Extracting Information from the *Business Outlook Survey*: A Principal-Component Approach

Lise Pichette and Lori Rennison, Canadian Economic Analysis

- *Since 1997, the* Business Outlook Survey *(BOS) has provided the Bank of Canada with valuable and timely information for the conduct of monetary policy.*
- *Recent work using principal-component analysis to extract information common to the BOS indicators is reviewed, as is testing of the out-of-sample forecasting performance of various models using this information.*
- *Results suggest that summarizing the common movements among BOS indicators may provide useful information for forecasting near-term growth in business investment. For growth in real gross domestic product, however, the survey's balance of opinion on future sales growth appears to be more informative.*

Since the autumn of 1997, the Bank of Canada's regional offices have conducted quarterly consultations with businesses across Canada. These consultations, referred to as the *Business Outlook Survey* (BOS), are structured around a questionnaire that covers topics of importance to the Bank, notably business activity, pressures on production capacity, prices and inflation, and credit conditions.[1] The responses to these qualitative questions (e.g., whether sales volumes will increase at a greater, lesser or the same rate over the next 12 months as over the past 12 months), together with the explanations that accompany them, allow senior economics staff at the Bank's regional offices to provide a macro-level assessment of the economy using the various demand- and supply-side signals from the survey. This assessment supplements the more quantitative approaches used by the Bank to evaluate the economic situation and outlook by providing insights into what businesses are seeing and planning.[2]

A key advantage of the BOS is its timeliness. Consultations take place around the middle of each quarter, and the results are published the week before the Bank's next fixed date for announcing monetary policy decisions. This is well ahead of the release of the National Income and Expenditure Accounts for that quarter. The high demand for timely information about the economy has led the survey results to become a well-monitored

1 For a detailed description of the survey, see Martin (2004) and *Backgrounder on Questions in the* Business Outlook Survey *Concerning Past Sales and Credit Conditions* (<http://www.bankofcanada.ca/wp-content/uploads/2011/07/bos_backgrounder_jan2008.pdf>).

2 See Macklem (2002) and Jenkins and Longworth (2002) for a description of how the BOS fits into the Bank's monetary policy decision-making process.

Table 1: *Business Outlook Survey* indicators

Survey question	Horizon	Type of signal
Balance of opinion[a] on past sales growth	Past 12 months	Demand-side
Balance of opinion on future sales growth	Next 12 months	Demand-side
Balance of opinion on investment in machinery and equipment	Next 12 months	Demand-side
Balance of opinion on employment	Next 12 months	Supply-side, indirect demand-side
Ability to meet an unexpected increase in demand[b]	Current	Supply-side, cost structure
Percentage of firms facing labour shortages	Current	Supply-side, cost structure
Balance of opinion on input prices	Next 12 months	Supply-side, cost structure
Balance of opinion on output prices	Next 12 months	Supply-side, margins
Balance of opinion on credit conditions	Past 3 months	Financial markets, demand-side

a. Percentage of firms responding "greater" or "higher" minus percentage of firms reporting "lesser" or "lower"
b. Percentage of firms responding "some" or "significant" difficulty

information source for the press and the financial community since their public release in 2004. Information published in the BOS helps to refine the Bank's view on the economic outlook, and is often cited in the Bank of Canada's *Monetary Policy Report*.

Our understanding of the statistical precision of individual BOS indicators has been strengthened by the research of de Munnik, Dupuis and Illing (2009) and de Munnik (2010) (**Box 1**). An evaluation of the survey's ability to predict economic variables, however, has been limited by its relatively short sample period. The initial assessment of the information content of the BOS by Martin (2004) was based on graphical and correlation analysis of 24 observations. Since then, the sample period has become long enough to include at least one full economic cycle with periods of expansion and slowdown, as well as a steep recession and a recovery, providing richer information for empirical analysis.

This article summarizes recent work that contributes to our understanding of the survey's information content by extending the early work by Martin (2004) in two key ways. First, since all BOS questions are designed to capture some aspect of economic activity and, therefore, are interrelated, principal-component analysis was used to extract the common underlying variations among the indicators. Second, the information content of these common movements was assessed, using regression analysis and a forecasting assessment. The first test of this measure's usefulness was whether it can help predict growth in real gross domestic product (GDP) and, if yes, whether it outperforms the survey question on future sales expectations—the question most closely tied to measuring GDP. The second test was whether the common movements of indicators provide clearer signals for any one component of economic activity. Given that the BOS is a survey of firms, business investment was a natural element to consider.

Extracting Common Information from BOS Indicators

For analytical and communication purposes, the responses to most BOS questions are expressed in terms of a balance of opinion or as a proportion of respondents (for questions on labour shortages and firms' ability to meet demand). This practice has been useful for interpreting the survey results with respect to specific aspects of economic activity. Each quarter, staff in the Bank's regional offices assess and amalgamate these signals from the survey regarding aggregate demand, aggregate supply and financial markets (**Table 1**), informed by the broader discussions that take place with firms during the interviews.

To statistically evaluate the survey's information content and its ability to predict real economic variables, common movements from the various BOS indicators were extracted using a data-reduction technique—principal-component analysis (PCA). In addition to capturing a common source of variation, using this shared underlying component to represent the fuller BOS data set in a forecasting assessment is an appealing alternative to using individual indicators, since it conserves degrees of freedom and lessens concerns about issues of multicollinearity.

Box 1: Statistical Accuracy of the *Business Outlook Survey*

Martin (2004) reports that the statistical properties of the *Business Outlook Survey* (BOS) are difficult to determine, given the small sample size of 100 firms and the non-random, quota-sampling approach, which involves setting objectives for the number of firms selected by region, industry and size in order to be representative of the Canadian economy. De Munnik, Dupuis and Illing (2009) and de Munnik (2010) have made significant progress in dealing with this issue.

Using a Monte Carlo simulation framework, de Munnik, Dupuis and Illing (2009) construct an artificial data set of firms and their responses and estimate the impact of the Bank's non-random sampling on the accuracy and coverage of the survey.[1] More specifically, they develop a method for modelling a complex, non-random sampling process and for computing relevant measures of the confidence intervals. This allows them to replicate the survey's firm-selection process. **Table 1-A** shows how each quota or constraint affects the estimate of the population parameter compared with the simple case of random sampling. When investigated individually, only the quotas for industry and firm size are found to widen the confidence intervals (rows 3 and 4). Results for the fully constrained model (row 8), however, show no evidence that the Bank's firm-selection process produces significantly biased estimates and/or wider confidence intervals than random selection. In other words, although the quota constraints result in biases on the parameter estimates when controlled individually, these biases are small and appear to be largely offset when the model is calibrated using average historical responses. With respect to survey coverage, the authors find that the BOS method of firm selection restricts the survey sample but does not create bias in the estimate.

De Munnik (2010) extends this analysis by outlining the statistical properties of questions expressed as population proportion versus balance of opinion, and demonstrates how the design of the question affects the calculation of the confidence intervals. He also shows that the confidence bands around both types of question can change from survey to survey when the underlying distribution of responses becomes more or less concentrated in particular categories (such as "higher," "the same" or "lower"). In particular, he illustrates that the confidence intervals around balance-of-opinion questions are larger when there is more dispersion in the responses.

These studies have improved understanding of the survey's statistical accuracy and, together with the qualitative stories that accompany the responses, allow staff at the Bank's regional offices to better interpret and describe survey results from quarter to quarter.

Table 1-A: Comparison of simulation results

Selection model	Bias vs. "pseudo-population"	Confidence interval 95% (66%)
1. Random sample	0.06	16.6 (8.2)
2. Regional quota	2.00	16.6 (8.2)
3. Industry quota	-2.07	17.5 (8.6)
4. Firm-size quota	-2.78	17.7 (8.7)
5. Rotation constraint	0.17	16.7 (8.2)
6. Familiarity constraint	-0.23	17.0 (8.4)
7. Non-response constraint	-0.10	16.7 (8.4)
8. Fully constrained model	-0.23	16.8 (8.3)

1 De Munnik, Dupuis and Illing (2009) do not analyze the coverage of the survey, but a revised version of the paper that includes this analysis is available upon request.

Box 2: Technical Details of Principal-Component Analysis

The first principal component is obtained by maximizing its contribution to the variance of a set of p variables (x). It is expressed as follows:

$$\propto_1' x = \propto_{11} x_1 + \propto_{12} x_2 + \cdots + \propto_{1p} x_p =$$

$$\sum_{j=1}^{p} \propto_{1p} x_p,$$

where x has a covariance matrix Σ. To derive this first principal component, the vector α_1 maximizes $var(\alpha_1' x) = \alpha_1' \Sigma \alpha_1$ subject to $\alpha_1' \alpha_1 = 1$. Using the technique of Lagrange multipliers:

$$\max_{\alpha_1} \alpha_1' \Sigma \alpha_1 - \lambda(\alpha_1' \alpha_1 - 1).$$

The first-order condition is

$$(\Sigma - \lambda I_p)\alpha_1 = 0,$$

where I_p is the $(p \times p)$ identity matrix, λ is an eigenvalue of Σ, and α_1 is the corresponding eigenvector. To maximize $var(\alpha_1' x)$, the vector of coefficients in the first principal component, α_1, is the eigenvector associated to the largest eigenvalue:

$$var(\alpha_1' x) = \alpha_1' \Sigma \alpha_1 = \alpha_1' \lambda \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda.$$

The *k*th principal component is derived by maximizing $var(\alpha_k' x)$ subject to $\alpha_k' \alpha_k = 1$ and $\text{cov}(\alpha_k' x, \alpha_l' x) = 0$ for all $k \neq l$.

Chamberlin (2007) makes similar use of PCA on survey measures of economic activity in the United Kingdom to develop an alternative forecast of GDP.

As explained by Jolliffe (2002, 1), "The central idea of principal component analysis is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set." This method generates a new set of variables—principal components—that are linear combinations of the original variables. Principal components are artificial variables that account for most of the variance in the observed variables contained in the data set, and they are all orthogonal to each other. (See **Box 2** for a brief technical description of PCA.)

By definition, the number of principal components that can be found is the same as the number of variables considered, but, in general, most of the variance in the data set will be accounted for by fewer principal components. For this analysis, only the first principal component (PC1) was retained. This underlying variable was extracted from the nine published BOS indicators that pertain to firms' views on their own business situation or plans (**Table 1**).[3]

While data for most indicators are available from 1998Q3, the questions on credit conditions and on the ability of firms to meet an unexpected increase in demand were added more recently. Extracting information from all nine indicators therefore required the length of the sample be limited to 2001Q4 to 2011Q2.[4, 5]

Correlation with Economic Data

Results from the correlation analysis of PC1 and real economic variables are shown in **Table 2**. For each economic variable, PC1 is compared with the survey indicator that most closely corresponds to the same economic concept.

Panel A presents the peak correlations with real GDP growth. The correlation results are moderate to moderately strong for both PC1 and the most relevant

3 This analysis excluded the question relating to firms' expectations regarding consumer price index inflation over the next two years.

4 The question on the ability to meet demand was added in 1999Q3 and that on credit conditions in 2000Q3. The question on credit was modified in 2001Q4 to reduce its horizon from the past 12 months to the past 3 months.

5 While the information content analyzed used the sample spanning 2001Q4 to 2011Q2, the sample was also extended back to 2000Q3, using the first formulation of the credit-conditions question. Since results were very similar, this extended sample is used for the charts to illustrate the behaviour of the underlying variable through the 2001 slowdown.

Table 2: Peak correlations
Sample: 2001Q4 to 2011Q2

BOS indicators	Quarter-over-quarter			Year-over-year		
	Panel A: Real GDP growth					
PC1	0.54 (t - 1)	0.56 (t)	0.50 (t + 1)	0.63 (t)	0.73 (t + 1)	0.63 (t + 2)
Balance of opinion on future sales	0.49 (t)	0.69 (t + 1)	0.54 (t + 2)	0.61 (t + 2)	0.68 (t + 3)	0.55 (t + 4)
	Panel B: Real consumption growth					
PC1	0.51 (t - 2)	0.45 (t - 1)	0.31 (t)	0.63 (t)	0.64 (t + 1)	0.47 (t + 2)
Balance of opinion on future sales	0.11 (t - 1)	0.32 (t)	0.27 (t + 1)	0.23 (t + 1)	0.33 (t + 2)	0.31 (t + 3)
	Panel C: Real business-investment growth					
PC1	0.61 (t - 1)	0.83 (t)	0.73 (t + 1)	0.79 (t + 1)	0.79 (t + 2)	0.63 (t + 3)
Balance of opinion on investment in machinery and equipment	0.59 (t)	0.67 (t + 1)	0.57 (t + 2)	0.55 (t)	0.70 (t + 1)	0.65 (t + 3)

indicator—the balance of opinion on future sales growth.[6] While the question on future sales asks firms to characterize the expected change in sales growth over the next 12 months (i.e., momentum as opposed to growth), it is reasonable to expect that firms' predictions about future momentum could also contain information about current or near-term growth. The strongest correlation between PC1 and quarterly GDP growth is reached contemporaneously (in quarter t), while, for the balance of opinion on future sales growth, the highest correlation coefficient is obtained one quarter ahead (t + 1). Given the 12-month horizon of many BOS questions, it is also worth noting that the correlation for the question on sales outlook with the year-over-year growth of real GDP reaches a peak three quarters ahead, while it is only one quarter ahead for PC1. **Chart 1** plots the balance of opinion on future sales and PC1, together with real GDP growth. It shows that both the individual question and the common factor seem to track developments in aggregate economic activity relatively well.

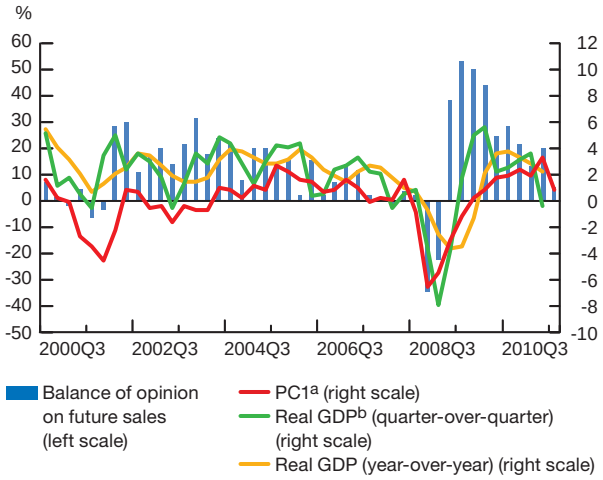
Panels B and C of **Table 2** report the correlation results for specific components of the GDP: consumption and business investment. The results for quarterly growth in real consumption are weak or moderate, while those for growth in real business investment are moderately strong to strong. The weak correlation with consumption may be partly explained by the survey sample, which comprises the business sector rather than the consumer sector, and even within the business sector, not only firms selling to consumers, but also firms selling to other businesses or exporting. The underlying variable, PC1, has a higher correlation with business investment than the survey question on the expected direction of change in investment in machinery and equipment over the next 12 months. This suggests that extracting the common movements from all the survey questions might lead to a better indicator for quarterly growth in business investment than this single question on investment intentions.

The strong correlation between PC1 and business investment is interesting, since very few indicators of investment are available

Chart 2 shows that PC1 closely tracks fluctuations in business investment over the sample period. The strong correlation between this underlying variable derived from the BOS results and business investment is interesting, since very few indicators of investment are available ahead of official statistical data. Moreover, the correlation coefficient continues to be moderately strong one quarter ahead, suggesting that the BOS also contains forward-looking information regarding business investment.

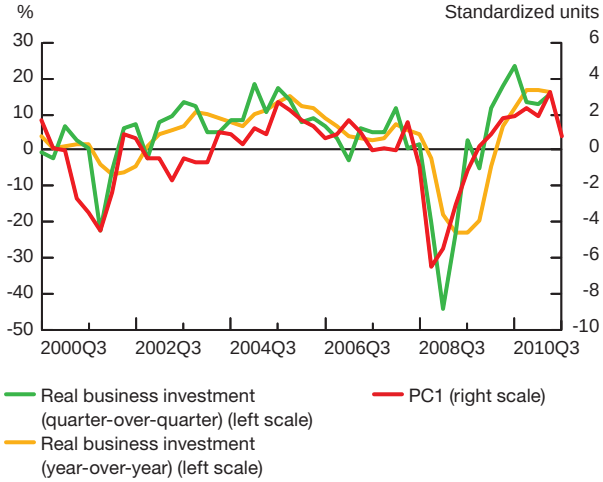
6 The scale used to evaluate the correlation coefficients is that used in Martin (2004): strong, > 0.80; moderately strong, 0.60 to 0.80; moderate, 0.40 to 0.60; weak, 0.20 to 0.40; insignificant, < 0.20.

Chart 1: Aggregate economic activity and BOS indicators



a. PC1 is measured in standardized units.
b. Real GDP is measured as a percentage.
Sources: Statistics Canada, Bank of Canada and authors' calculations
Last observation: 2011Q3

Chart 2: Business investment and the underlying BOS variable



Sources: Statistics Canada and authors' calculations
Last observation: 2011Q3

Forecasting Assessment

Regression analysis and a forecasting exercise were carried out to evaluate (i) whether the underlying variable extracted from the BOS indicators using PCA can provide information beyond that contained in the past values of the economic variables, and (ii) whether it provides more information than is contained in the answers to the individual survey questions on future sales growth and investment intentions.

Various simple models were examined and compared based on the root mean square errors (RMSE) computed using a series of one-step-ahead forecasts for each equation.[7] Specifically, each equation is estimated for a sample spanning the period 2001Q4 to 2006Q1, and a forecast is produced for 2006Q2.[8] One observation is then added to the estimation period for the next-quarter forecast, and this is repeated up to 2011Q2. The ratio of the RMSE for each equation, relative to a benchmark case that includes only the lags of the dependent variable, is reported. For example, an RMSE ratio below one implies that the inclusion of the common component obtained from BOS results improves the forecast derived from an equation that takes into account only the latest information on the variable of interest.

Table 3 summarizes the estimation results for quarterly real GDP growth. The first five rows report results for three different specifications (equations 1 to 3) estimated on the full sample (2001Q4 to 2011Q2). The coefficient on PC1 (equation 2) is significant, and the adjusted R^2 increases compared with equation 1, which includes only lagged GDP. PC1 is incorporated only at time t in this equation, because lags were not statistically significant. Nonetheless, since data from the BOS are available almost two months before the release of the national accounts, the results can be useful for forecasting. The results for equation 3, however, indicate that the balance of opinion on future sales growth remains a better indicator than the underlying variable, with the adjusted R^2 increasing to 0.56. The balance of opinion on future sales is significant only contemporaneously despite correlation results that suggested that expectations of future sales contained more forward-looking information.

The last row of **Table 3** reports the results of the out-of-sample forecast exercise. The RMSE ratios for both equations 2 and 3 are below one, indicating that the inclusion of information from the BOS improves the forecast from that of equation 1. However, the improvement is only marginal for PC1, and the difference between equations 2 and 3 is found to be significant, according to the Diebold-Mariano test. Thus, the underlying variable extracted from the set of BOS responses does not outperform the balance of opinion on future sales in forecasting real economic activity.

7 The prediction is for the current quarter before the release of the national accounts.
8 The data used in this exercise were the latest available, published on 31 August 2011.

Table 3: Estimation results for real GDP growth (quarter-over-quarter)
Sample: 2001Q4 to 2011Q2

Variables included	Equation 1	Equation 2	Equation 3
Constant	0.79 (1.84)[a]	1.11 (2.56)	-0.33 (-0.75)
GDP growth (*t* - 1)	0.60 (4.63)	0.42 (2.84)	0.58 (5.45)
PC1 (*t*)		0.41 (2.19)	
Balance of opinion on future sales (*t*)			0.07 (4.34)
Adjusted R^2	0.35	0.41	0.56
RMSE ratio	1.00	0.97	0.82

a. t-statistics are in parentheses.

Table 4: Estimation results for growth in real business investment (quarter-over-quarter)
Sample: 2001Q4 to 2011Q2

Variables included	Equation 1	Equation 2	Equation 3	Equation 4
Constant	1.76 (1.05)[a]	2.97 (2.51)	4.20 (3.71)	-2.36 (-1.22)
Growth in investment (*t* - 1)	0.66 (5.22)	0.26 (2.41)		0.40 (2.93)
PC1 (*t*)		4.16 (6.31)	3.72 (4.57)	
PC1 (*t* - 1)			1.82 (2.25)	
Balance of opinion on investment in machinery and equipment (*t*)				0.39 (3.32)
Adjusted R^2	0.41	0.71	0.71	0.54
RMSE ratio	1.00	0.70	0.71	0.89

a. *t*-statistics are in parentheses.

Table 4 reports the estimation results for growth in real business investment, using the same approach. In this case, the values of the adjusted R^2 and the RMSE ratio are quite impressive when the underlying variable extracted from the BOS results is included as an explanatory variable. Based on the estimates from equation 3, PC1 alone, without the lagged growth of business investment, produces results very similar to those of equation 2. While the survey question on investment intentions for machinery and equipment is found to have explanatory power (equation 4), results for equations 2 and 3 indicate that PC1 outperforms the survey question.[9] From these results, it appears that a measure of the underlying information from all BOS indicators provides more useful signals for monitoring the growth of near-term investment than the question regarding investment intentions for machinery and equipment over the next 12 months in isolation.

Robustness

Results for PCA, presented in the previous section, are relatively robust to weights. Indeed, alternative approaches to extracting common movements, including a simple average of the nine indicators and factor analysis, generated series that were highly correlated with PC1.[10] Moreover, real-time estimates of the first principal component were examined, since weights may fluctuate when the sample is changing. Sensitivity analysis suggests that, although the weights vary, the underlying variable extracted using PCA remains virtually the same.[11]

Discussion and Conclusions

The research reviewed here used principal-component analysis to evaluate the information content of the BOS. It also assessed the information content of the first principal component relative to that of individual survey questions in an out-of-sample forecasting exercise. This is the first empirical assessment of the BOS information content since the initial correlation analysis by Martin (2004), and it has provided several notable contributions.

First, the results suggest that the first principal component appears to be a useful indicator of economic activity, particularly for providing information on investment spending—a variable that is typically difficult to predict and for which there are very few indicators. This may not be surprising, since the BOS is a survey of firms, and all its questions provide some signals relating to the probability of investing. For instance, if the outlook of firms regarding sales, employment and investment improves, if more firms are operating at or above their production capacity, and if more firms report an easing in credit conditions, then it is reasonable to expect higher

9 The results from both equations 2 and 3 are statistically different from those of equation 4 at the 5 per cent significance level, according to the Diebold-Mariano test.
10 Correlation coefficients were greater than 0.98. (This is true for the series expressed both in levels or first differences.)
11 Results of this sensitivity analysis are available from the authors upon request.

investment activity. The outlook for prices can also play a role in firms' near-term investment spending; for example, if firms are expecting higher input costs and are therefore spending to become more efficient, or if higher prices are stimulating activity in particular sectors where investment projects become profitable (as was the case with the boom in commodity prices in the 2000s).

Second, this analysis found that the individual survey questions on future sales growth and intentions for investment in machinery and equipment provide useful information in a forecasting context for real GDP growth and the growth of business investment, respectively. The inclusion of these variables in their respective regressions improved upon a simple autoregressive model. In the case of business investment, however, the single question was found to be less informative than the measure of common movements.

This work contributes to our understanding of the survey's information content, but the reliability and robustness of the results will need to be evaluated over time as the sample period grows. Moreover, promising statistical assessments do not preclude careful examination of the movements in each BOS indicator every quarter, or the qualitative assessment of the messages that accompany firms' responses, both of which make a valuable contribution to monetary policy. Whether in terms of common movements or individual indicators, information gathered from business surveys is often best used with informed judgment rather than according to mechanical rules. The information obtained from individual survey indicators and from the qualitative assessment carried out by the Bank's regional offices remain important elements in BOS analysis. As emphasized in Martin (2004, 10), "The BOS interview format allows for a broader understanding of current business perceptions through confidential discussions with business representatives, which provide invaluable information that cannot be measured quantitatively."

Literature Cited

Chamberlin, G. 2007. "Forecasting GDP Using External Data Sources." *Economic & Labour Market Review* 1 (8): 18–23.

de Munnik, D. 2010. "Statistical Confidence Intervals for the Bank of Canada's Business Outlook Survey." Bank of Canada Discussion Paper No. 2010-7.

de Munnik, D., D. Dupuis and M. Illing. 2009. "Computing the Accuracy of Complex Non-Random Sampling Methods: The Case of the Bank of Canada's Business Outlook Survey." Bank of Canada Working Paper No. 2009-10.

Jenkins, P. and D. Longworth. 2002. "Monetary Policy and Uncertainty." *Bank of Canada Review* (Summer): 3–10.

Jolliffe, I. 2002. *Principal Component Analysis*. 2nd ed. New York: Springer.

Macklem, T. 2002. "Information and Analysis for Monetary Policy: Coming to a Decision." *Bank of Canada Review* (Summer): 11–18.

Martin, M. 2004. "The Bank of Canada's Business Outlook Survey." *Bank of Canada Review* (Spring): 3–18.