual, non aggregated confidence data, instead of using the available, aggregated confidence indices

Our findings reveal that confidence variables (both national and international) possess relevant predictive ability for future Canadian economic activity. For some forecasting horizons, the information contained in confidence allows it outperform any combination of datasets that abstract from confidence variables. The best forecasting performance is obtained by combining Canadian and US confidence with Canadian macroeconomic and financial variables in a large dataset to derive factors. The overall results suggest that disaggregated confidence survey data on investor, business and consumer both at national and international level are informative in the forecasting of national real economic activity.

A promising avenue for future research would be to evaluate the potential contribution of confidence data when producing distribution rather than mean forecasts; recent research has highlighted the interest of using density forecast to measure uncertainty around future macroeconomic outcomes and confidence data can potentially make this literature progress.

## References


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

Jushan Bai and Serena Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150, 2006.

Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

Jushan Bai and Serena Ng. Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629, 2009.

André Binette and Jae Chang. CSI: A model for tracking short-term growth in Canadian real GDP. *Bank of Canada Review*, 2013(Summer):3–12, 2013.

Giorgio Bodo, Roberto Golinelli, and Giuseppe Parigi. Forecasting industrial production in the Euro area. *Empirical economics*, 25(4):541–561, 2000.

Jean Boivin and Serena Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194, 2006.

Zhihong Chen, Azhar Iqbal, and Huiwen Lai. Forecasting the probability of US recessions: a probit and dynamic factor modelling approach. *Canadian Journal of Economics/Revue canadienne d'économique*, 44(2):651–672, 2011.

Tony Chernis and Rodrigo Sekkel. A dynamic factor model for nowcasting Canadian GDP growth. Technical report, February 2017. Bank of Canada Staff Working Paper 2017-02.

Calista Cheung and Frédérick Demers. Evaluating forecasts from factor models for Canadian GDP growth and core inflation. *Bank of Canada Working Paper*, 2007(8), 2007.

Charlotte Christiansen, Jonas Nygaard Eriksen, and Stig Vinther Møller. Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, 49:459–468, 2014.

Richard Curtin. What recession? What recovery? The arrival of the 21st century consumer: consumer ambiguity is likely to be a feature of the US economy for some time to come. *Business Economics*, 38(2):25–33, 2003.

Francis X Diebold and Roberto S Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.

Sandra Eickmeier and Tim Ng. Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting*, 27(2): 496–511, 2011.

Umar Faruqui, Paul Gilbert, and Wendy Kei. The Bank of Canada's senior loan officer survey. *Bank of Canada Review*, 2008(Autumn):57–64, 2008.

Laurent Ferrara, Massimiliano Marcellino, and Matteo Mogliani. Macroeconomic forecasting during the great recession: The return of non-linearity? *International Journal of Forecasting*, 31(3):664–679, 2015.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840, 2005.

O. Fortin-Gagnon, M. Leroux, D. Stevanovic, and S. Surprenant. A large Canadian database for macroeconomic analysis. Manuscript, July. UQAM, 2018.

Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, 2006.

Jesper Hansson, Per Jansson, and Marten Löf. Business survey data: do they help in forecasting GDP growth? *International Journal of Forecasting*, 21:377–389, 2005.

Yawen Hudson and Christopher J Green. Is investor sentiment contagious? International sentiment and UK equity returns. *Journal of Behavioral and Experimental Finance*, 5: 46–59, 2015.

RA Johnson and DW Wichern. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 2007.

K. Jurado, S. C. Ludvigson, and S. Ng. Measuring uncertainty. *The American Economic Review*, 105:1117–1216, 2015.

Alexandre Kopoin, Kevin Moran, and Jean-Pierre Paré. Forecasting regional GDP with factor models: How useful are national and international data? *Economics Letters*, 121 (2):267–270, 2013.

R. Kotchoni and D. Leroux, M. Stevanovic. Macroeconomic forecast accuracy in a data-rich environment. Manuscript, 2018.

Kajal Lahiri and George Monokroussos. Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, 29(4):644–658, 2013.

Kajal Lahiri, George Monokroussos, and Yongchen Zhao. Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*, 31:1254–1275, 2016.

Maxime Leroux, Rachidi Kotchoni, and Dalibor Stevanovic. Forecasting economic activity in data-rich environment. *CIRANO Scientific Series*, 2017(s-05), 2017.

Kjetil Martinsen, Francesco Ravazzolo, and Fredrik Wulfsberg. Forecasting macroeconomic variables using disaggregate survey data. *International Journal of Forecasting*, 30(1):65–77, 2014.

John G Matsusaka and Argia M Sbordone. Consumer confidence and economic fluctuations. *Economic Inquiry*, 33(2):296–318, 1995.

M. W. McCracken and Serena Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016a.

Michael W McCracken and Serena Ng. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016b.

Kevin Moran, Simplice Aimé Nono, and Imad Rherrad. Does confidence data help forecast business cycles? new evidence from Canada. *Applied Economics*, 2018. Accepted for Publication.

Lise Pichette and M. N. Robitaille. Assessing the *Business Outlook Survey* underlying indicator using real-time data. Bank of Canada Discussion Paper, 2017.

Teresa Santero and Niels Westerlund. Confidence indicators and their relationship to changes in economic activity. *OECD Economics Department Working Paper 170*, 1996.

Christian Schumacher. Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*, 26(4):271–302, 2007.

Christian Schumacher. Factor forecasting using international targeted predictors: the case of German GDP. *Economics Letters*, 107(2):95–98, 2010.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002a.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002b.

James H Stock and Mark W Watson. Forecasting with many predictors. *Handbook of economic forecasting*, 1:515–554, 2006.

Karl Taylor and Robert McNabb. Business cycles and the role of confidence: Evidence for Europe. *Oxford Bulletin of Economics and Statistics*, 69(2):185–208, 2007.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Table 3 – Forecasting Performance with no targeting

| | *Targeting method: None* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ | $h=7$ | $h=8$ |
| | *Panel A:* $MSFE_{relative}$ with respect to *the AR model* | | | | | | | |
| $CA$ | 0.863*** | 1.027* | 1.038** | 0.914*** | 0.995 | 0.982 | 0.985 | 0.989 |
| $CAc$ | 1.172*** | 1.040** | 1.040** | 0.976 | 1.013 | 0.994 | 0.944*** | 0.949** |
| $US$ | 1.042*** | 1.073*** | 1.122*** | 1.026* | 1.078*** | 1.025* | 0.964** | 0.961** |
| $USc$ | 1.181* | 1.051** | 1.056* | 1.191*** | 1.035 | 1.053** | 0.992*** | 0.989* |
| $CA+CAc$ | 0.834*** | 1.025* | 1.040** | 0.915*** | 0.994 | 0.985 | 0.989 | 0.969** |
| $CA+US$ | 0.841*** | 1.017 | 1.021 | 0.941*** | 0.994 | 0.983 | 0.986 | 0.973* |
| $CA+USc$ | 0.985* | 1.021** | 1.016* | 1.091*** | 1.005 | 1.023** | 0.962*** | 0.959* |
| $CA+CAc+US$ | 0.823*** | 1.016 | 1.025* | 0.939*** | 0.995 | 0.986 | 0.989 | 0.972* |
| $CA+CAc+USc$ | 0.812*** | 1.030 | 1.073*** | 0.927*** | 0.991 | 1.004 | 0.991 | 0.983 |
| $CA+US+USc$ | 0.891*** | 1.002 | 1.014 | 1.013 | 0.988 | 0.987 | 0.966*** | 0.957* |
| $CA+CAc+US+USc$ | 0.920*** | 1.017* | 1.049*** | 1.017* | 0.981* | 0.994 | 0.978** | 0.970* |
| | *Panel B:* $MSFE_{relative}$ with respect to *the CA model* | | | | | | | |
| $CA+CAc$ | 0.966*** | 0.998* | 1.002** | 1.001*** | 0.999 | 1.003 | 1.004 | 0.980** |
| $CA+US$ | 0.975*** | 0.990 | 0.984 | 1.030*** | 0.999 | 1.001 | 1.001 | 0.984* |
| $CA+USc$ | 1.141* | 0.994** | 0.979* | 1.194*** | 1.010 | 1.042** | 0.977*** | 0.970* |
| $CA+CAc+US$ | 0.954*** | 0.989 | 0.987* | 1.027*** | 0.999 | 1.004 | 1.004 | 0.983* |
| $CA+CAc+USc$ | 0.941*** | 1.003 | 1.034*** | 1.014*** | 0.996 | 1.022 | 1.006 | 0.994 |
| $CA+US+USc$ | 1.032*** | 0.976 | 0.977 | 1.108 | 0.993 | 1.005 | 0.981*** | 0.968* |
| $CA+CAc+US+USc$ | 1.006*** | 0.990* | 1.011*** | 1.113* | 0.986* | 1.012 | 0.993** | 0.981** |

*Notes*: Results from forecasting Canadian GDP growth using four separate datasets: the ones labelled *CA* and *US* contain standard Canadian and US macroeconomic and financial variables, respectively, while *CAc* and *USc* incorporate Canadian and US confidence data. Factors are extracted as indicated in the test and the forecasting equation are estimated and forecasts computed separately for each horizon, using a rolling window. Each table entry is the relative mean squared forecast error ($MSFE_{relative}$) ie. the ratio of the mean squared forecast error from the factor model to that obtained with the AR benchmark using only lags of the dependent variable as predictors (Panel A) or with the *CA* model, which uses only standard Canadian macroeconomic variables as predictors (Panel B). Entries under 1 suggests that the factor model displays better forecasting performance than the benchmark. The quotes *, ** and *** indicate that the null hypothesis of equal predictive accuracy is rejected at the 10%, 5% and 1% level, respectively according to the Giacomini and White (2006) test.

Table 4 – Forecasting Performance with Targeting I: Hard-Thresholding with $t^\star = 1.28$

| | *Targeting: Hard thresholding with target t-statistic = 1.28* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ | $h = 5$ | $h = 6$ | $h = 7$ | $h = 8$ |
| | *Panel A:* $MSFE_{relative}$ with respect to *the AR model* | | | | | | | |
| $CA$ | 0.883*** | 1.062*** | 1.045** | 0.928*** | 0.992 | 1.022 | 0.992 | 0.963** |
| $CAc$ | 1.220*** | 1.026 | 1.061** | 0.974 | 1.027 | 0.997 | 0.957** | 0.999* |
| $US$ | 1.048*** | 1.061*** | 1.106*** | 1.029* | 1.061*** | 1.016 | 0.943*** | 0.921*** |
| $USc$ | 1.270*** | 1.062** | 1.064 | 0.989*** | 1.011 | 0.949*** | 0.966*** | 1.009*** |
| | *Panel B:* $MSFE_{relative}$ with respect to *the CA model* | | | | | | | |
| $CA + CAc$ | 0.951*** | 0.956 | 0.996** | 0.999*** | 1.011 | 0.962 | 1.008 | 1.003** |
| $CA + US$ | 0.973*** | 0.985*** | 0.968 | 1.064 | 0.999 | 0.954* | 0.962*** | 0.970*** |
| $CA + USc$ | 1.047*** | 0.962** | 0.964 | 1.109*** | 1.005 | 0.949*** | 0.966*** | 1.009*** |
| $CA + CAc + US$ | 0.963*** | 0.999*** | 0.976 | 1.083 | 0.999 | 0.949* | 0.965** | 0.942*** |
| $CA + CAc + USc$ | 0.930*** | 0.962 | 0.975 | 1.079 | 1.010 | 0.955 | 0.970** | 1.070* |
| $CA + US + USc$ | 1.002*** | 0.967** | 0.967 | 1.091 | 0.997 | 0.982 | 0.976** | 0.973*** |
| $CA + CAc + US + USc$ | 0.992*** | 0.975** | 0.994** | 1.099 | 0.992 | 0.981 | 0.974** | 0.998** |

*Notes*: See endnotes from Table 3

Table 5 – Forecasting Performance with Targeting II: Hard-Thresholding with $t^\star = 1.65$

| | *Targeting: Hard thresholding with target t-statistic* = 1.65 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ | $h = 5$ | $h = 6$ | $h = 7$ | $h = 8$ |
| | *Panel A:* $MSFE_{relative}$ with respect to *the AR model* | | | | | | | |
| $CA$ | 0.893*** | 1.012*** | 1.044** | 0.929*** | 1.009 | 0.971 | 0.990 | 0.955** |
| $CAc$ | 1.152*** | 1.016 | 1.043** | 0.977 | 1.021 | 1.006 | 0.962** | 0.992* |
| $US$ | 1.022*** | 1.068*** | 1.132*** | 1.014* | 1.042*** | 1.020 | 0.950*** | 0.937*** |
| $USc$ | 1.167*** | 1.022** | 1.064 | 1.109*** | 0.998 | 1.010*** | 0.966*** | 1.009*** |
| | *Panel B:* $MSFE_{relative}$ with respect to *the CA model* | | | | | | | |
| $CA + CAc$ | 0.980*** | 1.009* | 1.005*** | 1.013*** | 0.998 | 0.996** | 1.012 | 0.986*** |
| $CA + US$ | 0.959*** | 1.048*** | 0.970 | 1.070 | 0.983 | 1.028 | 0.997 | 0.952*** |
| $CA + USc$ | 1.069*** | 1.048*** | 1.011*** | 1.115** | 0.999 | 1.001** | 0.974*** | 0.985*** |
| $CA + CAc + US$ | 0.961*** | 1.030*** | 0.985** | 1.071 | 0.975 | 1.030 | 0.996 | 1.051 |
| $CA + CAc + USc$ | 0.938*** | 1.007 | 0.975 | 1.039** | 0.999 | 1.006* | 0.962*** | 1.016** |
| $CA + US + USc$ | 0.972*** | 1.025** | 0.971 | 1.086 | 0.981 | 1.034 | 0.972** | 0.975*** |
| $CA + CAc + US + USc$ | 0.997*** | 1.024** | 0.998*** | 1.091 | 0.974 | 1.007* | 0.963*** | 0.970*** |

*Notes*: See endnotes from Table 3

Table 6 – Forecasting Performance with Targeting I: Hard-Thresholding with $t^{\star} = 2.58$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *Targeting method: Hard thresholding with target tstatistic* $= 2.58$ | | | | | | | |
| Dataset | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ | $h = 5$ | $h = 6$ | $h = 7$ | $h = 8$ |
| | *Panel A:* $MSFE_{relative}$ with respect to *the AR model* | | | | | | | |
| $CA$ | 0.811*** | 0.999 | 1.024 | 0.887*** | 0.988 | 0.980 | 0.984 | 0.954** |
| $CAc$ | 1.148*** | 1.016 | 1.065*** | 0.980 | 1.044*** | 1.009 | 0.988 | 1.006 |
| $US$ | 1.048*** | 1.051*** | 1.133*** | 1.030** | 1.049*** | 1.016 | 0.948*** | 0.954*** |
| $USc$ | 1.133*** | 1.049** | 1.091 | 1.073** | 1.060 | 1.025 | 1.004 | 1.024 |
| | *Panel B:* $MSFE_{relative}$ with respect to *the CA model* | | | | | | | |
| $CA + CAc$ | 1.017*** | 1.002 | 0.993 | 1.041*** | 1.001 | 0.996* | 0.992* | 0.996*** |
| $CA + US$ | 1.033*** | 1.035** | 0.963 | 1.063*** | 1.003 | 1.008 | 0.974*** | 0.968*** |
| $CA + USc$ | 1.033*** | 1.049** | 0.991 | 1.073** | 1.020 | 1.025 | 1.034 | 1.024 |
| $CA + CAc + US$ | 1.064*** | 1.035** | 0.965 | 1.066*** | 1.003 | 1.025 | 0.976** | 0.969*** |
| $CA + CAc + USc$ | 1.012*** | 0.999 | 0.992 | 1.060*** | 1.006 | 1.011 | 0.983** | 1.018* |
| $CA + US + USc$ | 1.074*** | 1.044*** | 0.964 | 1.081*** | 1.006 | 1.020 | 0.975*** | 0.966*** |
| $CA + CAc + US + USc$ | 1.079*** | 1.045*** | 0.974 | 1.057*** | 1.006 | 1.024 | 0.973*** | 0.971*** |

*Notes*: See endnotes from Table 3

Table 7 – Forecasting Performance with Predictor Soft-thresholding

| | $MSFE_{relative}$ with respect to *the CA* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ | $h=7$ | $h=8$ |
| | *Panel A: Targeting method LARS-EN soft threshold with* $N_{\mathcal{A}}=15$ | | | | | | | |
| $CA+CAc$ | 0.982*** | 1.080*** | 1.183*** | 0.976* | 0.968 | 0.973 | 1.023 | 1.010 |
| $CA+US$ | 1.013*** | 0.982 | 0.995 | 0.962** | 1.013 | 0.982 | 0.999 | 0.996 |
| $CA+USc$ | 1.011*** | 0.994 | 1.007 | 1.021 | 0.964 | 0.983 | 1.001* | 1.025 |
| $CA+CAc+US$ | 0.997*** | 1.012 | 1.003 | 0.981** | 1.022** | 0.989 | 1.012 | 1.009 |
| $CA+CAc+USc$ | 0.837 | 1.106** | 1.152*** | 0.901** | 1.143*** | 1.230*** | 0.913** | 1.429*** |
| $CA+US+USc$ | 0.898*** | 1.007 | 1.001* | 0.980*** | 1.022*** | 0.968** | 1.010** | 1.006 |
| $CA+CAc+US+USc$ | 0.975*** | 1.038*** | 1.012** | 1.012 | 1.022** | 0.972 | 1.005* | 1.014 |
| | *Panel B: Targeting method LARS-EN soft threshold with* $N_{\mathcal{A}}=30$ | | | | | | | |
| $CA+CAc$ | 0.852 | 1.162*** | 1.016*** | 0.936 | 1.042 | 1.188*** | 0.858** | 1.050 |
| $CA+US$ | 0.970*** | 0.997 | 0.816 | 0.918*** | 1.040 | 1.103 | 0.903 | 1.033 |
| $CA+USc$ | 0.971*** | 0.996 | 0.804 | 0.927** | 1.034 | 1.093 | 0.910 | 1.041 |
| $CA+CAc+US$ | 0.703*** | 1.034*** | 0.924*** | 1.263*** | 0.931** | 1.015*** | 0.929 | 0.988*** |
| $CA+CAc+USc$ | 0.745** | 1.042*** | 0.888*** | 1.132** | 1.080 | 1.117*** | 0.823 | 1.047*** |
| $CA+US+USc$ | 0.772*** | 0.973 | 0.849** | 1.039*** | 0.983** | 1.176*** | 0.851*** | 1.033 |
| $CA+CAc+US+USc$ | 0.775*** | 1.048*** | 0.939** | 1.312 | 0.921** | 1.002 | 0.915* | 0.993 |
| | *Panel C: Targeting method LARS-EN soft threshold with* $N_{\mathcal{A}}=75$ | | | | | | | |
| $CA+CAc$ | 0.895 | 1.130*** | 1.190*** | 0.780*** | 1.091 | 1.009 | 1.115 | 1.020 |
| $CA+US$ | 0.899** | 1.024* | 1.031*** | 0.924** | 1.030*** | 0.942*** | 1.081** | 0.994*** |
| $CA+USc$ | 1.035*** | 0.997 | 0.946 | 0.885*** | 1.083 | 0.990 | 1.114 | 1.039 |
| $CA+CAc+US$ | 0.959*** | 0.936*** | 1.026*** | 0.889*** | 1.113** | 0.914*** | 1.127 | 0.991*** |
| $CA+CAc+USc$ | 0.771** | 1.454*** | 1.336*** | 1.123** | 1.022 | 1.282*** | 1.086 | 0.780*** |
| $CA+US+USc$ | 0.784*** | 1.037* | 1.001** | 0.991** | 1.049 | 1.060*** | 0.998*** | 1.050 |
| $CA+CAc+US+USc$ | 0.923*** | 1.047* | 1.010** | 0.997** | 1.019 | 0.910*** | 0.998*** | 1.005 |

*Notes*: See endnotes from Table 3

Table 8 – Forecasting Performance with Factor Targeting

| | $MSFE_{relative}$ with respect to *the CA model* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Dataset | $h=1$ | $h=2$ | $h=3$ | $h=4$ | $h=5$ | $h=6$ | $h=7$ | $h=8$ |
| | *Panel A: Factor Hard thresholding with target tstatistic* $= 1.28$ | | | | | | | |
| $CA+CAc$ | 0.990*** | 0.974 | 1.029 | 0.982*** | 1.008* | 1.038* | 1.050 | 1.005** |
| $CA+US$ | 0.970*** | 0.997 | 1.033 | 1.045*** | 0.972 | 1.005 | 0.994 | 1.001* |
| $CA+USc$ | 1.101*** | 1.010 | 1.366 | 1.120** | 1.034 | 1.093 | 1.013 | 1.041 |
| $CA+CAc+US$ | 1.041*** | 1.013*** | 0.935*** | 1.403*** | 0.910** | 1.038*** | 0.936 | 1.016*** |
| $CA+CAc+USc$ | 1.026*** | 0.990 | 1.049** | 1.159*** | 0.974 | 1.026 | 0.971** | 0.9720** |
| $CA+US+USc$ | 1.029*** | 0.982 | 1.026 | 1.125 | 0.966 | 1.009 | 0.974*** | 1.033* |
| $CA+CAc+US+USc$ | 1.062*** | 0.997* | 1.062*** | 1.130* | 0.959* | 1.016 | 0.986** | 0.998* |
| | *Panel B: Factor Hard thresholding with target tstatistic* $= 1.65$ | | | | | | | |
| $CA+CAc$ | 0.963*** | 0.933*** | 1.093** | 1.005*** | 0.995 | 1.054*** | 1.022* | 0.990*** |
| $CA+US$ | 0.877*** | 1.024*** | 1.115** | 1.004** | 0.983 | 0.987** | 1.010 | 0.995** |
| $CA+USc$ | 0.929** | 1.124*** | 1.314*** | 1.238*** | 1.113*** | 0.928*** | 1.085** | 1.059** |
| $CA+CAc+US$ | 0.844*** | 0.981 | 1.101 | 0.988*** | 1.032** | 1.053* | 1.018 | 0.986** |
| $CA+CAc+USc$ | 0.822*** | 0.947 | 1.128*** | 1.069 | 0.979 | 1.048* | 0.923*** | 0.978 |
| $CA+US+USc$ | 0.910*** | 0.967 | 1.088 | 1.065 | 0.973 | 1.009 | 0.974*** | 1.050 |
| $CA+CAc+US+USc$ | 0.939*** | 0.982* | 1.127*** | 1.070* | 0.966* | 1.016 | 0.986*** | 1.010 |
| | *Panel C: Factor Hard thresholding with target tstatistic* $= 2.58$ | | | | | | | |
| $CA+CAc$ | 0.996*** | 1.000** | 1.009*** | 1.001*** | 0.999 | 1.003*** | 0.978*** | 0.995*** |
| $CA+US$ | 0.991*** | 1.001** | 1.018** | 1.021** | 0.961*** | 0.997*** | 0.963*** | 1.004*** |
| $CA+USc$ | 1.024* | 1.011** | 1.046* | 1.090*** | 0.975** | 1.027 | 0.935*** | 0.9637 |
| $CA+CAc+US$ | 0.926*** | 1.034*** | 1.005** | 1.022 | 0.996 | 1.000** | 0.962*** | 1.001*** |
| $CA+CAc+USc$ | 0.922*** | 1.009 | 1.069** | 1.076*** | 0.982* | 1.014 | 0.982 | 0.990 |
| $CA+US+USc$ | 0.951*** | 1.027*** | 1.039 | 1.052* | 0.989 | 0.997** | 0.960*** | 1.020 |
| $CA+CAc+US+USc$ | 0.956*** | 1.050*** | 1.083*** | 1.055* | 0.985 | 1.005* | 0.967** | 1.012 |

*Notes*: See endnotes from Table 3

# A Canadian Surveys on Sentiment

## A.1 Conference Board Consumer Confidence Survey

The Conference Board of Canada operates a monthly survey of Canadian households to measure levels of optimism regarding current and future economic conditions. Surveyed households are asked to give their views about their current and expected financial positions and the short-term employment outlook. In addition, they are also asked to assess whether now is a good time or a bad time to make a major purchase such as a house, car or other big-ticket items. Specifically, the four questions asked are as follows:

1. Considering everything, would you say that your family is better or worse off financially than six months ago?
2. Again, considering everything, do you think that your family will be better off, the same or worse off financially six months from now?
3. How do you feel the job situation and overall employment will be in this community six months from now?
4. Do you think that right now is a good or bad time for the average person to make a major outlay for items such as a home, car or other major item?

## A.2 Conference Board Business Confidence Survey

The Conference Board of Canada also operates a quarterly survey of Chief Executive Officers and Chief Financial Officers at Canadian business organizations. The survey measures perceptions of the economic environment and investment intentions. The questions are are as follows.

Do you expect overall economic conditions in Canada six months from now to be:

• Better • Worse • The same

Do you expect prices, in general, in Canada to increase over the next six months at an annual rate of:

• 1% • 2% • 3% • 4% • 5% • 6% • 7% • 8% • $> 8\%$

Over the next six months, do you expect your firm's financial position to:

• Improve • Worsen • Remain the same

Over the next six months, do you expect your firm's profitability to:

• Improve • Worsen • Remain the same

Would you say the present is a good or a bad time to undertake expenditures to expand your plant or add to your stock of machinery and equipment?

• Good • Bad • Not sure

What change in the level of your capital investment expenditures do you expect over the next 6 months?

• Up 20% • Up 10% to 19% • Up 1% to 9% • No change • Down 1% to 9% • Down 10% to 19% • Down 20% or more

In which region(s) of the country do you expect the bulk of your planned investment expenditures for the next six months to take place?

• Atlantic Provinces • Quebec • Ontario • Prairie Provinces • British Columbia • United States • International

How do you assess your current level of operations relative to optimal capacity?

• Above capacity • At or close to capacity • At, close to, or above • Slightly below capacity • Substantially below capacity

## A.3 Bank of Canada Business Outlook Survey

Questions included in the Business Outlook Survey are grouped in four broad categories:

1. Past Business Conditions

   Past sales: The growth of sales volume (adjusted for price changes) over the past 12 months (compared with growth over the previous 12 months) was:

   • Greater • Less • The same

2. Outlook for Business Activity

   Future sales: The growth of sales volume over the next 12 months (compared with growth over the past 12 months) is expected to be:

• Greater • Less • The same

Investment intentions for machinery and equipment: The level of investment spending on machinery and equipment over the next 12 months is expected to be:

• Higher • Lower • The same

Investment intentions for buildings: The level of investment spending on buildings over the next 12 months is expected to be:

• Higher • Lower • The same

Outlook for employment: The number of employees (full-time equivalent) employed by your organization over the next 12 months is expected to be:

• Higher • Lower • The same

3. Pressures on Production Capacity

   Labour shortages: The organization is facing shortages of labour that restrict the ability to meet demand:

   • Yes • No

   Ability to meet demand: Currently, the potential level of difficulty in meeting an unexpected increase in demand or sales would be:

   • No difficulty (operating below capacity) • Some difficulty (at or near full capacity)
   • Significant difficulty (operating above capacity)

4. Outlook for Wages, Prices, and Inflation

   Outlook for wages: The increase in labour costs (per hour) over the next 12 months is expected to be:

   • Greater • Less • The same

   Outlook for input prices: The increase in the prices of products or services purchased over the next 12 months is expected to be:

• Greater • Less • The same

Outlook for output prices: The increase in the prices of products or services that are sold over the next 12 months is expected to be:

• Greater • Less • The same

Inflation-expectations index: The firm's expectation for the average annual rate of inflation over the next two years as measured by the consumer price index is:

• Above 3% • 2% to 3% • 1% to 2% • Below 1%

### A.4 Bank of Canada Senior Loan Officer Survey

This survey asks the following question to representatives from financial institutions : How have your institution's general standards (i.e. your appetite for risk) and terms for approving credit changed in the past three months?

• Tightened • Eased • Remain unchanged

Respondents indicate how their practices about business lending conditions evolved by taking into account each of the following conditions:

a. Pricing of credit (spreads over base rates, fees),

b. General standards,

c. Limit of capital allocation,

d. Terms of credit (collateral, covenants, etc.).

Another question is asked about loans provided to corporate, commercial and small business firms. The responses for commercial and small business firms are provided for five regions: British Columbia, the Prairies, Ontario, Quebec, and the Atlantic provinces. Corporate, commercial, and small business firms are further differentiated by the size of the loans authorized for each, using the following suggested definitions: (a) Corporate: over $50 million, (b) Commercial: between $2 and $50 million, and (c) Small business: less than $2 million. Respondents may answer based on internal reporting definitions, which may differ from the definitions suggested.