CSI: A Model for Tracking Short-Term Growth in Canadian Real GDP

André Binette and Jae Chang, Canadian Economic Analysis

- The formulation of monetary policy requires central banks to assess the current state of the economy in a timely fashion. A variety of tools can be used to conduct this current analysis.
- Forecasting short-term growth in real GDP is a challenging task, given the wide range of potentially useful economic indicators and delays in the availability of data. Factor models offer a way to summarize the predictive content of many indicators without abandoning useful explanatory information in any of the series.
- Canada's Short-Term Indicator (CSI) is a new state-of-the-art indicator model for Canada that exploits the information content of 32 indicators to produce daily updates of real GDP growth forecasts for the two quarters following the latest release of official data.
- Although the forecast accuracy of this new model is encouraging, current analysis should not rely mechanically on predictions from a single model. Indeed, the Bank of Canada uses a wide range of models and information sources, as well as expert judgment, in producing its short-term forecasts.

The formulation of monetary policy relies, in part, on analysis of a variety of information about current economic conditions. Through current analysis,[1] economists try to understand and gauge the implications of the most recent economic conditions, including the impact of unpredictable events, such as natural disasters and work stoppages. Consequently, timely and accurate data are important for current analysis, since a clear understanding of current events is critical to better predict future developments. This in turn allows for the appropriate monetary policy response, given the forward-looking nature of the monetary policy approach.

The well-known maxim, "We need to know where we have been to know where we are going," highlights the value of short-term forecasting and early assessment, which are key facets of current analysis. To guide its monetary policy actions, the Bank of Canada devotes considerable time and

1 Current analysis is the process of collecting and analyzing a large amount of current information. This process is essential for monitoring and predicting short-term economic activity (see Coletti and Kozicki in this issue). Throughout this article, the terms "monitoring" and "short-term forecasting" are used interchangeably.

resources to monitoring and predicting short-term economic activity, as measured by real gross domestic product (GDP), and inflation. The Bank is continually developing new tools to improve its ability to predict economic developments over the short term, which is typically two quarters after the latest release of official GDP data.

The Bank of Canada is continually developing new tools to improve its ability to predict economic developments over the short term

Forecasting short-term growth in real GDP presents a number of challenges. Economists have a large number of data series at their disposal, ranging from National Accounts data to credit aggregates. From this profusion of data, they must extract the right information. As well, many indicators are published with lags, some of which are as long as two months. Economists need to find the best way to address the problems caused by these delays in the publication and revision of data. Another challenge is to develop tools that can use series with different frequencies, since data are published at daily, weekly, monthly or quarterly frequencies. The high-frequency data could provide useful information; for example, if economists consider quarterly data only, other information available over the course of the quarter, such as daily data on stock market indexes, could be lost. Another challenge involves truncated series resulting largely from redefinitions of variables.

The statistical agencies that produce these data face a trade-off between timeliness and the accuracy of the initial release. A lack of both timely and accurate economic data can lead to inaccurate conclusions about the state of the economy. Taken together, these challenges make current analysis a complex process. This article focuses on the forecasting aspect of current analysis and describes a recently developed state-of-the-art indicator model for tracking short-term growth in Canadian real GDP. This model can accommodate most challenges in current analysis at the same time and can complement other information, since the Bank uses a wide range of models and information sources as well as expert judgment in producing its short-term forecasts.[2]

The Bank uses a wide range of models and information sources as well as expert judgment in producing its short-term forecasts

Factor Models as Tools for Monitoring Economic Developments

Several statistical tools assist Bank staff in monitoring short-term economic developments. These tools are often econometric models, which are simplified mathematical approximations of a complicated and evolving reality. The variables included in these models are based on economic theory, and statistical techniques are used to identify relationships among them.

Research has demonstrated the potential of factor models to address the main challenge of current analysis—extracting useful information from abundant data on multiple indicators. Factor models describe the relationship among observed correlated variables in terms of a few unobserved variables, called factors. The premise of these models is that the factors explain the variation and common movement in a large number of observed variables. For example, movements in real GDP are correlated with changes in other measured variables such as employment and consumer confidence. Factor models formalize the idea that the true business cycle is not directly observed and is best measured by estimating the common movements of various economic time series (Burns and Mitchell 1946; Lucas 1977) (**Box 1**). By uncovering the underlying common movements, information from a variety of indicators can be used to forecast growth in real GDP.

Research has demonstrated the potential of factor models to extract useful information from abundant data on multiple indicators

2 Granziera, Luu and St-Amant (this issue) find that combining forecasts from different models generally improves forecast accuracy when compared with various benchmarks.

Box 1

Factor Models: Specification and Estimation

In general, factor models can be specified as:

$$x_{it} = \lambda_i f_t + e_{it},$$

where

$$f_t = \sum_{j=0}^{q} \varphi_j f_{t-j} + u_t.$$

x_{it} is one of N observed variables in the model and t represents the time period. Each variable $\{x_{it}\}_{i=1}^{N}$ is assumed to depend on a latent (unobserved) factor, f_t, and an idiosyncratic component, e_{it}. The term $\lambda_i f_t$ is the common component underlying x_{it}, with λ_i being the corresponding factor loading for variable i. The factor loading can be defined as the marginal effect of the unobserved factor f_t on x_{it}. All idiosyncratic components are assumed to be uncorrelated with each other and also uncorrelated with the unobserved common component. The factor f_t is assumed to follow a covariance-stationary autoregressive process. If f_t were known, parameter estimates of λ_i and φ_j could be obtained from regression analysis. Unfortunately, f_t, λ_i and φ_j are all unknown. The only known elements in the above system of equations are the observed data in $\{x_{it}\}_{i=1}^{N}$.

Since both λ_i and f_t are unobserved, the factor model is not identified, in the sense that some restrictions have to be imposed in order to estimate the model. For a relatively small number of variables, and assuming that residuals are normally distributed, maximum likelihood and the Kalman filter can be used to obtain estimates of the factor loadings and the common factor (Stock and Watson 1991; Kalman 1960). An alternative methodology for estimating the common latent factor is principal-component analysis (PCA), created in 1901 by British mathematician Karl Pearson (Pearson 1901). Empirical evidence suggests that, for purely predictive purposes on a given data set, factor models estimated by the Kalman filter generally have a similar performance to those estimated by PCA (Boivin and Ng 2005).

Employing these models to uncover useful patterns is called factor analysis, which was first developed in 1904 by the British psychologist Charles Spearman in the field of intelligence research (Spearman 1904). Spearman theorized that seemingly disparate cognitive test scores could be explained by a single general intelligence factor. Geweke (1977) and Sargent and Sims (1977) were among the earliest researchers to model economic time series with factor models.

Factor models offer a way to summarize the predictive content of many indicators without abandoning the relevant information in any of the series. If all indicators at a given point in time move together, the model will easily discover the general upward or downward trend in the series. When many indicators move in different directions and there is no obvious upward or downward trend, the conflicting signals are resolved by a weighted average, with series that are more informative (based on historical correlations) receiving more weight than less-informative series. Essentially, more-volatile series are often given less weight.

Canada's Short-Term Indicator Model

The Bank of Canada's factor model—Canada's Short-Term Indicator, or CSI—closely follows the approach of Camacho and Perez-Quiros (2010), which accommodates missing observations resulting from delays in the release of data, as well as data samples that represent short time spans, monthly and quarterly indicators (mixed frequencies), and different transformations of the data (monthly, quarterly and year-over-year growth rates). In addition, Camacho and Perez-Quiros (2010) include a way to deal with multiple GDP releases, which allows the information in the monthly real GDP figures for Canada to be exploited. The model is self-contained, since

it makes internal predictions for each indicator, enabling assessment of the impact of each new data release on the model's forecast of real GDP growth.

Forecasting with CSI follows three important steps: (i) collect information for a wide variety of economic indicators;[3] (ii) conduct a complete evaluation of the available information (at this stage, CSI analyzes the indicators and determines weights to assign to each of them); and (iii) calculate the common component and the forecast of real GDP growth.

Main features of CSI

CSI is a monthly, dynamic, single-factor model built on the principle that any series can be divided into two components: a component that is common to all variables in the model and an idiosyncratic component. All indicators in CSI are projected based on a common component and on their own individual dynamics, as described by autoregressive (AR) processes in which the current values of the indicators are explained by using only their past values. The empirical analysis uses data available from 1982 through to 2012.

CSI is built on the principle that any series can be divided into two components: a component that is common to all variables in the model and an idiosyncratic component

Although CSI is a monthly model, its indicators include quarterly variables. The model simply considers these variables to be monthly series with missing observations. The quarterly indicators are linked to the monthly factor using a mathematical relationship that expresses quarterly growth rates as monthly growth rates in both the current quarter (the quarter being measured) and the previous quarter (Statistics Canada 2011).[4] This relationship implies that about 66 per cent of the quarterly growth rate of a series is known after the release of the first month of a given quarter, and about 90 per cent is known after two months.

Unlike statistical agencies in the United States and some other countries, which publish preliminary and advance estimates of quarterly real GDP, Statistics Canada provides monthly GDP figures for Canada. Monthly and quarterly real GDP series are not conceptually identical: monthly figures are published at basic prices, while quarterly real GDP is expressed at market prices, which include net taxes on products. Notwithstanding the conceptual difference, the growth rates of the two measures of real GDP often exhibit a similar dynamic at a quarterly frequency. Consequently, it is assumed that, after the first and second month of the quarter, information on monthly real GDP reflects early estimates of quarterly real GDP at market prices. Thus, this key monthly indicator is treated in the same way that preliminary and advance estimates of quarterly GDP are dealt with by researchers using data for other countries. As noted in Camacho and Perez-Quiros (2010), these early estimates of GDP are incomplete and the difference between them and the final quarterly release is unpredictable.

CSI indicators

While factor models, in theory, could process the information content of a very large number of indicators, Boivin and Ng (2006) show that larger data sets do not necessarily generate more-accurate forecasts in empirical applications. The choice of indicators used in CSI has therefore been guided

3 Armah (this issue) discusses the rapid growth in the number of potential indicators resulting from the development of information technology.

4 The first month of the current quarter has the largest impact, with a weight of 1; the previous month and the following month have the next-largest impact (weight = 2/3); and the last month of the current quarter and the second month of the previous quarter have the least impact (weight = 1/3).

by the following criteria: (i) the variables should be directly related to the Canadian economy; and (ii) forecasts over the past decade should be more accurate than simple benchmarks found in the literature.[5]

The current specification of CSI includes 32 indicators, most of which are well-known statistics for Canada

Over time, Bank staff have evaluated the ability of various indicators to predict the growth rate of real GDP. The current specification of CSI includes 32 indicators (**Appendix 1**),[6] most of which are well-known statistics for Canada, such as total hours worked (from the Labour Force Survey), retail trade and housing starts. Other indicators include soft information (such as consumer confidence), financial data and international variables. U.S. data series and the global purchasing managers' index (PMI) for manufacturing are used to proxy foreign demand for Canadian exports.[7] As previously mentioned, after the first and second month, monthly GDP information provides early estimates of quarterly GDP. A quarterly momentum indicator related to monthly GDP is also incorporated in the model to capture the early dynamic of a new quarter.[8] The addition of timely soft information and financial indicators gives the model early information about the quarter of interest and potentially improves forecast accuracy.[9]

CSI performance

CSI explains about 75 per cent of the variation in the quarterly growth rate of real GDP over the 1982–2012 period

CSI is based on the premise that common movements (i.e., the common factor) that affect all indicators are linked to the business cycle, as measured by growth in real GDP. The common factor should therefore have a profile similar to GDP growth. In fact, the model performs relatively well, since it explains about 75 per cent of the variation in the quarterly growth rate of real GDP over the 1982–2012 period (**Chart 1**).

The estimation results (factor loadings) also suggest that all of the indicators retained in the model exhibit a positive correlation with the common factor.[10] Nevertheless, the strength of the correlation varies among indicators (**Chart 2**). As expected, the momentum indicator, as well as early estimates of GDP and quarterly GDP, present the strongest relationships with the common factor and therefore have the greatest impact on the model's forecast. For monthly variables, the link with the common factor varies by the type of indicator (hard, soft or financial). Most of the variables with above-average correlation are standard statistics (hard indicators), with the exception of the global PMI for manufacturing. While 15 monthly indicators have a below-average relationship with the common factor, 8 of them are timely, with very short publication lags. These indicators have been included in an attempt to improve the forecast performance early in the forecast cycle.

5 Forecasts generated by an AR model and the unconditional mean of the series are the benchmarks against which we have compared CSI forecasts.

6 We initially considered about 50 indicators and retained in the model only the indicators that provide information not available in other series.

7 Morel (2012) presents a measure of foreign activity that tracks historical export data relatively well. The measure includes U.S. consumption, U.S. residential investment, U.S. business investment and foreign GDP outside of the United States. As proxies for U.S. activity, CSI uses U.S. retail sales, U.S. car sales, U.S. housing starts and U.S. industrial production.

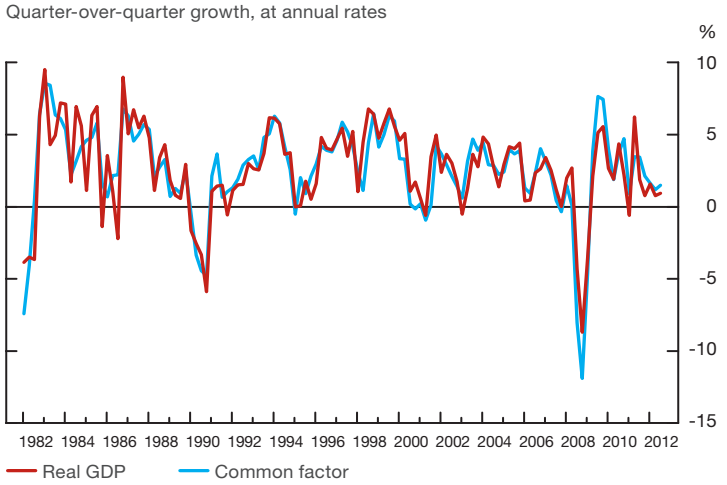
8 The momentum indicator is a quarterly series that exploits monthly GDP information from the last two months of the previous quarter to assess the vigour of economic activity at the start of a new quarter. The unpublished months of the momentum indicator and the early estimates are currently forecast with a moving average of the growth rate from the previous three months.

9 Most indicators are included in a difference-of-log format (growth rate), while some are incorporated in log-level format. The data transformations ensure that all indicators used in the model are stationary series and provide the best forecasting performance beyond the sample period.

10 A factor loading measures the change in an observed variable following a one-unit change in the common factor.

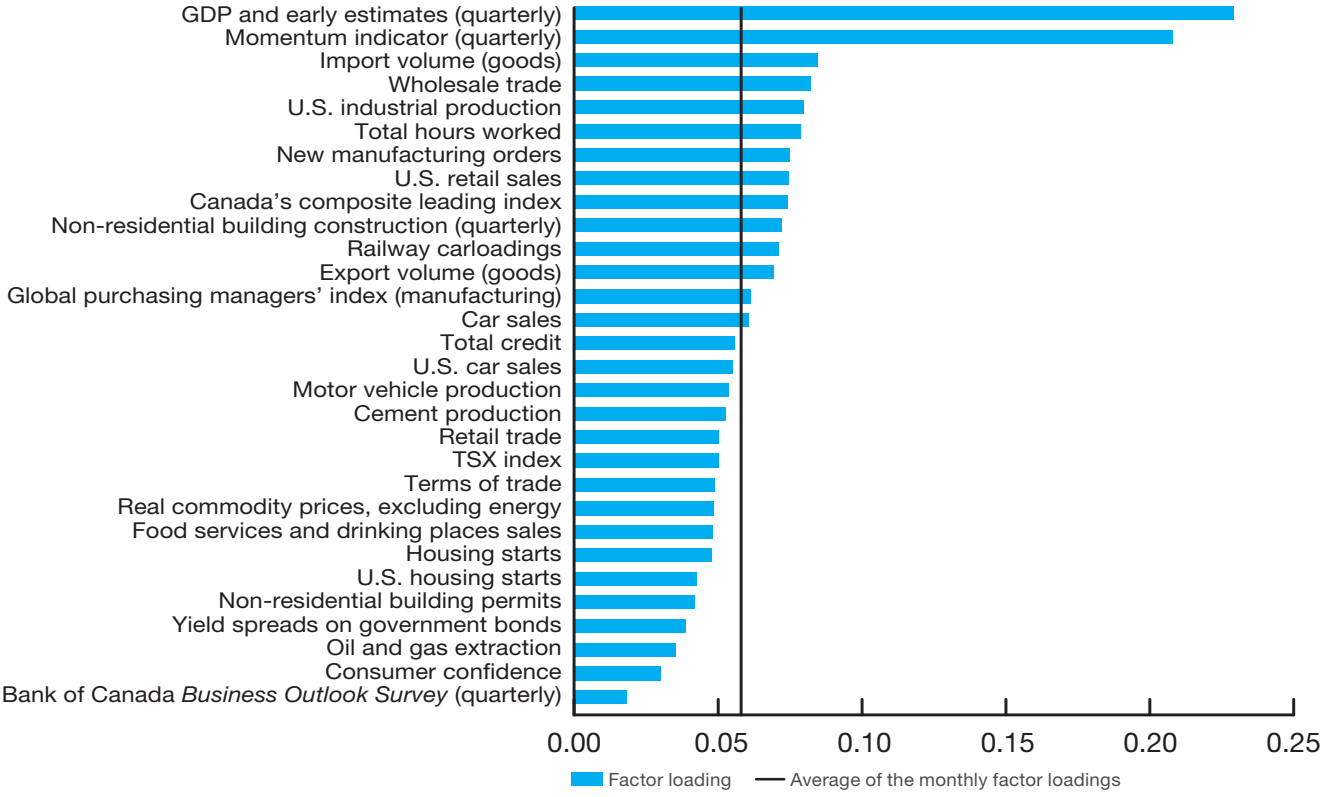
Chart 1: Quarterly growth in real GDP and the common factor

Quarter-over-quarter growth, at annual rates



Sources: Statistics Canada and Bank of Canada calculations Last observation: 2012Q4

Chart 2: CSI factor loadings

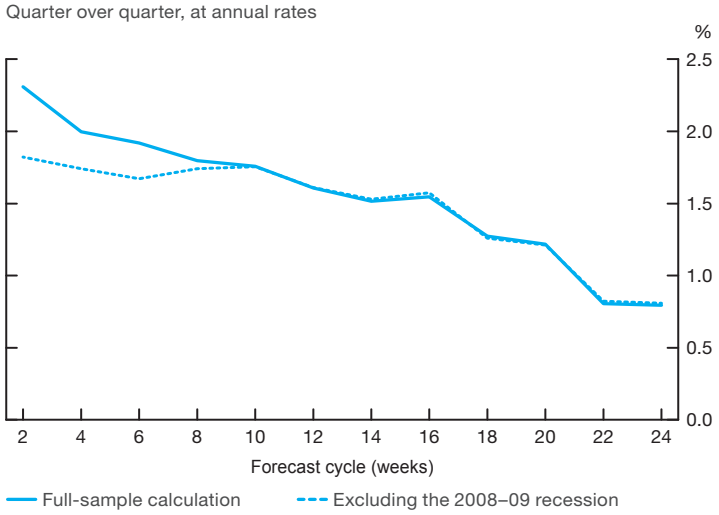


Note: See Box 1 for a discussion of factor loadings.
Source: Bank of Canada calculations

To assess the performance of CSI beyond the sample period, a quasi-real-time exercise is performed in which the model uses only the information available at the time that it makes its predictions. This approach mimics the actual conditions faced by analysts at the Bank. The exercise is conducted in quasi-real time, since the original unrevised data are not available for

Chart 3: Root-mean-square prediction errors, 2000Q1–2012Q4

Quarter over quarter, at annual rates



Source: Bank of Canada calculations

all the indicators.[11] The model's performance is assessed 12 times over the forecast cycle. For a given quarter, the cycle covers six months, representing a prediction every two weeks. For example, the initial forecast for the fourth quarter of 2012 was made in early September 2012, while the last prediction was made in the second half of February 2013 (i.e., just before the release of real GDP growth for the fourth quarter).[12]

Overall, the CSI model performs as anticipated, with accuracy increasing as more information becomes available

Overall, the CSI model performs as anticipated. The initial forecasts are not very accurate, with root-mean-square prediction errors (RMSPEs) above 2 per cent (**Chart 3**), in part, because of the model's inability to predict the severe economic downturn in 2008–09. The accuracy of CSI increases, however, as more information becomes available, and significant improvements occur in weeks 18 and 22 with the release of the monthly GDP data for the first two months of the quarter (early estimates). This should come as no surprise, since GDP at basic prices and GDP at market prices are highly correlated at the quarterly frequency, despite the small conceptual difference.

Another insightful measure of performance is the model's forecast memory, which is the ratio of the RMSPE to the standard deviation of quarterly GDP growth. When this ratio is above one, model forecasts are less accurate than a simple forecast that assumes GDP growth will equal the average of the series (i.e., the unconditional mean). Thus, the forecast memory indicates the horizon at which the indicators provide useful signals. As **Chart 4** shows, CSI provides valuable information above the unconditional mean as early as one month before the start of the quarter under consideration. For example, forecasts made by CSI in September (weeks 2 and 4) for the fourth quarter of a given year are, on average, more accurate than a forecast based on the unconditional mean of real GDP growth.

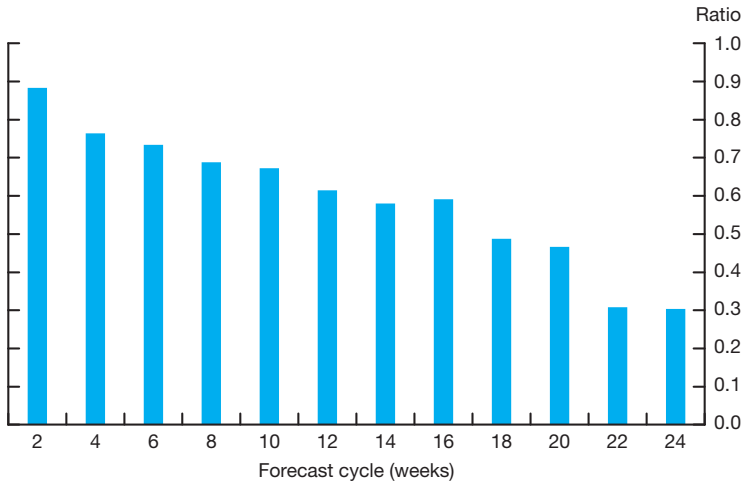
While forecast precision is important, the direction of a prediction is also crucial. The "hit ratio" indicates how often a model correctly predicts an increase or a decrease in the growth rate of any series. Although the CSI

11 Although real-time data would provide a better sense of the model's performance, this type of analysis is left for future work.

12 The first four weeks of the forecast cycle (i.e., before the start of the quarter being considered) are often called the forecasting period. Predictions made during the quarter (weeks 5 to 16) are referred to as "nowcasting," while those made after the end of the quarter (but before the quarterly data are released) are called "backcasting."

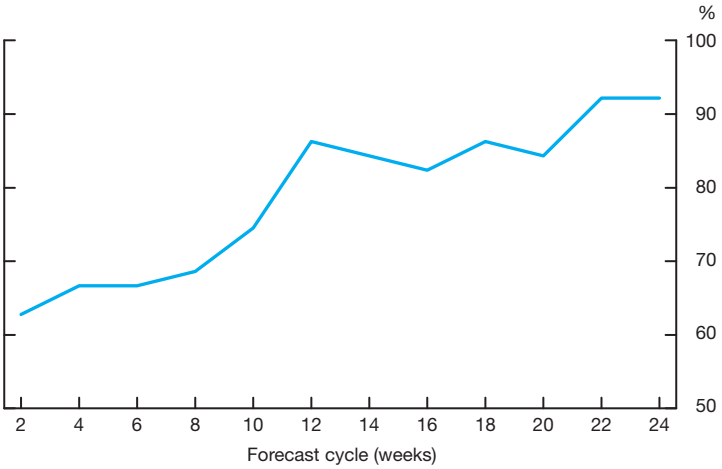
Chart 4: Forecast memory of CSI, 2000Q1–2012Q4

Ratio of the root-mean-square prediction error to the standard deviation of quarterly GDP growth



Source: Bank of Canada calculations

Chart 5: Hit ratio of CSI, 2000Q1–2012Q4



Source: Bank of Canada calculations

forecasts are not very accurate early in the forecasting cycle (**Chart 3**), their direction is correct more than 60 per cent of the time (**Chart 5**). Furthermore, as more data become available during the forecast cycle, CSI correctly predicts an increase or a decrease in real GDP growth with a hit ratio of about 90 per cent.

Conclusion

The main objective of CSI is to offer a data-intensive, judgment-free approach to short-term forecasting. CSI provides a way to extract information more systematically from some indicators that had been previously used with judgment to forecast GDP. A factor model can process a large number of indicators and, as implemented in CSI, is able to produce a new prediction of real GDP growth almost immediately following the latest release of data for an indicator. While these results are encouraging, current analysis should not rely mechanically on predictions from a single model. The Bank of Canada uses a wide range of models and information sources, as well as expert judgment, in producing its short-term forecasts. As such,

CSI is considered to be a good complement to other forecasting tools, providing valuable information about the direction of economic growth during the current quarter and the next. Further assessment of its real-time performance is needed, however, to better ascertain the model's full potential.

Appendix 1

CSI Indicators

Indicator	Source[a]	Frequency
1. Early estimate 1 (first month of GDP)†	STC	Quarterly
2. Early estimate 2 (second month of GDP)†	STC	Quarterly
3. Quarterly GDP†	STC	Quarterly
4. Momentum indicator†	STC	Quarterly
5. U.S. industrial production†	FED	Monthly
6. Total hours worked (Labour Force Survey)†	STC	Monthly
7. Canada's composite leading index†	STC and MLI	Monthly
8. U.S. retail sales†	USCB	Monthly
9. Global purchasing managers' index (manufacturing)††	J.P. Morgan	Monthly
10. Real commodity prices, excluding energy prices†††	BoC	Monthly
11. Terms of trade†	STC	Monthly
12. TSX index†††	STC	Monthly
13. Wholesale trade†	STC	Monthly
14. Consumer confidence††	CBoC	Monthly
15. Car sales†	STC	Monthly
16. Import volume (goods)†	STC	Monthly
17. Export volume (goods)†	STC	Monthly
18. Retail trade†	STC	Monthly
19. New manufacturing orders†	STC	Monthly
20. U.S. car sales†	WA	Monthly
21. Food services and drinking places sales†	STC	Monthly
22. Oil and gas extraction†	STC	Monthly
23. Bank of Canada *Business Outlook Survey*†† (average balance of opinion on past sales growth, future sales growth, investment in machinery and equipment, and output price pressures; some or significant difficulty in meeting demand; and labour shortages)	BoC	Quarterly
24. Railway carloadings†	STC	Monthly
25. Housing starts†	CMHC	Monthly
26. U.S. housing starts†	USCB	Monthly
27. Motor vehicle production†	WA	Monthly
28. Cement production†	STC	Monthly
29. Non-residential building construction†	STC	Quarterly
30. Total credit (household and business)†††	BoC	Monthly
31. Non-residential building permits†	STC	Monthly
32. Yield spreads on government bonds††† (Government of Canada bond yields: yield of a 5–year bond minus the yield of a 3-month treasury bill)	STC	Monthly

† Hard indicator †† Soft indicator ††† Financial indicator

a. The indicators used in Canada's Short-Term Indicator (CSI) model are taken from the following sources: Statistics Canada (STC), Bank of Canada (BoC), Canada Mortgage and Housing Corporation (CMHC), Conference Board of Canada (CBoC), Macdonald-Laurier Institute (MLI), WardsAuto (WA), United States Census Bureau (USCB), Board of Governors of the Federal Reserve System (FED) and J.P. Morgan.

Literature Cited

Armah, N. A. 2013. "Big Data Analysis: The Next Frontier." *Bank of Canada Review* (Summer): 32–39.

Boivin, J. and S. Ng. 2005. "Understanding and Comparing Factor-Based Forecasts." *International Journal of Central Banking* 1 (3): 117–51.

———. 2006. "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics* 132 (1): 169–94.

Burns, A. F. and W. C. Mitchell. 1946. *Measuring Business Cycles*. Cambridge, MA: National Bureau of Economic Research.

Camacho, M. and G. Perez-Quiros. 2010. "Introducing the Euro-Sting: Short-Term Indicator of Euro Area Growth." *Journal of Applied Econometrics* 25 (4): 663–94.

Coletti, D. and S. Kozicki. 2013. "Introduction: Tools for Current Analysis at the Bank of Canada." *Bank of Canada Review* (Summer): 1–2.

Geweke, J. 1977. "The Dynamic Factor Analysis of Economic Time Series Models." In *Latent Variables in Socioeconomic Models*, 365–83. Edited by D. Aigner and A. Goldberger. Amsterdam: North Holland.

Granziera, E., C. Luu and P. St-Amant. 2013. "The Accuracy of Short-Term Forecast Combinations." *Bank of Canada Review* (Summer): 13–21.

Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* 82 (1): 35–45.

Lucas, R. E. 1977. "Understanding Business Cycles." *Carnegie-Rochester Conference Series on Public Policy* 5 (1): 7–29.

Morel, L. 2012. "A Foreign Activity Measure for Predicting Canadian Exports." Bank of Canada Discussion Paper No. 2012-1.

Pearson, K. 1901. "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* 2 (6): 559–72.

Sargent, T. J. and C. A. Sims. 1977. "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory." Federal Reserve Bank of Minneapolis Working Paper No. 55.

Spearman, C. 1904. "The Proof and Measurement of Association Between Two Things." *American Journal of Psychology* 15 (1): 72–101.

Statistics Canada. 2011. "The Relationship Between Monthly, Quarterly, and Annual Growth Rates." *Canadian Economic Observer* 24 (6), Cat. No. 11-010-X (June).

Stock, J. H. and M. W. Watson. 1991. "A Probability Model of the Coincident Economic Indicators." In *Leading Economic Indicators: New Approaches and Forecasting Records*, 63–90. Edited by K. Lahiri and G. H. Moore. Cambridge, U.K.: Cambridge University Press.