



CIRANO

Allier savoir et décision

A Large Canadian Database for Macroeconomic Analysis

OLIVIER FORTIN-GAGNON

MAXIME LEROUX

DALIBOR STEVANOVIC

STÉPHANE SURPRENANT

2018S-25
WORKING PAPER

CS

2018s-25

A Large Canadian Database for Macroeconomic Analysis

*Olivier Fortin-Gagnon, Maxime Leroux,
Dalibor Stevanovic, Stéphane Surprenant*

Série Scientifique
Scientific Series

Montréal
Août/August 2018

© 2018 Olivier Fortin-Gagnon, Maxime Leroux, Dalibor Stevanovic, Stéphane Surprenant. Tous droits réservés.
All rights reserved. Reproduction partielle permise avec citation du document source, incluant la notice ©. Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaires corporatifs

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Laurentienne
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Énergir
Hydro-Québec
Innovation, Sciences et Développement économique Canada
Intact Assurance
Investissements PSP
Ministère de l'Économie, de la Science et de l'Innovation
Ministère des Finances du Québec
Mouvement Desjardins
Power Corporation du Canada
Rio Tinto
Ville de Montréal

Partenaires universitaires

École de technologie supérieure
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 2292-0838 (en ligne)

A Large Canadian Database for Macroeconomic Analysis*

*Olivier Fortin-Gagnon [†], Maxime Leroux [‡],
Dalibor Stevanovic [§], Stéphane Surprenant ^{**}*

Résumé/Abstract

This paper describes a large-scale Canadian macroeconomic database in monthly frequency. The dataset contains hundreds of Canadian and provincial economic indicators observed from 1981. It is designed to be updated regularly through StatCan database and is publicly available. It relieves users to deal with data changes and methodological revisions. We show five useful features of the dataset for macroeconomic research. First, the factor structure explains a sizeable part of variation in Canadian and provincial aggregate series. Second, the dataset is useful to capture turning points of the Canadian business cycle. Third, the dataset has substantial predictive power when forecasting key macroeconomic indicators. Fourth, the panel can be used to construct measures of macroeconomic uncertainty. Fifth, the dataset can serve for structural analysis through the factor-augmented VAR model.

Mots clés/Keywords: Big Data; Factor Model; Forecasting; Structural Analysis

* The third author acknowledges financial support from the Fonds de recherche sur la société et la culture (Québec) and the Social Sciences and Humanities Research Council.

[†] Desjardins

[‡] Département des sciences économiques, Université du Québec à Montréal. 315, Ste-Catherine Est, Montréal, QC, H2X 3X2

[§] Corresponding Author: dstevanovic.econ@gmail.com. Département des sciences économiques, Université du Québec à Montréal. 315, Ste-Catherine Est, Montréal, QC, H2X 3X2

^{**} Département des sciences économiques, Université du Québec à Montréal. 315, Ste-Catherine Est, Montréal, QC, H2X 3X2

1 Introduction

Large datasets are now very popular in the empirical macroeconomic research. Stock and Watson (2002a) have initiated the breakthrough by providing the econometric theory and simple methods to deal with a large amount of time series. Stock and Watson (2002b) have shown the benefits in terms of macroeconomic forecasting, while Bernanke et al. (2005) have inspired the literature on impulse response analysis in the so-called data-rich environment. Since then, many theoretical and empirical improvements have been made, see Stock and Watson (2016) for a recent overview. The most of the literature is built on the US datasets. Therefore, McCracken and Ng (2016) proposed a standardized version of a large monthly US dataset that is regularly updated and publicly available at the FRED (Federal Reserve Economic Data) website. No such developments have been made with Canadian macroeconomic data, so the objective of this work is to fill the gap and provide a user-friendly version of a large Canadian dataset suitable for many types of macroeconomic research.

In this paper we construct a large-scale Canadian macroeconomic database in monthly frequency and show how it can be useful for empirical macroeconomic analysis with several illustrative examples. The dataset contains hundreds of Canadian and provincial economic indicators observed from 1981. It is designed to be updated regularly through the StatCan database and is publicly available. It relieves users to deal with data changes and methodological revisions. We provide a balanced and stationary panel suitable for work in business cycle fluctuations.

Early attempts to construct large Canadian macroeconomic datasets are Gosselin and Tkacz (2001) and Galbraith and Tkacz (2007). Boivin et al. (2010) updated and merged data from those previous studies yielding a panel that covered the period 1969 - 2008 and had 348 monthly and 87 quarterly series. Then, Bedock and Stevanovic (2017) constructed a new dataset of 124 monthly variables observed from 1981 to 2012. The selection of series was based on Canadian counterparts of US series in Bernanke et al. (2005) and constrained by availability of StatCan tables. More recently, Sties (2017) has built a large quarterly dataset. Stephen Gordon has also been updating some relevant Canadian indicators¹. Our data selection is mainly inspired by McCracken and Ng (2016) when it comes to major groups of economic variables. Given that Canada is a small open economy, we follow previous papers and add more series in the international trade group as one usually finds in the US applications.

We illustrate five useful features of this dataset for macroeconomic research. First, we show that our panel is likely to present a factor structure and that common factors explain a sizeable portion of variation in Canadian and provincial aggregate series. The principal component analysis of the dataset identifies few driving forces of the Canadian economy such as housing

¹See Project Link at <https://www.ecn.ulaval.ca/~sgor>.

and bond markets, production, wages, exchange rates and energy prices.

Second, the dataset is useful to capture turning points of the Canadian business cycle. Using Probit, Lasso and factor models we show that this dataset has substantial explanatory power in addition to the standard term spread predictor.

Third, the dataset has substantial predictive power when forecasting key macroeconomic indicators. Factor models and regularized complete subset regressions show good performance in forecasting real activity variables such as industrial production, employment and unemployment rate, as well as CPI and Core CPI inflation. Gains are still present, though less important, in case of the housing market.

Fourth, the panel can be used to construct measures of macroeconomic uncertainty. Using Jurado et al. (2015) methodology, we estimate the Canadian macroeconomic uncertainty which suggests two significant episodes of uncertainty: 1981 and 2009 recessions.

Fifth, the dataset can serve for structural impulse response analysis. We identify and estimate the effects of uncertainty shocks using the factor-augmented VAR methodology à la Boivin et al. (2018) with our measure of macroeconomic uncertainty. The results suggest that an increase in uncertainty is associated with a decline in real activity and financial markets. The findings are in line with the literature using US data.

The rest of the paper is organized as follows. Section 2 describes the construction of datasets and performs the factor analysis. Section 3 shows the informational content of this dataset in detecting recession dates. In Section 4 we conduct a pseudo-out-of-sample forecasting exercise to test the capability of the dataset to help predicting main Canadian macroeconomic variables. The measures of macroeconomic uncertainty are proposed in Section 5. Section 6 performs an impulse response analysis through. Section 7 concludes.

2 Datasets

In this section we start by describing the construction of the dataset, and in particular how we dealt with several issues related to availability and statistical properties of the data. Then, we explore the factor structure of this dataset.

2.1 Construction of datasets

The Canadian monthly database comprises eight different groups of variables: production, labour, housing, manufacturers' inventories and orders, money and credit, international trade, prices and stock markets. Whenever available, we included regional data covering the Atlantic provinces, Québec, Ontario, the Prairies and British Columbia, as well as provincial data. Thus, a strictly Canadian, as well as an extended version including provincial data are to be made

public. The complete list of series is available in the data appendix. Note that we decided to include a lot of housing series following Leamer (2015) who argue that housing cycle is leading the business cycle.

In building this database, we encountered several problems. Some tables have unfortunately been discontinued and the new tables seldom go sufficiently far back in time to afford us a sizeable time frame. Therefore, we combined old and new time series to cope with this problem. This happened with data on production, housing, orders and import and export. For instance, GDP data for the period starting in January 1981 and ending today is split across two tables: 379-0027, going from 1981/01 to 2012/01 and 379-0031, starting only on 1997/01. For all combined series, we chose to adjust the level of the old series and use the earliest shared period as the junction point so as to leave all the newest observations intact. The adjustment factor is computed at the junction point in all instances.

For imports and export series, there usually was a need to aggregate old series before splicing since old and new trade data do not share a common classification system. The example provided below of exported consumption goods aggregated section 2 data on food, feed, beverages and tobacco, major group 4.23 on textile fabricated materials and major group 5.11 on other consumption goods to approximate the consumer good class of the North American Product Classification System (NAPCS).

As is evident from the examples provided in Figure 1, viewing the old time series as noisy indexes of new time series seems justified by the high correlations in the overlapping periods. Note that figures for all instances of splicing are systematically produced by the accompanying code.

Another problem concerns the seasonal behaviour of most unemployment duration series. They were not readily available in a seasonally adjusted format. To deal with this, we opted to simply apply a 72 months moving average filter:

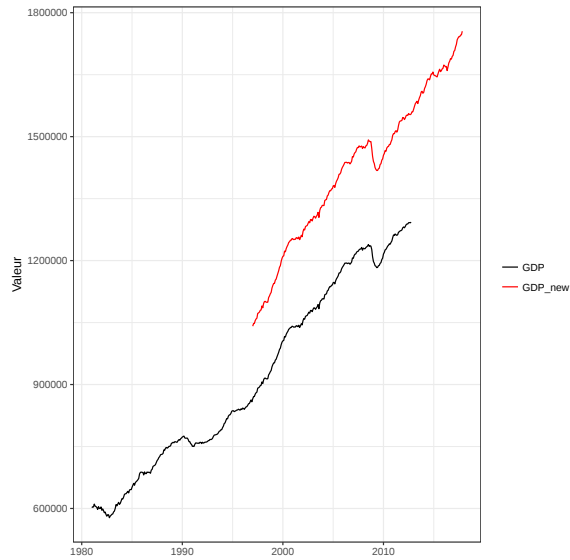
$$y_t^{adj} = \frac{1}{72} \sum_{j=0}^{71} y_t^{raw}. \quad (1)$$

Figure 2 gives an example of raw and seasonally adjusted series of unemployment duration growth.

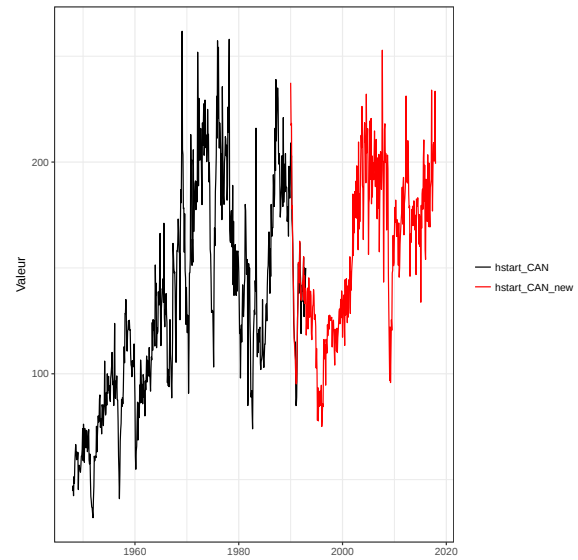
Most of the series included in the database must be transformed to induce a (weakly) stationary behaviour. We roughly follow McCracken and Ng (2016): most I(1) series are transformed in the first difference of logarithms; a first difference of levels is applied to unemployment rate and interest rates; first difference of logarithms is used for all price indexes; and housing data is featured in logarithms. The transformation codes are reported in data appendix.

Our last concern is to balance the resulting panel. Some series have missing observations.

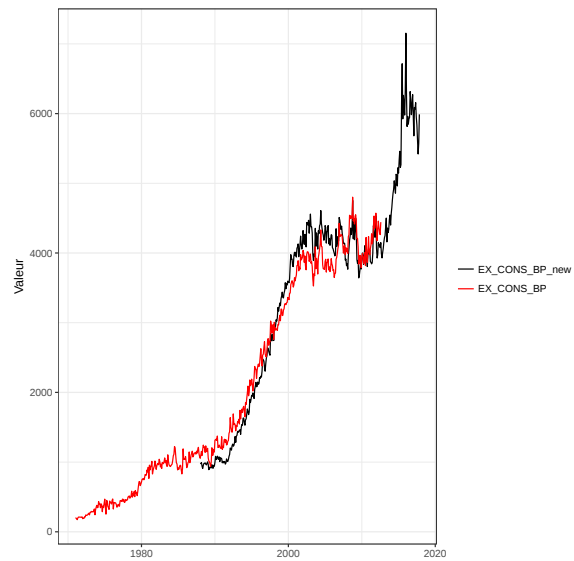
Figure 1: Examples of data splicing



(a) Gross domestic product (Canada)

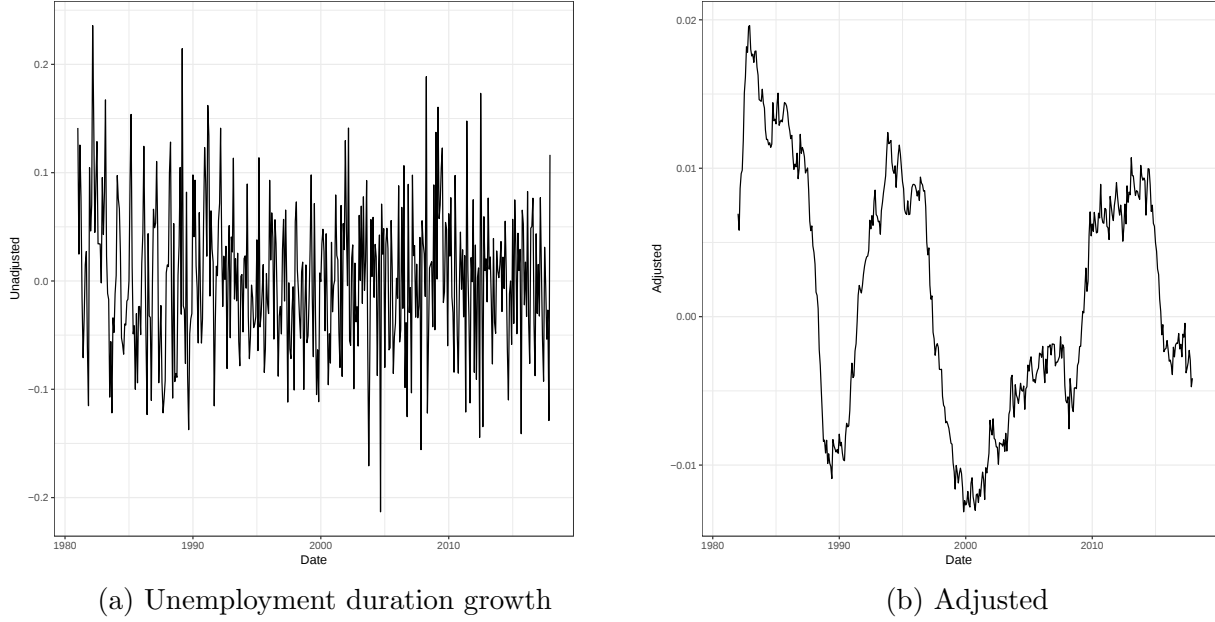


(b) Housing starts (units, Canada)



(c) Exports of consumer goods (Canada)

Figure 2: Example of seasonal adjustment



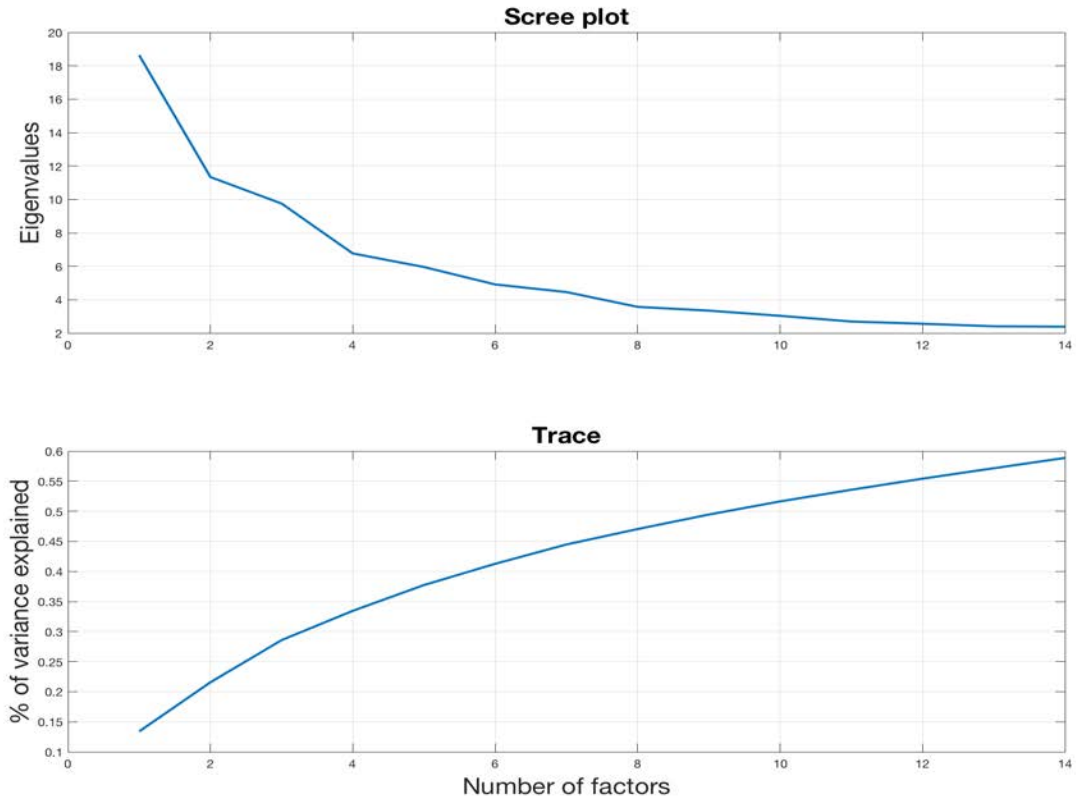
We opted to apply an expectation-maximization type of algorithm by assuming a factor model to fill in the blanks. Using the transformed data resulting from the previous steps, all variables are demeaned and reduced by their respective standard deviation. We initialize the algorithm by replacing missing observations with their unconditional mean and then proceed to estimate a factor model by principal component. The fitted values of this model are used to replace missing observations. The algorithm alternates between replacement and estimation until convergence. Examples of missing values include export and import series since the old tables went back only to 1988/01.

2.2 Number of Factors

Estimating the number of factors is an empirical challenge. Usually the first step is to plot the eigenvalues of the correlation matrix of data (scree plot) as well as the average explanatory power of consecutive principal components (trace). These are reported in Figure 3. The pictures are typical for macroeconomic panels. There is no clear cut separation among eigenvalues, except after the first, and the explanatory power grows slowly with the number of factors. However, we remark that 9 principal components explain 50% of variance of all 139 variables, which is quite satisfactory.

Many statistical decision procedures have been proposed to pin down the number of factors but are known to be unstable, see Mao Takongmo and Stevanovic (2015). Table 1 reports the number of factors estimated by the following methods: (BN02) Bai and Ng (2002) IC_{p2}

Figure 3: Eigenvalues and explanatory power of factors



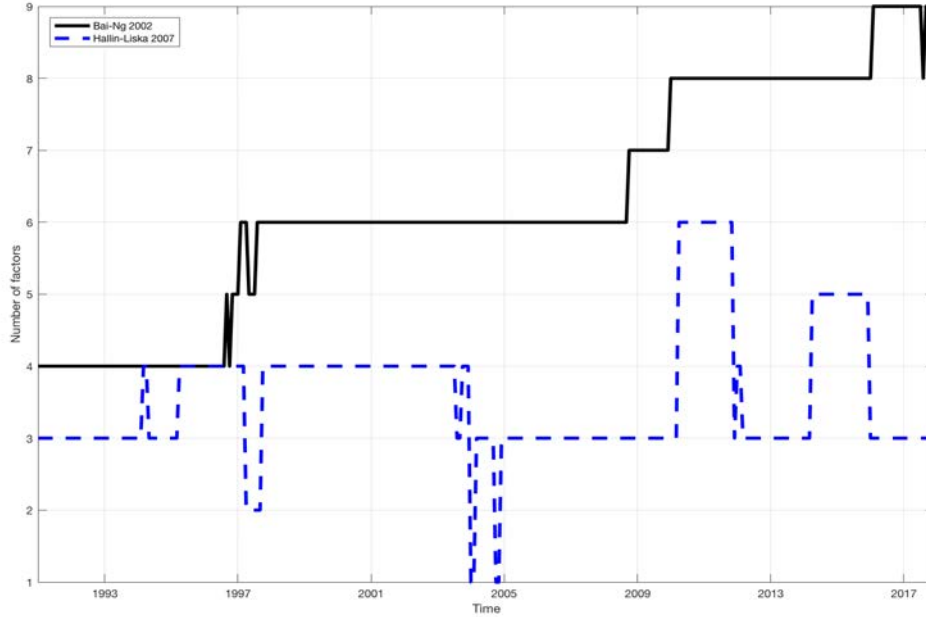
information criterion; (ABC) modified version of (BN02) by Alessi et al. (2010); (ON) Onatski (2010) test based on the empirical distribution of eigenvalues; (AH) Ahn and Horenstein (2013) eigenvalue ratio test; (HL) Hallin and Liska (2007) test for the number of dynamic factors; (BN07) Bai and Ng (2007) test for the number of dynamic factors; and finally (AW) Amengual and Watson (2007) information criterion for the number of dynamic factors. (ON) and (AH) are known to be very conservative and they indeed identify only one source of common variation. (BN02) and (ABC) suggest 9 static factors for the aggregate panel and 7 in the case of the panel including additional regional series. The number of factors is estimated between 6 and 9 according to (BN07) and (AW), but only 3 and 6 following Hallin and Liska (2007) procedure.

It is also common in the literature to verify the stability of the factor structure over time, in terms of the number of factors. Figure 4 plots the number of factors that has been estimated recursively by Bai and Ng (2002) and Hallin and Liska (2007) methods. We observe that the number of static factors is increasing since 1990, while the number of dynamic factors is rather stable around 3. Many explanations on the time-varying nature of the number of static factors are plausible: structural changes in terms of the correlation structure, presence of group-specific factors, sensitivity of the Bai and Ng (2002) procedure, and so on. We are not investigating

Table 1: Estimating the number of factors in CAN_MD

	Canada	Canada and regions
BN02	9	7
ABC	9	7
ON	1	1
AH	1	1
HL	3	6
BN07	7	9
AW	6	6

Figure 4: Number of factors over time



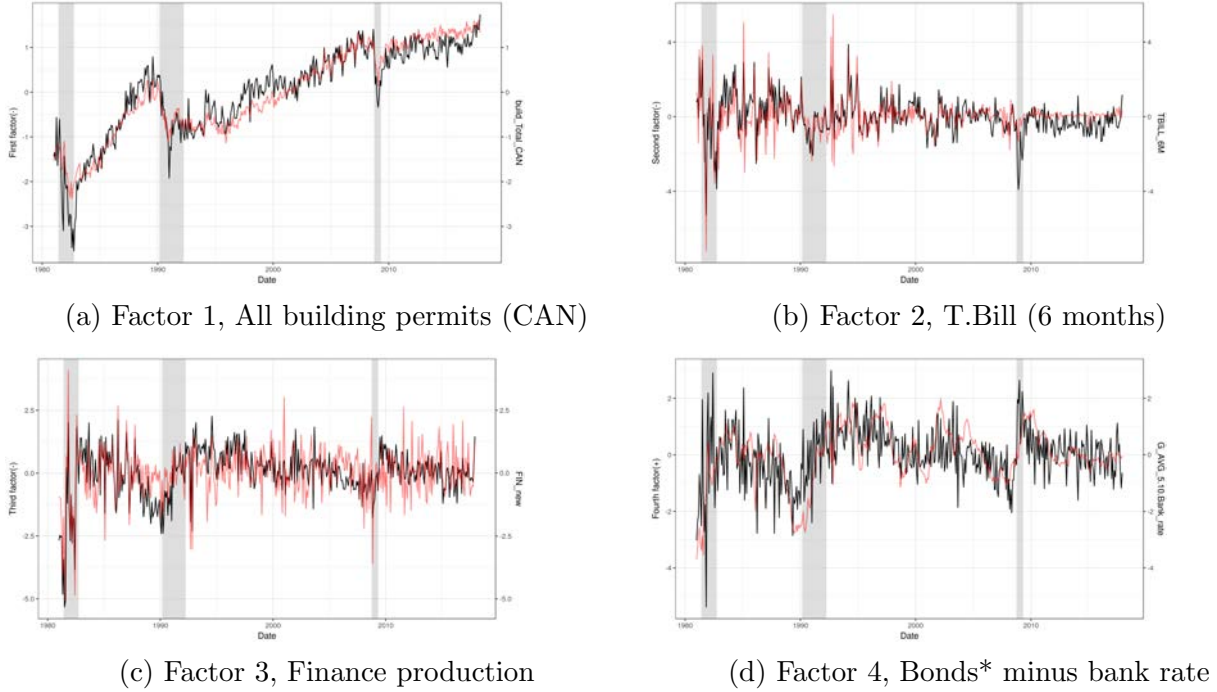
those possibilities but practitioners should be aware of this instability.

2.3 Estimated Factors

The factors estimated over the full sample by principal components are depicted in Figures 5 and 6 alongside their main series. The first factor seems to closely track the evolution of building permits in Canada, capturing much of the movements related to business cycle frequencies. Both variable have a slight upward trend.

The variable best explained by the second factor is the yield on 6-month treasury bills. This variable seems to dominate the movement in the factor until the last crisis when monetary policy begun holding rates very low, reducing its volatility. Production in the finance, real

Figure 5: Factors 1 to 4 and their main series

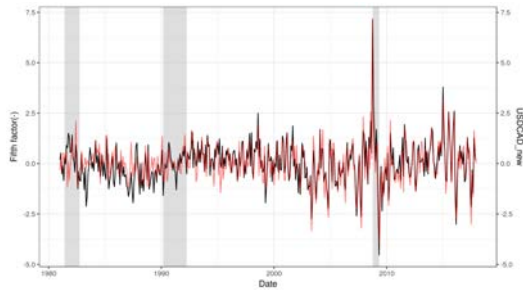


Notes: Factors are displayed in black and their main components in red. Factors have been estimated over the full sample and the chosen rotation is indicated by (+) or (-). Factors and series have been reduced by their respective standard deviation. (*): Average yield on governmental bonds of 5 to 10 year maturity.

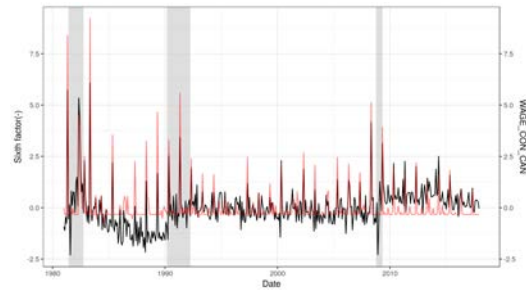
estate and insurance sectors are closely related to the third factor. Same for the late 1980s, both series are very similar. The fourth factor and eighth factors capture variations in the slope of the yield curve at two different points, the eighth factor leaning on more heavily on shorter maturities. Unsurprisingly, that factor becomes less variable after last crisis due to the aforementioned actions by the Bank of Canada. USD to CAD exchange rate movements seem to dominate the fifth factor, much like energy prices, especially from 2000 onwards, dominates the seventh factor. As the main series of the seventh factor shows, energy prices became much more volatile in the last two decades, hence the closer tracking in the later half of the sample. The sixth factor incorporates information about wages in Canada through the construction wage index. This variable rises in steps, hence the large peaks, which are well integrated in the behaviour of the corresponding factor.

Here, we study the factor interpretation through time by estimating the factor model recursively. Using the matrix $\{X_t\}_{t=1981M1}^{\tau}$, at each date $\tau = 1990M12, 1991M1, \dots, 2017M2$, we estimate factors and loadings by principal components. Squared loadings can be used as a measure of how heavily a series is weighted in the composition of each estimated factor and this procedure gave us a vector for all series at each date τ . The results of this procedure form the basis of the heatmaps shown in Figures 7 and 8. For convenience, variables are grouped

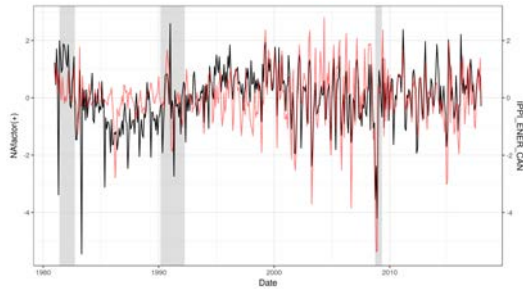
Figure 6: Factors 5 to 8 and their main series



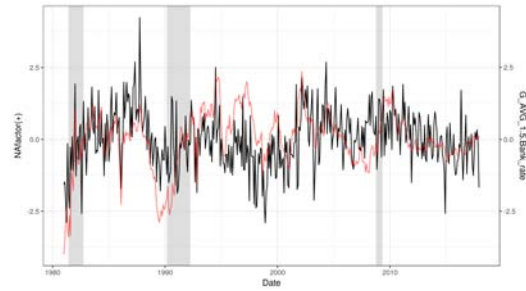
(a) Factor 5, USDCAD



(b) Factor 6, Construction wage (CAN)



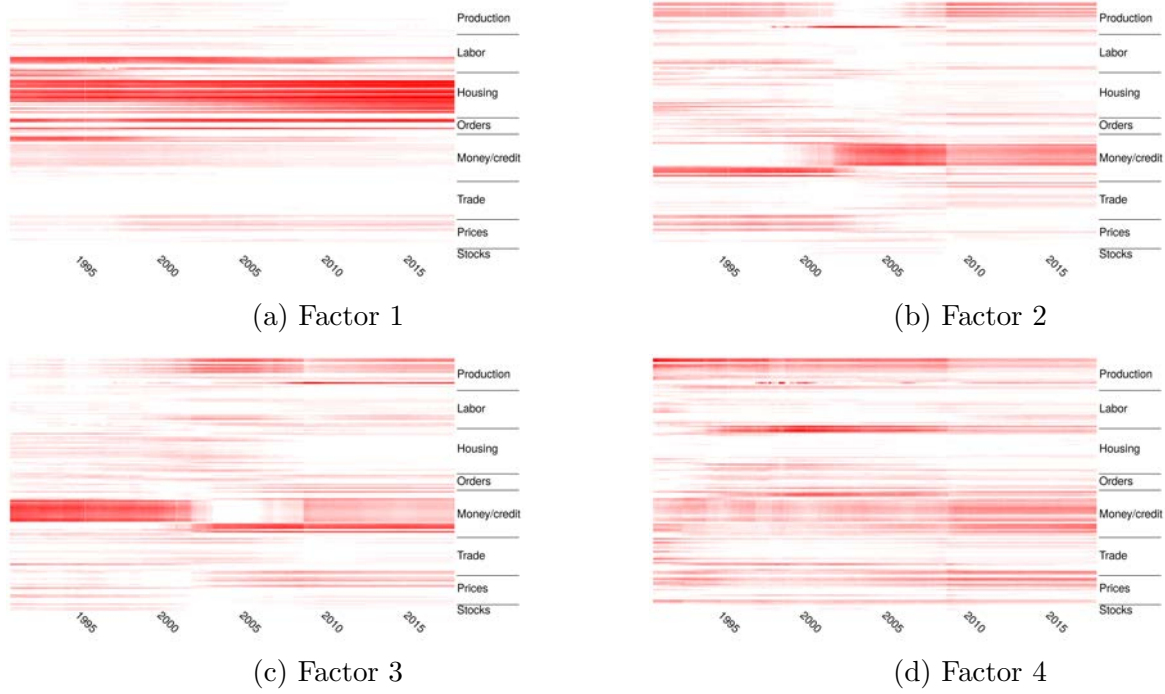
(c) Factor 7, IPPI Energy (CAN)



(d) Factor 8, Bonds** minus bank rate

Notes: Factors are displayed in black and their main components in red. Factors have been estimated over the full sample and the chosen rotation is indicated by (+) or (-). Factors and series have been reduced by their respective standard deviation. (**): Average yield on governmental bonds of 1 to 5 year maturity.

Figure 7: Heatmaps for factors 1 to 4



Notes: Factors and loadings are estimated recursively using an expanding window. Displayed shades of red capture squared loadings.

in categories, the exact composition of which are given in the data appendix. Tables 2 and 3 offer a more granular look in the interpretation and stability of factors, reporting the top ten series in terms of average squared loadings over subperiods. The subperiods have been chosen to match visual changes in some of the heatmaps, facilitating the parallel between the two.

The first factor weighs heavily on housing, manufacturing orders and to a smaller extent labour market variables. The factor appears overall very stable and this can be confirmed by the ranking of series in three selected subperiods reproduced in Table 2. The top three series, total building permits in different regions, are preserved throughout the whole sample, as well as their order. We also notice the fading of labour series from the factor by noting that the two unemployment series disappear from the top ten in the later two subperiods.

The heatmaps of factors two and three suggest a near inversion of the order of factors between 2000 and 2005. Recall that principal component orders factors by their share in explained variance for the whole database, hence this suggests that the type of dominant variables changed in the mid-2000 when compared to prior or later periods. Although these factors are more diffuse than the first, we can see that one captures a combination of production, monetary aggregates, credit measures and some prices, while the other is more heavily centred on interest rates and their differentials. This analysis is confirmed in Table 2. The second factor

incorporates information on credit measures in the top series in the first half of the sample while the third integrated more information on interest rates. The second factor becomes more like the early third factor in the second subperiod and vice-versa. These factors both seem more diffuse and less stable than the first.

The fourth factor is much harder to interpret, except perhaps for the earlier half of the sample when it captured mostly production and labour market movements. It is, however, interesting to note that a rather neat difference shows up in the squared loadings right before the last financial crisis in this factor, just as in the second and third factor. This change is in fact also present in other factors as shown in Figure 8. The top series featured in Table 2 confirm the difficulty of pinning down an interpretation for this factor in the time dimension as the main series change a lot across subperiods. However, we do note that the squared loadings have been remarkably more stable since the last crisis.

The fifth factor displayed in Figure 8 seems very much centred on labour market movements early in the sample. The interpretation becomes unclear between about 1995 and the last crisis. In the last subperiod, it captures the behaviour of several key variables of international trade and price indexes. Table 3 reveal that exchange rates were important variables since the second subperiod, but came to dominate only by the end of the sample. The next factor has clearer and more stable interpretation, capturing labour market movements. More specifically, it consists essentially of the observed behaviour of construction wage indexes. The seventh factor captured trade and stock variables until the financial crisis, after which it came to capture labour and price index movements. The last factor was much less stable than all others, even within each subperiod, and became very diffuse after 2005. It seems to have been centred on housing and manufacturing orders prior to becoming very hard to read.

3 Predicting Recessions

In this section we verify the ability of the dataset in analyzing the Canadian business cycle. To begin, we need an operational definition of a recession. We assume peaks and troughs are observed and they coincide with the dates from Cross and Bergevin (2012).² Since 1981, C.D. Howe committee has identified three recessions: June 1981 - October 1982, March 1990 - April 1992 and October 2008 - May 2009. Hence, these are fairly rare events in our dataset so we will not be able to do a pseudo-out-of-sample forecasting evaluation. We will focus only on the in-sample capability to correctly identify the turning points and to discover important leading indicators of the business cycle.

We adopt a static Probit (and Logit) to model the probability of recession since this the

²OCDE also produces recession dates for Canada but they suggest an unrealistically large number of downturns.

Table 2: Top ten explained series for factors 1 to 4

Factor 1		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
build_Total_CAN	build_Total_CAN	build_Total_CAN
build_Total_ONT	build_Total_ONT	build_Total_ONT
build_Total_QC	build_Total_QC	build_Total_QC
build_Total_ATL	build_Comm_CAN	build_Comm_CAN
DUR_INV_RAT_new	build_Total_ATL	build_Total_ATL
UNEMP_DURA.1.4.CAN	N_DUR_INV_RAT_new	N_DUR_INV_RAT_new
MANU_INV_RAT_new	build_Ind_CAN	build_Ind_CAN
build_Comm_CAN	MANU_INV_RAT_new	build_Comm_ONT
UNEMP_DURA.5.13.CAN	build_Comm_ONT	build_Comm_QC
build_Comm_QC	build_Comm_QC	build_Total_BC
Factor 2		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
FIN_new	TBILL.6M	GPI_new
G_AVG_10p.TBILL.3M	TBILL.3M	TBILL.6M
G_AVG_5.10.Bank_rate	GOV_AVG_1.3Y	TBILL.3M
G_AVG_1.5.Bank_rate	PC.PAPER.3M	IP_new
G_AVG_1.3.Bank_rate	BANK_RATE_L	PC.PAPER.3M
CRE.BUS	GOV_AVG_3.5Y	BANK_RATE_L
CPI.ALL.CAN	PC.PAPER.2M	GOV_AVG_1.3Y
CPI.MINUS.FEN.CAN	GOV_AVG_5.10Y	BSI_new
CPI.SHEL.CAN	MORTG.1Y	GDP_new
CRED.T	GOV_AVG_10pY	MORTG.1Y
Factor 3		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
GOV_AVG_1.3Y	G_AVG_5.10.Bank_rate	G_AVG_5.10.Bank_rate
TBILL.6M	G_AVG_10p.TBILL.3M	G_AVG_10p.TBILL.3M
TBILL.3M	G_AVG_1.5.Bank_rate	FIN_new
BANK_RATE_L	G_AVG_1.3.Bank_rate	G_AVG_1.5.Bank_rate
GOV_AVG_3.5Y	FIN_new	G_AVG_1.3.Bank_rate
PC.PAPER.3M	BSI_new	BSI_new
GOV_AVG_5.10Y	GDP_new	GDP_new
PC.PAPER.2M	IP_new	CRED.T
GOV_AVG_10pY	DM_new	CRE.BUS
MORTG.1Y	GPI_new	CPI.SERV_CAN
Factor 4		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
GDP_new	hstart_QC_new	CPI.MINUS.FOO.CAN
BSI_new	GDP_new	GOV_AVG_1.3Y
hstart_QC_new	hstart_CAN_new	G_AVG_5.10.Bank_rate
hstart_CAN_new	BSI_new	G_AVG_1.5.Bank_rate
hstart_ONT_new	CPI.MINUS.FOO.CAN	G_AVG_1.3.Bank_rate
SPI_new	SPI_new	G_AVG_10p.TBILL.3M
GPI_new	CRED.MORT	CPI.ALL.CAN
USDCAD_new	hstart_ONT_new	GOV_AVG_3.5Y
CRED.MORT	CPI.ALL.CAN	TBILL.6M
CRED.HOUS	GOV_AVG_1.3Y	CPI.SHEL.CAN

Notes: Factor loadings estimated recursively with an expanding window. Rankings are based on mean squared loadings over the indicated period.

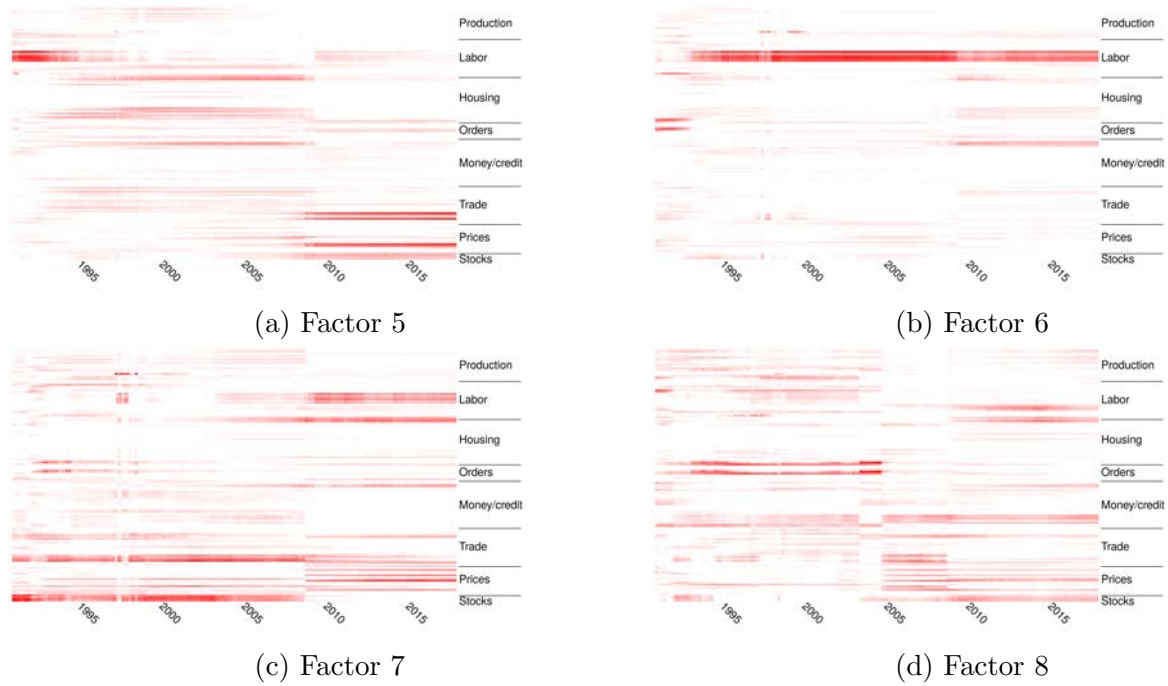


Figure 8: Heatmaps for factors 5 to 8

Notes: Factors and loadings are estimated recursively using an expanding window. Displayed shades of red capture squared loadings.

Table 3: Top ten explained series for factors 5 to 8

Factor 5		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
hstart_QC_new	hstart_QC_new	USDCAD_new
WAGE_CON_CAN	CRED_MORT	CERI_new
WAGE_CON_PRA	IPPLMOTOR_CAN	IPPLMOTOR_CAN
WAGE_CON_BC	USDCAD_new	IPPLMACH_CAN
WAGE_CON_ONT	CRED_HOUS	DJ_LO
build_Total_PRA	hstart_ATL_new	JPYCAD_new
CRED_MORT	IPPL_CAN	DJ_HI
build_Comm_BC	CERI_new	SP500
CRED_HOUS	hstart_CAN_new	DJ_CLO
build_Ind_PRA	Exp_BP_new	MANU_UNFIL_new
Factor 6		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
WAGE_CON_CAN	WAGE_CON_CAN	WAGE_CON_CAN
WAGE_CON_ONT	WAGE_CON_ONT	WAGE_CON_PRA
WAGE_CON_BC	WAGE_CON_PRA	WAGE_CON_ONT
WAGE_CON_PRA	WAGE_CON_BC	WAGE_CON_BC
WAGE_CON_QC	WAGE_CON_QC	WAGE_CON_QC
WAGE_CON_ATL	WAGE_CON_ATL	CRED_HOUS
FIN_new	CRED_HOUS	WAGE_CON_ATL
GOOD_HRS_CAN	CRED_MORT	CRED_MORT
MANU_TOT_INV_new	IPPL_CAN	CRED_CONS
DUR_N_ORD_new	CRED_T	hstart_QC_new
Factor 7		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
CERI_new	CERI_new	IPPLENER_CAN
DJ_CLO	GBPCAD_new	hstart_QC_new
DJ_LO	hstart_ATL_new	hstart_ATL_new
SP500	USDCAD_new	WAGE_CON_CAN
DJ_HI	hstart_QC_new	CPLGOOD_CAN
GBPCAD_new	SP500	WAGE_CON_ONT
IPPLMOTOR_CAN	DJ_LO	CPLMINUS_FOO_CAN
JPYCAD_new	JPYCAD_new	WTISPLC
USDCAD_new	DJ_CLO	WAGE_CON_QC
IPPL_CAN	IPPLMOTOR_CAN	CPLALL_CAN
Factor 8		
1991M1-2005M1	2005M1-2010M1	2010M1-2017M12
DUR_TOT_INV_new	IPPLENER_CAN	hstart_QC_new
MANU_TOT_INV_new	WTISPLC	UNEMP_DURA_14.25_CAN
MANU_UNFIL_new	CERI_new	SP500
DUR_UNFIL_new	EOIL_BP_new	G_AVG_10p_TBILL_3M
EMP_MANU_CAN	G_AVG_1.3.Bank_rate	UNEMP_DURA_5.13_CAN
RES_TOT	G_AVG_1.5.Bank_rate	IPPL_CAN
RES_USD	CPLGOOD_CAN	G_AVG_5.10.Bank_rate
IPPLMETAL_CAN	USDCAD_new	IPPLENER_CAN
EMP_CAN	CPLMINUS_FOO_CAN	DJ_CLO
RT_new	IPPLMOTOR_CAN	G_AVG_1.5.Bank_rate

Notes: Factor loadings estimated recursively with an expanding window. Rankings are based on mean squared loadings over the indicated period.

standard approach in the literature. Let $Z_{h,t}$ be a latent lead indicator:

$$Z_{h,t} = \alpha_h + \beta_h X_t + u_{h,t}, \quad (2)$$

with $u_{h,t} \sim N(0, 1)$, and which satisfies:

$$R_{t+h} = \begin{cases} 1 & \text{if } Z_{h,t} > 0 \\ 0 & \text{otherwise} \end{cases},$$

where h is the forecasting horizon. Since Estrella and Mishkin (1998) it is standard practice to consider the slope of the yield curve as the only predictor. It is usually proxied by the term spread (TS) which is the difference between the 10-year and 3-month Treasury bills. This will be our benchmark model. Therefore, the probability of recession is

$$P(R_{t+h} = 1 | TS_t) = \Phi(\alpha_h + \beta_h TS_t). \quad (3)$$

Then, we consider two ways of including the information from our large macroeconomic dataset in predicting business cycle turning points. The first is the static Probit where instead of X_t we consider factors estimated as principal components of X_t :

$$Z_{h,t} = F_t \gamma_h + u_{h,t} \quad (4)$$

$$X_t = \Lambda F_t + e_t \quad (5)$$

The probability of recession is then

$$P(R_{t+h} = 1 | F_t) = \Phi(\alpha_h + \beta_h F_t). \quad (6)$$

This is a two-step procedure. First, K principal components are constructed. Second, Probit model is estimated with those K factors as inputs. Note that this considered as *dense* modelling sense that all series in X_t are used in the model ($\hat{F}_t$ are linear combinations of all elements of X_t).

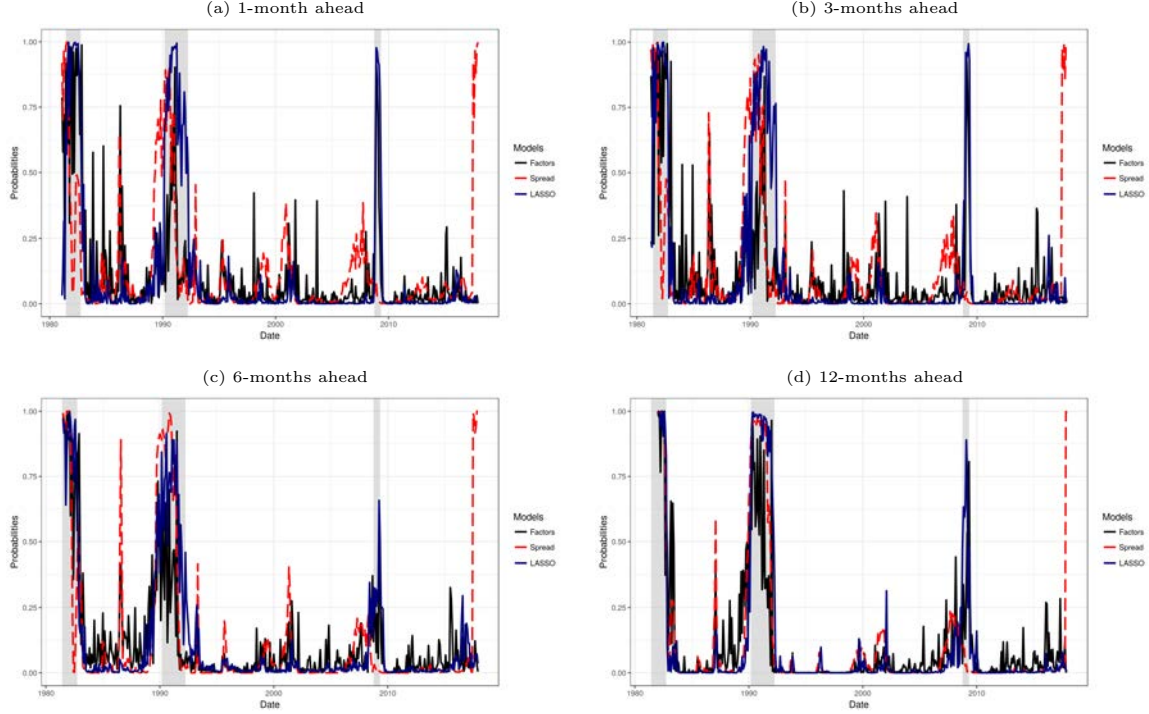
Another popular way to include a large number of predictors is through a Lasso model. Following Sties (2017), we use the Logit Lasso model:

$$P(R_{t+h} = 1 | X_t) = \frac{e^{(\alpha_h + \beta_h X_t)}}{1 + e^{(\alpha_h + \beta_h X_t)}}, \quad (7)$$

with

$$\arg \min_{\alpha_h, \beta_h} \left[RSS + \lambda \sum_{i=1}^N |\beta_{h,i}| \right].$$

Figure 9: Predicting recessions: full sample probabilities



Note: This Figure reports the estimated probabilities of recessions from all three models and for horizons 1, 3, 6 and 12 months ahead. The shaded areas correspond to C.D. Howe recession dates.

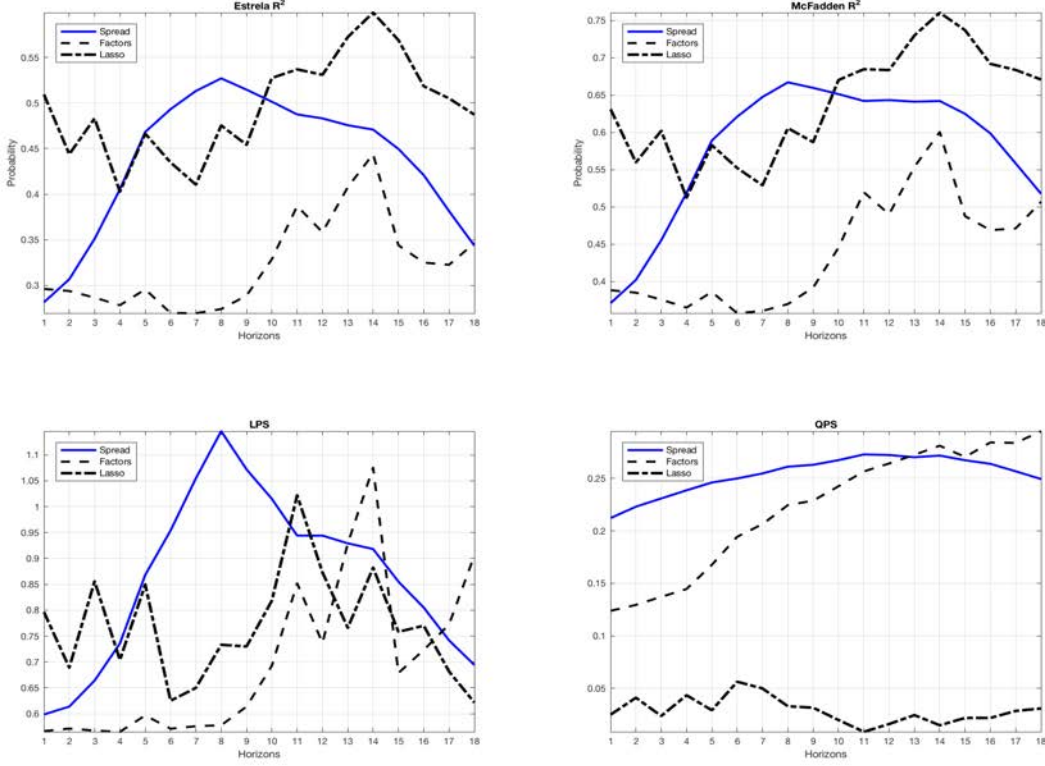
This is known as *sparse* modeling since many elements of β_h are set to zero. As opposed to the factor-Probit model 6, this is a one-step procedure.

Those three models are evaluated through Estrella and McFadden pseudo- R^2 s, quadratic probability score (QPS) and log probability score (LPS). The forecasting horizons are 1 to 18 months. Figure 9 shows the full-sample estimated probabilities for horizons 1, 3, 6 and 12-month ahead. Overall, all three models produce high probabilities during the C.D. Howe recession dates. Spread and factors' probit models produce more volatile probabilities and present a lot of 'false' signals.³ Some of them are interpretable. The peak in 1987 is caused by the stock market crash, while the 2001 increase in recession probability reflects the U.S. recession. On the other hand, the Lasso model probabilities are much smoother across all horizons.

Figure 10 shows goodness of fit measures across horizons for all three models. In terms of pseudo- R^2 , Spread model performance is maximized around 8-month ahead which has been already reported in the literature at least for US economy. Factors have better explanatory power at short and long horizons. In terms of LPS and QPS, the Lasso model is preferred to

³We call a false signal when the estimated probability is high while the C.D. Howe did not classify that observation as a recession. Of course, this false signal may also reveal some economic disturbances that were not pervasive or big enough to be judged as recession by the committee.

Figure 10: Predicting recessions: goodness of fit



Note: This Figure shows several in-sample goodness-of-fit measures for all three models and for all horizons.

the Probit alternatives, especially in case of the quadratic probability score.

Table 4 reports the 10 most important series of X_t selected by Lasso procedure for horizons 1, 3, 6 and 12 months ahead. For short-term horizons, 1 and 3 months, the most important predictor is the manufacturing inventory to sales ratio, followed by housing starts in Prairies and British Columbia. Several labour market indicators are relevant such as the unemployment duration 27+ weeks and the aggregate employment growth. There are also financial series like the spread between commercial paper and bank rate, stock market as well as the consumer credit and money supply.

For 6-month horizon the most important predictor is durables inventory to sales ratio, followed by housing starts and unemployment duration. Standard term spreads starts entering the top 10, while they become the most relevant series in case of one year ahead prediction. Overall, we find that inventory to sales ratios, housing starts, labour market and term spreads are the most important leading indicators when monitoring the Canadian business cycle.

Table 4: Predicting recessions: top 10 series in Lasso

	h=1	h=3	h=6	h=12
1	MANU-INV-RAT-new	MANU-INV-RAT-new	DUR-INV-RAT-new	G-AVG-5.10.Bank-rate
2	hstart-PRA-new	hstart-PRA-new	hstart-BC-new	TBILL-6M.Bank-rate
3	UNEMP-DURA-27.-CAN	hstart-BC-new	hstart-PRA-new	PC-PAPER-1M
4	EMP-CAN	UNEMP-DURA-27.-CAN	UNEMP-DURA-27.-CAN	hstart-BC-new
5	CRED-CONS	M3	G-AVG-10p.TBILL-3M	hstart-ONT-new
6	hstart-BC-new	build-Comm-PRA	TBILL-6M.Bank-rate	UNEMP-DURA-27.-CAN
7	PC-3M.Bank-rate	EMP-CAN	build-Comm-BC	CPI-DUR-CAN
8	BSI-new	build-Ind-BC	IPPI-METAL-CAN	RT-new
9	EMP-CONS-CAN	EMP-SALES-CAN	PC-3M.Bank-rate	hstart-PRA-new
10	SP500	WTISPLC	DJ-HI	CRED-MORT

Note: This Table reports 10 most important predictors selected by Lasso.

4 Forecasting Economic Activity

In order to explore the potential for predictive modelling of the CAN-MD database, we perform a standard out-of-sample forecasting exercise. Let Y_t be the variable of interest. If Y_t is stationary, the goal is to forecast its average over h periods:

$$y_{t+h}^{(h)} = (1200/h) \sum_{k=1}^h y_{t+k}, \quad (8)$$

where $y_t \equiv \ln Y_t$. If Y_t is an $I(1)$ series, then we forecast the average annualized growth rate as in McCracken and Ng (2016):

$$y_{t+h}^{(h)} = (1200/h) \ln(Y_{t+h}/Y_t). \quad (9)$$

4.1 Forecasting Models

A large number of forecasting techniques has been proposed to deal with large datasets, see Leroux et al. (2017) for a review and comparison. The goal of this section is to verify whether the CAN-MD dataset has some relevant and significant forecasting power in predicting key Canadian macroeconomic series. Therefore, we will use only a subset of standard time series models and data-rich methods.

Autoregressive Direct (ARD) The benchmark time series model is the *autoregressive direct* (ARD) model, which is specified as:

$$y_{t+h}^{(h)} = \alpha^{(h)} + \sum_{l=1}^{p_y^h} \rho_l^{(h)} y_{t-l+1} + e_{t+h}, \quad t = 1, \dots, T, \quad (10)$$

where $y_t \equiv \ln Y_t - \ln Y_{t-1}$, $h \geq 1$ and $p_y^h < \infty$. A direct prediction of $y_{T+h}^{(h)}$ is deduced from the model above as follows:

$$\hat{y}_{T+h|T}^{(h)} = \hat{\alpha}^{(h)} + \sum_{l=1}^{p_y^h} \hat{\rho}_l^{(h)} y_{T-l+1},$$

where $\hat{\alpha}^{(h)}$ and $\hat{\rho}^{(h)}$ are OLS estimators of $\alpha^{(h)}$ and $\rho^{(h)}$. The optimal p_y^h is selected using the Bayesian Information Criterion (BIC) for every out-of-sample (OOS) period.

Diffusion Indices (ARDI) The first data-rich model is the (direct) autoregression augmented with diffusion indices from Stock and Watson (2002b):

$$y_{t+h}^{(h)} = \alpha^{(h)} + \sum_{l=1}^{p_y^h} \rho_l^{(h)} y_{t-l+1} + \sum_{l=1}^{p_f^h} F_{t-l+1} \beta_l^{(h)} + e_{t+h}, \quad t = 1, \dots, T \quad (11)$$

$$X_t = \Lambda F_t + u_t \quad (12)$$

where X_t is the N -dimensional large informational set, F_t are $K^{(h)}$ static factors and the superscript h stands for the value of K when forecasting h periods ahead. The optimal values of p_y^h , p_f^h and $K^{(h)}$ are simultaneously selected by BIC. The h -step ahead forecast is obtained as:

$$\hat{y}_{T+h|T}^{(h)} = \hat{\alpha}^{(h)} + \sum_{l=1}^{p_y^h} \hat{\rho}_l^{(h)} y_{T-l+1} + \sum_{l=1}^{p_f^h} F_{T-l+1} \hat{\beta}_l^{(h)}.$$

The feasible ARDI model is obtained after estimating F_t as the first $K^{(h)}$ principal components of X_t .⁴

Targeted Diffusion Indices (T-ARDI) A critique of the ARDI model is that not necessarily all series in X_t are equally important to predict $y_{t+h}^{(h)}$. The ARDIT model of Bai and Ng (2008) takes this aspect into account. The idea is first to pre-select a subset X_t^* of the series in X_t , that are relevant for forecasting $y_{t+h}^{(h)}$. Then, predict $y_{t+h}^{(h)}$ with ARDI model using X_t^* to calculate factors. X_t^* can be constructed as follows:

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left[RSS + \lambda \sum_{i=1}^N |\beta_i| \right] \quad (13)$$

$$X_t^* = \{X_i \in X_t \mid \beta_i^{lasso} \neq 0\} \quad (14)$$

The approach uses the LASSO technique to select X_t^* by regressing $y_{t+h}^{(h)}$ on all elements

⁴See Stock and Watson (2002a) for technical details on the estimation of F_t as well as their asymptotic properties.

of X_t and using LASSO penalty to discard uninformative predictors.⁵ The λ parameter is selected by cross-validation for every out-of-sample period, while the ARDI hyperparameters are selected by BIC.

Regularized Data-Rich Model Averaging Leroux et al. (2017) proposed a new class of data-rich model averaging techniques that combines pre-selection and regularization with the complete subset regressions (CSR) of Elliott et al. (2013). The idea of CSR is to generate a large number of predictions based on different subsets of X_t and then construct the final forecast as the simple average of the individual forecasts:

$$\hat{y}_{t+h,m}^{(h)} = c + \rho y_t + \beta X_{t,m} + \varepsilon_{t,m} \quad (15)$$

$$\hat{y}_{T+h|T}^{(h)} = \frac{\sum_{m=1}^M \hat{y}_{T+h|T,m}^{(h)}}{M} \quad (16)$$

where $X_{t,m}$ contains L series for each model $m = 1, \dots, M$.⁶

Leroux et al. (2017) modified the original model in two ways. First, they propose to preselect a subset of relevant predictors (first step) before applying the CSR algorithm (second step). This model is labelled Targeted CSR (**T-CSR**). This initial step is meant to discipline the behaviour of the CSR algorithm ex ante. In this paper, the set of predictors X is reduced into a subset X_t^* at the first step by Lasso as in 13. We consider T-CSR with 10 and with 20 regressors, labelled T-CSR,10 and T-CSR,20 respectively.

Alternatively, one may choose to use the entire set of predictors but discipline the CSR algorithm ex post using a Ridge penalization. This model is labelled (**CSR-R**). Each CSR predictive regression is estimated as follows

$$\hat{y}_{t+h,m} = c + \rho y_t + \beta X_{t,m} \quad (17)$$

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left[RSS + \lambda \sum_{i=1}^N \beta_i^2 \right], \quad (18)$$

and then the final forecast is constructed as usual:

$$\hat{y}_{T+h|T}^{(h)} = \frac{\sum_{m=1}^M \hat{y}_{T+h|T,m}^{(h)}}{M}$$

Ridge penalization permits to elude multicollinearity that can occur in some subsets of predictors, and produces a well-behaved forecast from every subsample. We consider two specifica-

⁵Bai and Ng (2008) have also considered the hard threshold case, where a univariate predictive regression is performed for each variable X_{it} at the time. The subset X_t^* is then obtained by gathering those series whose coefficients $\beta_i^{(h)}$ have their t -stat larger than the critical value t_c .

⁶ L is usually set to 1, 10 or 20 and M is the total number of models considered (up to 5,000 in this paper).

tions of CSR-R: one based on 10 regressors and another on 20 regressors, labelled CSR-R,10 and CSR-R,20 respectively.

4.2 Pseudo-Out-of-Sample Experiment Design

The pseudo-out-of-sample period is 1991:01 - 2017:12. The forecasting horizons considered are 1, 3, 6, 9 and 12 months. All models are estimated on the rolling window through the out-of-sample period. For each model, the hyperparameters are selected by BIC, except for the Lasso parameter that is tuned by cross validation, for every evaluation period and forecasting horizon.

We consider the following variables: industrial production, employment, unemployment rate, consumer price index, core consumer price index, housing starts and the 5-y mortgage rate. These are typical macroeconomic aggregates that have been forecasted in the previous literature. All the series except housing starts are modelled as $I(1)$, hence we forecast the annualized growth rates.

The forecasting performance of the above models will be compared on the basis of the Root Mean Square Prediction Error (RMSPE) as is often the case in forecasting literature. Other metrics could be used but for the sake of simplicity and under space constraints we stick to the most common one.

4.3 Results

Tables 5 - 7 summarize the results. We report the value of RMSPE, ratio with respect to the reference ARD model as well as the p-value of Diebold-Mariano test. We group the variables in three categories: real activity (industrial production, employment and unemployment rate), inflation (CPI and core CPI), housing market (housing starts and 5-year mortgage rate).

Using our large database improves substantially the predictions of real activity series. For instance, when forecasting industrial production one month ahead, ARDI and CSR-R models outperform significantly the autoregressive model. For $h = 3$, improvements are even larger. At longer horizons, only the CSR-R models show significant ameliorations. The forecasting gains are between 8,6 and 18%. We get similar results for two labour market series. CSR-R models are in general the best with improvements between 6 and 17%.

Table 6 shows that using the large panel greatly improves the prediction of inflation series, especially for longer horizons. ARDI and CSR-R model are particularly resilient. For instance, one-year ahead forecasting of CPI inflation is ameliorated by as much as 32% with the CSR-R,20 model. The improvements are in general significant.

Finally, Table 7 reports the results for the housing market. Improvements provided by our large dataset and the data-rich forecasting techniques are smaller in this case. In case of housing

Table 5: Forecasting real activity

Models	Industrial Production											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,010	1		0,006	1	0,005	1		0,004	1		
ARDI	0,009	0,929	0,007	0,006	0,858	0,001	0,005	0,946	0,177	0,004	1,066	0,297
T-ARDI	0,010	0,990	0,429	0,006	0,944	0,179	0,005	1,075	0,195	0,005	1,267	0,033
T-CSR,10	0,010	0,992	0,445	0,006	0,873	0,005	0,005	0,947	0,229	0,004	1,012	0,438
T-CSR,20	0,010	1,020	0,364	0,006	0,891	0,033	0,005	1,054	0,256	0,004	1,055	0,275
CSR-R,10	0,009	0,929	0,009	0,006	0,856	0,000	0,005	0,904	0,005	0,004	0,914	0,065
CSR-R,20	0,009	0,914	0,004	0,006	0,820	0,000	0,005	0,901	0,024	0,004	0,940	0,228
	Employment											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,002	1		0,001	1	0,001	1		0,001	1		
ARDI	0,002	0,939	0,118	0,001	1,077	0,238	0,001	1,187	0,033	0,001	1,173	0,082
T-ARDI	0,002	1,072	0,116	0,001	1,089	0,119	0,001	1,256	0,024	0,001	1,319	0,023
T-CSR,10	0,002	1,077	0,133	0,001	0,996	0,477	0,001	1,001	0,494	0,001	1,016	0,440
T-CSR,20	0,002	1,160	0,014	0,001	1,099	0,121	0,001	1,075	0,217	0,001	1,150	0,140
CSR-R,10	0,002	0,964	0,252	0,001	0,921	0,117	0,001	0,909	0,043	0,001	0,875	0,083
CSR-R,20	0,002	0,959	0,233	0,001	0,908	0,143	0,001	0,905	0,095	0,001	0,886	0,149
	Unemployment rate											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,185	1		0,110	1	0,088	1		0,080	1		
ARDI	0,179	0,930	0,119	0,101	0,845	0,119	0,085	0,940	0,297	0,085	1,149	0,131
T-ARDI	0,182	0,962	0,239	0,103	0,869	0,090	0,089	1,035	0,394	0,092	1,337	0,014
T-CSR,10	0,181	0,953	0,169	0,098	0,792	0,017	0,086	0,962	0,335	0,080	0,997	0,491
T-CSR,20	0,187	1,019	0,364	0,101	0,833	0,047	0,090	1,061	0,283	0,083	1,090	0,253
CSR-R,10	0,177	0,916	0,003	0,098	0,790	0,002	0,081	0,855	0,007	0,073	0,827	0,025
CSR-R,20	0,177	0,909	0,017	0,096	0,750	0,005	0,080	0,832	0,021	0,073	0,837	0,068

Note: This Table reports the root mean squared predictive error (RMSPE), the ratio with respect to the reference ARD model (Relative) and the p-value of Diebold-Mariano test (DM).

Table 6: Forecasting inflation

	CPI											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,004	1		0,002	1	0,002	1		0,001	1		
ARDI	0,004	0,971	0,200	0,002	0,964	0,218	0,002	0,938	0,281	0,001	0,715	0,028
T-ARDI	0,004	1,016	0,379	0,002	0,970	0,350	0,002	1,110	0,223	0,001	0,934	0,335
T-CSR,10	0,004	0,998	0,479	0,002	0,951	0,261	0,002	1,040	0,385	0,001	0,815	0,086
T-CSR,20	0,004	1,020	0,321	0,002	0,980	0,398	0,002	1,091	0,278	0,001	0,838	0,162
CSR-R,10	0,004	0,915	0,002	0,002	0,903	0,016	0,002	0,965	0,382	0,001	0,718	0,002
CSR-R,20	0,004	0,898	0,003	0,002	0,904	0,036	0,002	0,968	0,405	0,001	0,680	0,012
	Core CPI											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,003	1		0,002	1	0,001	1		0,001	1		
ARDI	0,003	0,944	0,093	0,002	0,881	0,069	0,001	0,775	0,052	0,001	0,794	0,147
T-ARDI	0,003	1,108	0,010	0,002	0,942	0,249	0,001	1,024	0,437	0,001	1,084	0,325
T-CSR,10	0,003	1,053	0,175	0,002	1,048	0,304	0,001	0,934	0,333	0,001	0,907	0,291
T-CSR,20	0,003	1,049	0,203	0,002	1,050	0,295	0,001	1,004	0,491	0,001	0,936	0,378
CSR-R,10	0,003	0,941	0,139	0,002	0,940	0,186	0,001	0,778	0,005	0,001	0,736	0,011
CSR-R,20	0,003	0,921	0,086	0,002	0,926	0,187	0,001	0,759	0,024	0,001	0,703	0,041

Note: See Table 5.

Table 7: Forecasting housing market

	Housing starts											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,100	1		0,130	1	0,163	1		0,221	1		
ARDI	0,100	1,000	0,497	0,135	1,067	0,086	0,171	1,094	0,144	0,214	0,942	0,298
T-ARDI	0,103	1,078	0,056	0,143	1,205	0,002	0,188	1,321	0,008	0,238	1,158	0,117
T-CSR,10	0,101	1,031	0,242	0,132	1,018	0,374	0,170	1,085	0,171	0,223	1,023	0,402
T-CSR,20	0,105	1,109	0,058	0,136	1,094	0,074	0,177	1,171	0,041	0,222	1,014	0,442
CSR-R,10	0,101	1,023	0,278	0,128	0,962	0,207	0,159	0,946	0,149	0,212	0,921	0,060
CSR-R,20	0,100	1,013	0,393	0,128	0,966	0,239	0,159	0,952	0,239	0,213	0,932	0,182
	5-Y mortgage rate											
	h=1			h=3			h=6			h=12		
	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM	RMSPE	Relative	DM
ARD	0,272	1		0,178	1	0,118	1		0,084	1		
ARDI	0,270	0,986	0,413	0,181	1,024	0,349	0,124	1,121	0,083	0,097	1,335	0,011
T-ARDI	0,256	0,886	0,106	0,178	0,992	0,463	0,135	1,323	0,037	0,115	1,871	0,000
T-CSR,10	0,256	0,882	0,082	0,183	1,049	0,259	0,134	1,308	0,025	0,106	1,585	0,000
T-CSR,20	0,263	0,931	0,223	0,186	1,084	0,151	0,135	1,330	0,015	0,115	1,876	0,000
CSR-R,10	0,260	0,913	0,002	0,172	0,928	0,032	0,116	0,974	0,280	0,088	1,093	0,077
CSR-R,20	0,257	0,892	0,016	0,172	0,931	0,098	0,116	0,975	0,334	0,093	1,219	0,007

Note: See Table 5.

starts, no model can beat the autoregressive reference at one-month horizon. There are small improvements at horizons 3 and 6 months, but they are not significant. The only substantial increase in forecasting power is at the one-year horizon with CSR-R,10 model, which improves the prediction by 8%. In case of 5-y mortgage rate, all data-rich models outperform the ARD at $h = 1$. The best is T-CSR,10 with a 12% improvement. CSR-R,10 is the best at three-month horizon, but for longer horizons the autoregressive model is generally the best option.

5 Measuring macroeconomic uncertainty

In this section we demonstrate how our panel can be used to measure macroeconomic uncertainty. To do so, we rely on Jurado et al. (2015) who construct the uncertainty measures as the common stochastic volatility of forecasting errors of many macroeconomic indicators.

The uncertainty is defined as the conditional volatility of forecasting error of variable y_j :

$$U_{jt}^y(h) = \sqrt{E[(y_{jt+h} - E[y_{jt+h}|I_t])^2|I_t]} \quad (19)$$

where $E[y_{jt+h}|I_t]$ is the optimal h -step ahead forecast and I_t measures the information available to economic agents at time t . Since this can be constructed for any variable $j = 1, \dots, N$, it is natural to aggregate individual uncertainties into a common index:

$$U_t^y(h) =_{N_y \rightarrow \infty} \sum_{j=1}^{N_y} \omega_j U_{jt}^y(h) \equiv E_\omega[U_{jt}^y(h)]. \quad (20)$$

To implement the procedure, one needs to construct the optimal forecasts for many variables, produce the conditional volatilities of the forecasting errors and then aggregate them up. For the first part, the following forecasting model is considered:

$$\begin{aligned} X_t &= \Lambda F_t + u_t \\ X_t^2 &= \Lambda^W W_t + u_t \\ y_{jt+h} &= \rho(L)y_{jt} + \beta(L)F_t + \gamma(L)F_{1,t}^2 + \delta(L)W_t + e_{jt+h} \end{aligned}$$

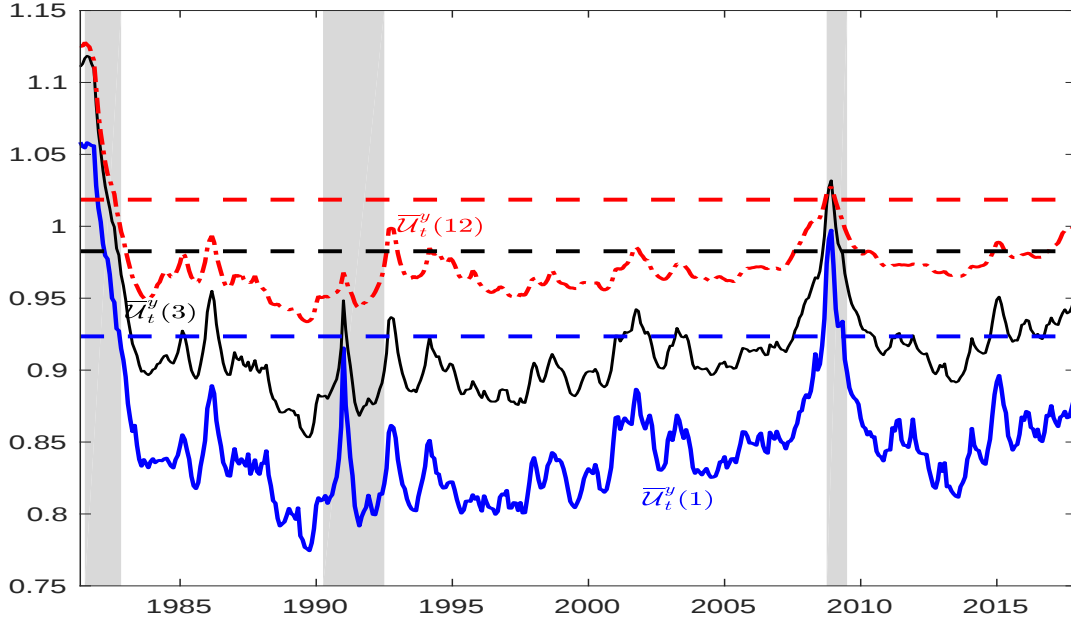
This is essentially a factor-augmented predictive regression as the ARDI model from the previous section. The only difference is the inclusion of nonlinearities by considering factors from squared data, W_t , and adding the square of the first *level* factor directly in the predictive regression, $F_{1,t}^2$.

In this application, X_t is our large dimensional Canadian dataset as used in the previous section. The number of factors in F_t and W_t is selected by Bai and Ng (2002). The lag orders of $\rho(L)$ and $[\beta(L), \gamma(L), \delta(L)]$ fixed to 4 and 2. The forecasting horizons are $h = (1, 2, \dots, 12)$ months.⁷ The target y_{jt+h} is the j^{th} variable in the dataset. Once the in-sample forecast is obtained, an univariate stochastic volatility model is estimated on every time series of forecasting errors to get $\hat{U}_{jt}^y(h)$. Finally, the measure of macroeconomic uncertainty is constructed by aggregating individual uncertainties $\hat{U}_{jt}^y(h)$. This can be done by taking the cross-sectional average or by extracting the first principal component. Also, we can aggregate across all series, which provides the aggregate macroeconomic uncertainty, or condition across sectors or regions.

As an example, the Figure 11 shows the macroeconomic uncertainty constructed as the cross-sectional average of all individual uncertainties. Several observations are worth mentioning. Uncertainty has a clear countercyclical behavior since it systematically increases during recessions. Horizontal lines indicate 1.65 standard deviation above the mean of each series. There are only few episodes where the uncertainty exceeds the 1.65 standard deviation above the mean: 1981 and 2009 recessions. Recall that shaded areas are C.D. Howe committee recession dates. The committee did not classify 2001 as a recession like in U.S., neither the 2015 when oil prices went down. There are, however, some moderate increases in the measured uncertainty during those periods. Note that the level of uncertainty increases with the forecasting horizon h , as well as the variability since the forecast tends to unconditional mean when $h \rightarrow \infty$.

⁷We followed Jurado et al. (2015) for this choice of hyperparameters.

Figure 11: Measure of macroeconomic uncertainty



Note: This Figure shows aggregate uncertainty measures for horizons $h = 1, 3, 12$. Horizontal lines indicate 1.65 standard deviation above the mean of each series.

6 Impulse Response Analysis

In this section we use the CAN-MD dataset to conduct a structural impulse response analysis. The empirical model is a factor-augmented VAR as in Boivin et al. (2018). Start with the following state-space factor representation

$$X_t = \Lambda F_t + u_t, \quad (21)$$

$$F_t = \Phi(L)F_{t-1} + e_t, \quad (22)$$

where X_t contains N economic and financial indicators, F_t represents K unobserved factors ($N \gg K$), Λ is a $N \times K$ matrix of factor loadings, u_t are idiosyncratic components of X_t that are uncorrelated at all leads and lags with F_t and with the factor innovations e_t . We assume a stationary and an approximate factor model, as some limited cross-section correlations are allowed among the idiosyncratic components.⁸ The estimation of the model (21)–(22) is based on a two-step principal components procedure, where factors are approximated in the first step, and the factors' VAR process is estimated in the second step.⁹

⁸See Stock and Watson (2016) for an overview of the modern factor analysis literature, and the distinction between exact and approximate factor models.

⁹Stock and Watson (2002a) prove the consistency of such an estimator in the approximate factor model when both cross-section and time sizes, N and T , go to infinity, and without restrictions on N/T . Moreover, they justify using $\hat{F}_t$ as regressor without adjustment. Bai and Ng (2006) furthermore show that PCA estimators

The core of IRF analysis is identification of structural shocks. To do so, we apply the contemporaneous timing restrictions procedure as in Stock and Watson (2016) that identifies shocks by restricting only the impact response of a small number of economic indicators. Following Boivin et al. (2018), start from the moving-average representation of X_t :

$$\begin{aligned} X_t &= B(L)e_t + u_t, \\ &= \Lambda[I - \Phi(L)L]^{-1}e_t + u_t. \end{aligned} \quad (23)$$

Assume factor innovations e_t are linear combinations of structural shocks $\varepsilon_t = He_t$ where H is a nonsingular square matrix and $E[\varepsilon_t \varepsilon_t'] = I$. Thus, the structural moving-average representation of X_t is obtained by replacing e_t in 23:

$$\begin{aligned} X_t &= B * (L)e_t + u_t, \\ &= \Lambda[I - \Phi(L)L]^{-1}H^{-1}\varepsilon_t + u_t. \end{aligned} \quad (24)$$

To identify the structural shocks ε_t , assume a recursive structure implying that $K - 1$ indicators do not respond on impact to certain shocks. Specifically, the contemporaneous impact matrix of X_t , $B^*(0)$, takes the following form where zeros represent timing restrictions

$$B_0^* \equiv B^*(0) = \begin{bmatrix} x & 0 & \cdots & 0 \\ x & x & \ddots & 0 \\ x & x & \ddots & 0 \\ x & x & \cdots & x \\ \vdots & \vdots & \ddots & \vdots \\ x & x & \cdots & x \end{bmatrix}, \quad (25)$$

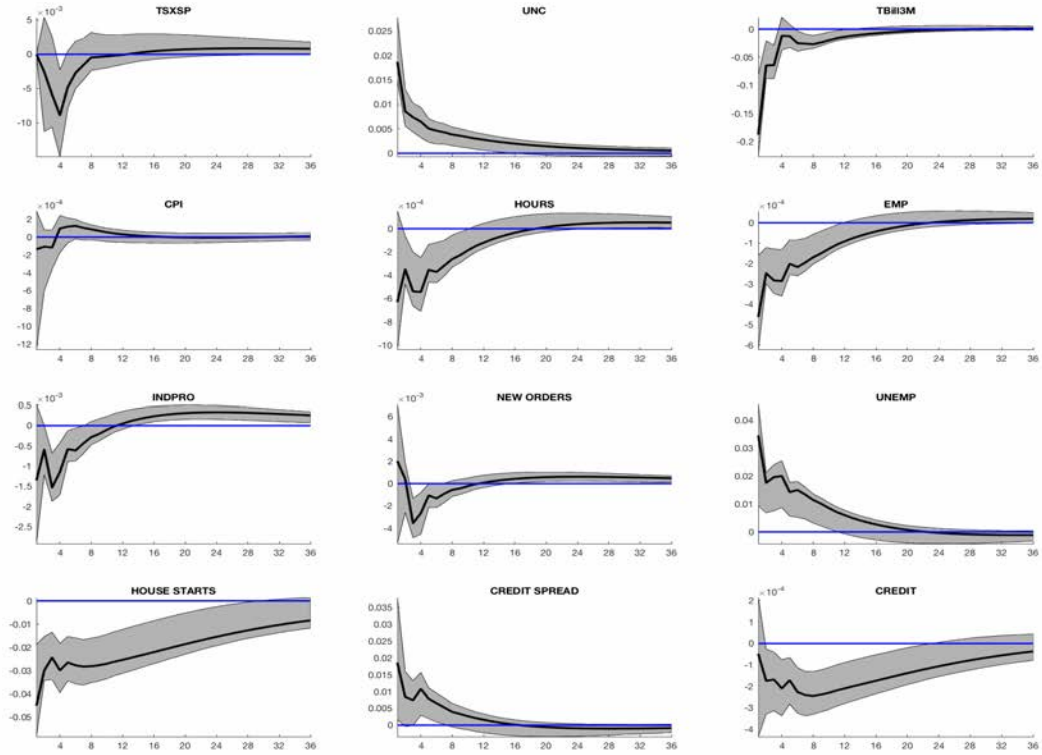
To estimate the matrix H , we proceed as in Stock and Watson (2016), noting that $B_{0:K}^* \varepsilon_t = B_{0:K} e_t$ implies $B_{0:K}^* B_{0:K}^{*'} = B_{0:K} \Sigma_e B_{0:K}'$, where $B_{0:K}$ contains the first K rows of $B_0 \equiv B(0) = \Lambda$, $B_{0:K}^* = B_{0:K} H^{-1}$, and Σ_e is the covariance matrix of e_t . Since $B_{0:K}^*$ is a $K \times K$ lower triangular matrix, then it must be the case that $B_{0:K}^*$ can be obtained by performing a Choleski decomposition of $(B_{0:K} \Sigma_e B_{0:K}')$, i.e.: $B_{0:K}^* = \text{Chol}(B_{0:K} \Sigma_e B_{0:K}')$. It follows that $H = (B_{0:K}^*)^{-1} B_{0:K}$, or

$$H = [\text{Chol}(B_{0:K} \Sigma_e B_{0:K}')]^{-1} B_{0:K}. \quad (26)$$

These restrictions allow to just-identify the matrix H and therefore identify the structural impulse responses in 24. The estimate of H is obtained by replacing $B_{0:K}$ and Σ_e with their

are $\sqrt{T}$ consistent and asymptotically normal if $\sqrt{T}/N \rightarrow 0$. Inference should take into account the effect of generated regressors, except when T/N goes to zero.

Figure 12: Dynamic responses to uncertainty shock



Note: This Figure shows dynamic responses of selected variables to an uncertainty shock. The grey areas indicate the 90% confidence intervals computed using 5000 bootstrap replications.

estimates in (26). See Boivin et al. (2018) for further details and properties of this identification procedure.

The goal in this application is to estimate the effects of uncertainty shocks on Canadian economy. The uncertainty is proxied by the cross-sectional average of all individual uncertainties reported in Figure 11. Following Bloom (2009), we impose the following recursive ordering: $[TSX, UNC, TBill3M, CPI, HOURS, EMP, INDPRO]$, where TSX stands for the TSX returns, UNC is the uncertainty measure, $TBill3M$ is the 3-month treasury bill, CPI is the CPI inflation, $Hours$ are worked hours, while EMP and $INPROD$ are employment and industrial production growth rates respectively. Hence, there are 7 latent factors in F_t . The uncertainty shock is ordered second in ε_t . The VAR lag order in equation (22) is set to 3.

The results for selected variables are reported in Figure 12. An increase in uncertainty causes a decline in financial markets as indicated from the negative reaction of TSX and treasury bills. The credit spread increases suggesting higher uncertainty is related to higher external financing premium. As a result, the volume of credit decrease. On the real side, employment, industrial production hours worked, housing starts and manufacturing new orders decline, while the

Table 8: Variance decomposition and R^2

Series	VD: common shocks	R2: common comp.	VD: total
TSXSP	0,012	0,141	0,002
UNC	0,547	0,186	0,102
TBill3M	0,318	0,740	0,235
CPI	0,008	0,709	0,006
HOURS	0,183	0,232	0,043
EMP	0,317	0,439	0,139
INDPRO	0,102	0,618	0,063
NEW ORDERS	0,038	0,341	0,013
UNEMP	0,334	0,252	0,084
HOUSE STARTS	0,364	0,629	0,229
CREDIT SPREAD	0,148	0,118	0,017
CREDIT	0,081	0,526	0,043

Notes: The second column reports the contribution of the uncertainty shock to the variance of the common component of the series at a 36-month horizon. The third column contains the fraction of the total variability of the series explained by the common component. The fourth column is the product of the previous two and represents the fraction of the total variance of the series.

unemployment increases, so that uncertainty could be seen as a contributor to the business cycle. The prices, however, do not react significantly. Overall, these results are in line with Bloom (2009) and Jurado et al. (2015) who used VAR models with similar recursive orderings on US data.

Table 8 reports the importance of uncertainty shock in explaining macroeconomic fluctuations since 1981. The second column shows the contribution of the uncertainty shock to the variance of the common component (all 7 common shocks in ε_t) of the series at a 36-month horizon. The third column reports the fraction of the variance of those series explained by the common component, which is simply the R^2 . Finally, the fourth column is the product of the second and third columns, and presents the contribution of the uncertainty shock to the total variance of the series. In other words, the variance decomposition in the fourth column takes into account the idiosyncratic shocks of the series.

Uncertainty shocks are particularly important, among common shocks, for treasury bills, labour market and housing starts where they explain around 30% of the variance. They are also relevant for hours, industrial production, credit spreads and credit volume with variance decomposition between 8% and 18%. The third column shows that the common component explains a sizeable fraction of the variability of those series, particularly for treasury bills, inflation, employment, industrial production, housing starts and credit.

7 Conclusion

In this paper we proposed a monthly large-scale Canadian macroeconomic database containing hundreds of Canadian and provincial economic indicators observed from 1981. It is designed to be updated regularly through the StatCan database and is publicly available. It relieves users to deal with data changes and methodological revisions, and is already balanced and transformed to induce stationarity.

The panel seems to have a factor structure where the common factors explain a sizeable portion of variation in Canadian and provincial aggregate series. The principal component analysis of the dataset revealed few driving forces of the Canadian economy such as housing and bond markets, production, wages, exchange rates and energy prices.

The advantages of this dataset for macroeconomic research are shown in several examples. The panel is useful to capture turning points of the Canadian business cycle. Using Probit, Lasso and factor models we showed that this dataset has substantial explanatory power in addition to the standard term spread predictor.

The information contained in the dataset has substantial predictive power when forecasting key macroeconomic indicators. Factor models and regularized complete subset regressions were very resilient in forecasting real activity variables such as industrial production, employment and unemployment rate, as well as CPI and Core CPI inflation. Gains were less important for the housing market.

The panel was also used to construct measures of macroeconomic uncertainty à la Jurado et al. (2015). The estimated Canadian macroeconomic uncertainty suggests two significant episodes of uncertainty: 1981 and 2009 recessions. Finally, we used that measure to show how the dataset can serve for structural impulse response analysis. We estimated the effects of uncertainty shocks using the factor-augmented VAR methodology à la Boivin et al. (2018) with our measure of macroeconomic uncertainty. The results suggested that an increase in uncertainty is associated with a decline in real activity and financial markets. The findings are in line with the literature using US data.

References

- Ahn, S. K. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.
- Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters*, 80(23-24):1806–1813.
- Amengual, D. and Watson, M. (2007). Consistent estimation of the number of dynamic factors in large N and T panel. *Journal of Business and Economic Statistics*, 25(1):91–96.

- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.
- Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74:1133–1150.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks. *Journal of Business and Economic Statistics*, 25(1):52–60.
- Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.
- Bedock, N. and Stevanovic, D. (2017). An empirical study of credit shock transmission in a small open economy. *Canadian Journal of Economics*, 50(2):541–570.
- Bernanke, B., Boivin, J., and Elias, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120:387–422.
- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.
- Boivin, J., Giannoni, M., and Stevanović, D. (2010). Monetary transmission in a small open economy: more data, fewer puzzles. Technical report, Columbia Business School, Columbia University.
- Boivin, J., Giannoni, M., and Stevanović, D. (2018). Dynamic effects of credit shocks in a data-rich environment. *Journal of Business and Economic Statistics*.
- Cross, P. and Bergevin, P. (2012). Turning points: Business cycles in Canada since 1926. Technical Report 366, C.D. Howe Institute, Vancouver, British Columbia.
- Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357–373.
- Estrella, A. and Mishkin, F. (1998). Predicting us recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80:45–61.
- Galbraith, J. W. and Tkacz, G. (2007). How far can forecasting models forecast? Forecast content horizons for some important macroeconomic variables. Technical report, Bank of Canada Working Paper No. 2007-1.
- Gosselin, M.-A. and Tkacz, G. (2001). Evaluating factor models: An application to forecasting inflation in Canada. Technical report, Bank of Canada Working Paper No. 2001-18.
- Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617.
- Jurado, K., Ludvigson, S., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.
- Leamer, E. E. (2015). Housing really is the business cycle: What survives the lessons of 2008-09? *Journal of Money, Credit and Banking*, 47(S1):43–50.
- Leroux, M., Kotchoni, R., and Stevanovic, D. (2017). Macroeconomic forecast accuracy in a data-rich environment. Technical report, CIRANO, 2017s-05.
- Mao Takongmo, C. and Stevanovic, D. (2015). Selection of the number of factors in presence

- of structural instability: a monte carlo study. *Actualité Économique*, 91:177–233.
- McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, 34(4):574–589.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.
- Sties, M. (2017). Forecasting recessions in a big data environment. Technical report, Department of Economics, University of Alberta.
- Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.
- Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.
- Stock, J. H. and Watson, M. W. (2016). Factor models and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics*, 2A:415–526.

Appendix: Data Sets

The transformation codes are: 1 - no transformation; 2 - first difference; 4 - logarithm; 5 - first difference of logarithm.

No.	Series Code	Series Description	T-Code
PRODUCTION			
1	GDP-new	Gross domestic product	5
2	BSI-new	Business sector production	5
3	NBSI-new	Non business sector production	5
4	GPI-new	Goods production	5
5	SPI-new	Services production	5
6	IP-new	Industrial production	5
7	NDM-new	Non durable manufacturing production	5
8	DM-new	Durable manufacturing production	5
9	OILP-new	Oil production	5
10	CON-new	Construction production	5
11	RT-new	Retail trade production	5
12	WT-new	Wholesale trade production	5
13	PA-new	Public administration production	5
14	FIN-new	Finance, insurance and real estate production	5
15	OIL-CAN	Refined petroleum production (Canada)	5
LABOR MARKET			
16	EMP-CAN	Total employment	5
17	EMP-SERV-CAN	Services employment	5
18	EMP-FOR-OIL-CAN	Forestry,oil,etc. employment	5
19	EMP-CONS-CAN	Construction employment	5
20	EMP-SALES-CAN	Wholesale and retail trade employment	5
21	EMP-FIN-CAN	Finance, insurance and real estate production	5
22	EMP-PART-CAN	Part time employment	5
23	EMP-MANU-CAN	Manufacturing employment (Canada)	5
24	UNEMP-CAN	Unemployment rate	2
25	WAGE-CON-CAN	Construction wage index (Canada)	5
26	WAGE-CON-ATL	Construction wage index (Atlantic)	5
27	WAGE-CON-QC	Construction wage index (Quebec)	5
28	WAGE-CON-ONT	Construction wage index (Ontario)	5
29	WAGE-CON-PRA	Construction wage index (Prairies)	5
30	WAGE-CON-BC	Construction wage index (British Columbia)	5
31	UNEMP-DURA-1-4-CAN	Unemployment duration (1-4 weeks)	5
32	UNEMP-DURA-5-13-CAN	Unemployment duration (5-13 weeks)	5
33	UNEMP-DURA-14-25-CAN	Unemployment duration (14-25 weeks)	5
34	UNEMP-DURA-27+-CAN	Unemployment duration (27+ weeks)	5
35	UNEMP-DURA-avg-CAN	Unemployment duration (average)	5
36	TOT-HRS-CAN	Total hours worked	5
37	GOOD-HRS-CAN	Goods sector hours worked	5
HOUSING AND CONSTRUCTION			
38	hstart-CAN-new	Housing starts (Canada)	4
39	hstart-ATL-new	Housing starts (Atlantic)	4
40	hstart-QC-new	Housing starts (Quebec)	4
41	hstart-ONT-new	Housing starts (Ontario)	4
42	hstart-PRA-new	Housing starts (Prairies)	4
43	hstart-BC-new	Housing starts (British Columbia)	4
44	build-Total-CAN	Total building permits (Canada)	4
45	build-Ind-CAN	Industrial building permits (Canada)	4
46	build-Comm-CAN	Commercial building permits (Canada)	4
47	build-Total-ATL	Total building permits (Atlantic)	4
48	build-Ind-ATL	Industrial building permits (Atlantic)	4
49	build-Comm-ATL	Commercial building permits (Atlantic)	4
50	build-Total-QC	Total building permits (Quebec)	4
51	build-Ind-QC	Industrial building permits (Quebec)	4
52	build-Comm-QC	Commercial building permits (Quebec)	4
53	build-Total-ONT	Total building permits (Ontario)	4
54	build-Ind-ONT	Industrial building permits (Ontario)	4
55	build-Comm-ONT	Commercial building permits (Ontario)	4
56	build-Total-PRA	Total building permits (Prairies)	4
57	build-Ind-PRA	Industrial building permits (Prairies)	4
58	build-Comm-PRA	Commercial building permits (Prairies)	4
59	build-Total-BC	Total building permits (British Columbia)	4
60	build-Ind-BC	Industrial building permits (British Columbia)	4
61	build-Comm-BC	Commercial building permits (British Columbia)	4
MANUFACTURING, SALES AND INVENTORIES			
62	MANU-N-ORD-new	Manufacturing new orders	5
63	MANU-UNFIL-new	Manufacturing unfilled orders	5
64	MANU-TOT-INV-new	Manufacturing total inventories	5
65	MANU-INV-RAT-new	Manufacturing inventory to sales ratio	1
66	N-DUR-INV-RAT-new	Non durable inventory to sales ratio	1
67	DUR-N-ORD-new	Durable manufacturing new orders	5
68	DUR-UNFIL-new	Durable manufacturing unfilled orders	5
69	DUR-TOT-INV-new	Durable total inventories	5
70	DUR-INV-RAT-new	Durable inventory to sales ratio	1
MONEY AND CREDIT			
71	M3	M3	5
72	M2p	M2+	5
73	M-BASE1	Monetary base	5
74	CRED-T	Total credit	5
75	CRED-HOUS	Household credit	5
76	CRED-MORT	Mortgage credit	5
77	CRED-CONS	Consumer credit	5
78	CRE-BUS	Business credit	5
79	BANK-RATE-L	Bank rate	2
80	PC-PAPER-1M	Commercial paper yield (1 month)	2
81	PC-PAPER-2M	Commercial paper yield (2 months)	2
82	PC-PAPER-3M	Commercial paper yield (3 months)	2
83	GOV-AVG-1-3Y	Governmental bonds average yield (1-3 years)	2
84	GOV-AVG-3-5Y	Governmental bonds average yield (3-5 years)	2

85	GOV-AVG-5-10Y	Governmental bonds average yield (5-10 years)	2
86	GOV-AVG-10pY	Governmental bonds average yield (10+ years)	2
87	MORTG-1Y	Mortgage rate 1 year	2
88	MORTG-5Y	Mortgage rate 3 years	2
89	TBILL-3M	Treasury bill yield (3 months)	2
90	TBILL-6M	Treasury bill yield (6 months)	2
91	PC-3M-Bank-rate	Spread: commercial paper (3 months) minus bank rate	1
92	G-AVG-1-3-Bank-rate	Spread: Governmental bonds average yield (1-3 years) minus bank rate	1
93	G-AVG-1-5-Bank-rate	Spread: Governmental bonds average yield (3-5 years) minus bank rate	1
94	G-AVG-5-10-Bank-rate	Spread: Governmental bonds average yield (5-10 years) minus bank rate	1
95	TBILL-6M-Bank-rate	Spread: treasury bill yield (6 months) minus bank rate	1
96	G-AVG-10p-TBILL-3M	Spread: Governmental bonds average yield (10+ years) minus treasury bill yield (3 months)	1
INTERNATIONAL TRADE			
97	RES-TOT	Total international reserves	5
98	RES-USD	USD reserves	5
99	RES-IMF	IMF reserve position	5
100	Imp-BP-new	Total imports, balance of payment basis	5
101	IOIL-BP-new	Oil imports, balance of payment basis	5
102	Exp-BP-new	Total exports, balance of payment basis	5
103	EOIL-BP-new	Oil exports, balance of payment basis	5
104	EX-ENER-BP-new	Energy exports, balance of payment basis	5
105	EX-MINER-BP-new	Minerals exports, balance of payment basis	5
106	EX-METAL-BP-new	Metal exports, balance of payment basis	5
107	EX-IND-EQUIP-BP-new	Industrial machinery and equipment exports, balance of payment basis	5
108	EX-TRANSP-BP-new	Motor vehicles and parts exports, balance of payment basis	5
109	EX-CONS-BP-new	Consumer goods exports, balance of payment basis	5
110	IMP-METAL-BP-new	Metal imports, balance of payment basis	5
111	IMP-IND-EQUIP-BP-new	Industrial machinery and equipment imports, balance of payment basis	5
112	IMP-TRANSP-BP-new	Motor vehicles and parts imports, balance of payment basis	5
113	IMP-CONS-BP-new	Consumer goods imports, balance of payment basis	5
114	USDCAD-new	USD to CAD exchange rate	5
115	JPYCAD-new	Yen to CAD exchange rate	5
116	GBPCAD-new	Pound to CAD exchange rate	5
117	CERI-new	CERI exchange rate	5
PRICES			
118	CPI-ALL-CAN	Consumer price index (CPI) all	5
119	CPI-SHEL-CAN	CPI shelter	5
120	CPI-CLOT-CAN	CPI clothing	5
121	CPI-HEA-CAN	CPI healthcare	5
122	CPI-MINUS-FOO-CAN	CPI minus food	5
123	CPI-MINUS-FEN-CAN	CPI minus food and energy	5
124	CPI-DUR-CAN	CPI durable goods	5
125	CPI-GOOD-CAN	CPI goods	5
126	CPI-SERV-CAN	CPI services	5
127	IPPI-CAN	Industrial production price index (IPPI) all industries	5
128	IPPI-ENER-CAN	IPPI energy	5
129	IPPI-WOOD-CAN	IPPI lumber and wood	5
130	IPPI-METAL-CAN	IPPI metal	5
131	IPPI-MOTOR-CAN	IPPI motor vehicles and parts	5
132	IPPI-MACH-CAN	IPPI machinery	5
133	WTISPLC	Western texas intermediate petrol price (from FRED, spliced)	5
STOCK MARKETS			
134	TSX-val	Toronto Stock Exchange (TSX) value index	5
135	TSX-vol	TSX volume index	5
136	DJ-HI	Dow Jones (DJ) high index	5
137	DJ-LO	DJ low index	5
138	DJ-CLO	DJ close index	5
139	SP500	Standard and Poor's 500 index	5
PROVINCIAL/REGIONAL SERIES			
PRODUCTION			
140	OIL-QC	Refined petroleum production (Quebec)	5
141	OIL-ALB	Refined petroleum production (Alberta)	5
LABOR MARKET			
142	EMP-NF	Total employment (New Foundland)	5
143	EMP-SERV-NF	Services employment (New Foundland)	5
144	EMP-FOR-OIL-NF	Forestry,oil,etc. employment (New Foundland)	5
145	EMP-CONS-NF	Construction employment (New Foundland)	5
146	EMP-SALES-NF	Wholesale and retail trade employment (New Foundland)	5
147	EMP-FIN-NF	Finance, insurance and real estate production (New Foundland)	5
148	EMP-PEI	Total employment (Prince Edward Island)	5
149	EMP-SERV-PEI	Services employment (Prince Edward Island)	5
150	EMP-FOR-OIL-PEI	Forestry,oil,etc. employment (Prince Edward Island)	5
151	EMP-CONS-PEI	Construction employment (Prince Edward Island)	5
152	EMP-SALES-PEI	Wholesale and retail trade employment (Prince Edward Island)	5
153	EMP-FIN-PEI	Finance, insurance and real estate production (Prince Edward Island)	5
154	EMP-NS	Total employment (Nova Scotia)	5
155	EMP-SERV-NS	Services employment (Nova Scotia)	5
156	EMP-FOR-OIL-NS	Forestry,oil,etc. employment (Nova Scotia)	5
157	EMP-CONS-NS	Construction employment (Nova Scotia)	5
158	EMP-SALES-NS	Wholesale and retail trade employment (Nova Scotia)	5
159	EMP-FIN-NS	Finance, insurance and real estate production (Nova Scotia)	5
160	EMP-NB	Total employment (New Bruinswick)	5
161	EMP-SERV-NB	Services employment (New Bruinswick)	5
162	EMP-FOR-OIL-NB	Forestry,oil,etc. employment (New Bruinswick)	5
163	EMP-CONS-NB	Construction employment (New Bruinswick)	5
164	EMP-SALES-NB	Wholesale and retail trade employment (New Bruinswick)	5
165	EMP-FIN-NB	Finance, insurance and real estate production (New Bruinswick)	5
166	EMP-QC	Total employment (Quebec)	5
167	EMP-SERV-QC	Services employment (Quebec)	5
168	EMP-FOR-OIL-QC	Forestry,oil,etc. employment (Quebec)	5
169	EMP-CONS-QC	Construction employment (Quebec)	5

170	EMP-SALES-QC	Wholesale and retail trade employment (Quebec)	5
171	EMP-FIN-QC	Finance, insurance and real estate production (Quebec)	5
172	EMP-ONT	Total employment (Ontario)	5
173	EMP-SERV-ONT	Services employment (Ontario)	5
174	EMP-FOR-OIL-ONT	Forestry,oil,etc. employment (Ontario)	5
175	EMP-CONS-ONT	Construction employment (Ontario)	5
176	EMP-SALES-ONT	Wholesale and retail trade employment (Ontario)	5
177	EMP-FIN-ONT	Finance, insurance and real estate production (Ontario)	5
178	EMP-MAN	Total employment (Manitoba)	5
179	EMP-SERV-MAN	Services employment (Manitoba)	5
180	EMP-FOR-OIL-MAN	Forestry,oil,etc. employment (Manitoba)	5
181	EMP-CONS-MAN	Construction employment (Manitoba)	5
182	EMP-SALES-MAN	Wholesale and retail trade employment (Manitoba)	5
183	EMP-FIN-MAN	Finance, insurance and real estate production (Manitoba)	5
184	EMP-SAS	Total employment (Saskatchewan)	5
185	EMP-SERV-SAS	Services employment (Saskatchewan)	5
186	EMP-FOR-OIL-SAS	Forestry,oil,etc. employment (Saskatchewan)	5
187	EMP-CONS-SAS	Construction employment (Saskatchewan)	5
188	EMP-SALES-SAS	Wholesale and retail trade employment (Saskatchewan)	5
189	EMP-FIN-SAS	Finance, insurance and real estate production (Saskatchewan)	5
190	EMP-ALB	Total employment (Alberta)	5
191	EMP-SERV-ALB	Services employment (Alberta)	5
192	EMP-FOR-OIL-ALB	Forestry,oil,etc. employment (Alberta)	5
193	EMP-CONS-ALB	Construction employment (Alberta)	5
194	EMP-SALES-ALB	Wholesale and retail trade employment (Alberta)	5
195	EMP-FIN-ALB	Finance, insurance and real estate production (Alberta)	5
196	EMP-BC	Total employment (British Columbia)	5
197	EMP-SERV-BC	Services employment (British Columbia)	5
198	EMP-FOR-OIL-BC	Forestry,oil,etc. employment (British Columbia)	5
199	EMP-CONS-BC	Construction employment (British Columbia)	5
200	EMP-SALES-BC	Wholesale and retail trade employment (British Columbia)	5
201	EMP-FIN-BC	Finance, insurance and real estate production (British Columbia)	5
202	UNEMP-NF	Unemployment rate (New Foundland)	2
203	UNEMP-PEI	Unemployment rate (Prince Edward Island)	2
204	UNEMP-NS	Unemployment rate (Nova Scotia)	2
205	UNEMP-NB	Unemployment rate (New Brunswick)	2
206	UNEMP-QC	Unemployment rate (Quebec)	2
207	UNEMP-ONT	Unemployment rate (Ontario)	2
208	UNEMP-MAN	Unemployment rate (Manitoba)	2
209	UNEMP-SAS	Unemployment rate (Saskatchewan)	2
210	UNEMP-ALB	Unemployment rate (Alberta)	2
211	UNEMP-BC	Unemployment rate (British Columbia)	2
212	UNEMP-DURAvg-NF	Unemployment duration (average) (New Foundland)	5
213	UNEMP-DURAvg-PEI	Unemployment duration (average) (Prince Edward Island)	5
214	UNEMP-DURAvg-NS	Unemployment duration (average) (Nova Scotia)	5
215	UNEMP-DURAvg-NB	Unemployment duration (average) (New Brunswick)	5
216	UNEMP-DURAvg-QC	Unemployment duration (average) (Quebec)	5
217	UNEMP-DURAvg-ONT	Unemployment duration (average) (Ontario)	5
218	UNEMP-DURAvg-MAN	Unemployment duration (average) (Manitoba)	5
219	UNEMP-DURAvg-SAS	Unemployment duration (average) (Saskatchewan)	5
220	UNEMP-DURAvg-ALB	Unemployment duration (average) (Alberta)	5
221	UNEMP-DURAvg-BC	Unemployment duration (average) (British Columbia)	5
MANUFACTURING, SALES AND INVENTORIES			
222	MANU-NF-new	Manufacturing total inventories (New Foundland)	5
223	DUR-NF-new	Durable manufacturing inventories (New Foundland)	5
224	MANU-PEI-new	Manufacturing total inventories (Prince Edward Island)	5
225	DUR-PEI-new	Durable manufacturing inventories (Prince Edward Island)	5
226	MANU-NS-new	Manufacturing total inventories (Nova Scotia)	5
227	DUR-NS-new	Durable manufacturing inventories (Nova Scotia)	5
228	MANU-NB-new	Manufacturing total inventories (New Brunswick)	5
229	DUR-NB-new	Durable manufacturing inventories (New Brunswick)	5
230	MANU-QC-new	Manufacturing total inventories (Quebec)	5
231	DUR-QC-new	Durable manufacturing inventories (Quebec)	5
232	MANU-ONT-new	Manufacturing total inventories (Ontario)	5
233	DUR-ONT-new	Durable manufacturing inventories (Ontario)	5
234	MANU-MAN-new	Manufacturing total inventories (Manitoba)	5
235	DUR-MAN-new	Durable manufacturing inventories (Manitoba)	5
236	MANU-SAS-new	Manufacturing total inventories (Saskatchewan)	5
237	DUR-SAS-new	Durable manufacturing inventories (Saskatchewan)	5
238	MANU-ALB-new	Manufacturing total inventories (Alberta)	5
239	DUR-ALB-new	Durable manufacturing inventories (Alberta)	5
240	MANU-BC-new	Manufacturing total inventories (British Columbia)	5
241	DUR-BC-new	Durable manufacturing inventories (British Columbia)	5
PRICES			
242	CPI-ALL-NF	CPI all (New Foundland)	5
243	CPI-MINUS-FEN-NF	CPI minus food and energy (New Foundland)	5
244	CPI-ALL-PEI	CPI all (Prince Edward Island)	5
245	CPI-MINUS-FEN-PEI	CPI minus food and energy (Prince Edward Island)	5
246	CPI-ALL-NS	CPI all (Nova Scotia)	5
247	CPI-MINUS-FEN-NS	CPI minus food and energy (Nova Scotia)	5
248	CPI-ALL-NB	CPI all (New Brunswick)	5
249	CPI-MINUS-FEN-NB	CPI minus food and energy (New Brunswick)	5
250	CPI-ALL-QC	CPI all (Quebec)	5
251	CPI-MINUS-FEN-QC	CPI minus food and energy (Quebec)	5
252	CPI-ALL-ONT	CPI all (Ontario)	5
253	CPI-MINUS-FEN-ONT	CPI minus food and energy (Ontario)	5
254	CPI-ALL-MAN	CPI all (Manitoba)	5
255	CPI-MINUS-FEN-MAN	CPI minus food and energy (Manitoba)	5
256	CPI-ALL-SAS	CPI all (Saskatchewan)	5
257	CPI-MINUS-FEN-SAS	CPI minus food and energy (Saskatchewan)	5
258	CPI-ALL-ALB	CPI all (Alberta)	5
259	CPI-MINUS-FEN-ALB	CPI minus food and energy (Alberta)	5
260	CPI-ALL-BC	CPI all (British Columbia)	5
261	CPI-MINUS-FEN-BC	CPI minus food and energy (British Columbia)	5

262	CPI-GOOD-NF	CPI goods (New Foundland)	5
263	CPI-SERV-NF	CPI services (New Foundland)	5
264	CPI-GOOD-PEI	CPI goods (Prince Edwards Island)	5
265	CPI-SERV-PEI	CPI services (Prince Edwards Island)	5
266	CPI-GOOD-NS	CPI goods (Nova Scotia)	5
267	CPI-SERV-NS	CPI services (Nova Scotia)	5
268	CPI-GOOD-NB	CPI goods (New Bruinswick)	5
269	CPI-SERV-NB	CPI services (New Bruinswick)	5
270	CPI-GOOD-QC	CPI goods (Quebec)	5
271	CPI-SERV-QC	CPI services (Quebec)	5
272	CPI-GOOD-ONT	CPI goods (Ontario)	5
273	CPI-SERV-ONT	CPI services (Ontario)	5
274	CPI-GOOD-MAN	CPI goods (Manitoba)	5
275	CPI-SERV-MAN	CPI services (Manitoba)	5
276	CPI-GOOD-SAS	CPI goods (Saskatchewan)	5
277	CPI-SERV-SAS	CPI services (Saskatchewan)	5
278	CPI-GOOD-ALB	CPI goods (Alberta)	5
279	CPI-SERV-ALB	CPI services (Alberta)	5
280	CPI-GOOD-BC	CPI goods (British Columbia)	5
281	CPI-SERV-BC	CPI services (British Columbia)	5