Testing and evaluation of Dynamic Factor Models for Canadian GDP forecasting

Ismael Diamoutene

Project report submitted to the Faculty of Graduate and Postdoctoral Studies

In partial fulfilment of the requirements for the MSc. degree in Mathematics and Statistics[1]

Department of Mathematics and Statistics

Faculty of Science

University of Ottawa.

June 29, 2022, Ottawa, Canada

[1]The MSc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

This report summarizes our findings, from an exploratory analysis on Statistics Canada's internal data, with regard to the usefulness of dynamic factor models for Canadian GDP forecasting. We evaluate 3 aspects of the model: its goodness of fit, its empirical forecasting performance and whether it can be used for variable selection. In particular, we noticed that for all these aspects, there are better alternatives to a dynamic factor model. We use Principal Component Analysis and Time Series Regression as benchmarks to evaluate the model.

Acknowledgements

I would like to thank my supervisor, Professor Rafal Kulik, for its guidance, its patience, and for entrusting me with this project. I would also like to thank the team at Statistics Canada for their continuous feedbacks and insightful comments.

I'm very grateful to my family and friends, whose presence and support helped me cope with the challenges I encountered during my Master's.

Contents

List of Figures

Introduction

A Dynamic Factor Model (DFM) is a dimensionality reduction technique. The assumption behind the model is that the dynamics in large multivariate time series can be explained by a small number of unobservable time series called factors. Dynamic Factor Models have gained prominence this last few years and showed encouraging results for GDP forecasting ([17], [18]).

In this project we deal with a dataset provided by Statistics Canada. The goal is to use a multivariate time series consisting of many economical variables to forecast the Canadian GDP using a Dynamic Factor Model (DFM). The main issue with the available data is non-stationarity.

Through exploratory data analysis, we tried to gain a better understanding of the model by answering the following questions:

- Should we use dynamic factor models instead of simpler alternatives? (comparison with Principal Component Analysis and Time Series Regression).
- How well do the factors describe the GDP? (Goodness of fit).
- How well does the model work when it comes to forecasting? (forecasting accuracy and robustness).
- How does the number of factors impact the estimation and forecasting performances of the model?
- Is the model useful for variable (predictors) selection?
- How does the performance of the model vary with some characteristics of the dataset (time dependence, level of noise...)?

We use Principal Component Analysis (PCA) and Time Series Regression (TS Regression) as benchmarks to evaluate the Dynamic Factor Model.

Our analysis is mainly based on the data provided by Statistics Canada, but we eventually use some simulated data to hypothesize on some properties of the model and make some suggestions.

The goal of this project is not to demonstrate some general theoretical properties of dynamic factor models. Our objective is rather to evaluate, in the specific context of forecasting the Canadian GDP, with the data available and the DFM we consider, whether Statistics Canada should move forward with adopting dynamic factor models. The conclusions we draw are highly dependent on the dataset we use, as well as the assumptions we make and the estimation method we use.

We proceed as follows in our analysis. First, we give a detailed description of the model, its estimation, and the forecasting approaches we consider (Section 1). Next, we describe and explore the provided dataset (Section 2). We then tackle the goodness of fit of the model, that is how well it describes the movements of the GDP (Section 3). After that we evaluate the empirical forecasting performance of DFM (Section 4). Finally we discuss the use of dynamic factor models for variable selection (Section 5).

The main conclusions of our analysis are as follows:

- For the Canadian dataset, the DFM does not have a better forecasting performance than the benchmarks we consider.
- As a dimensionality reduction tool, PCA might be preferable to DFM.
- The DFM has no mechanism to select the variables best suited for the model.
- Some characteristics of the original dataset, like a strong dependence in time of the features, could translate in a better forecasting performance of the factors.

All the code used for this project is available in the following Github repository: `https://github.com/ismod/Dynamic_factor_model_CANGDP`.

1 DFM: Model description, estimation and forecasting

1.1 Model description

We consider a simpler and restricted version of the dynamic factor model described in [3].

$$X_t = \Lambda F_t + \xi_t, \tag{1}$$

$$F_t = \sum_{i=1}^{p} A_i F_{t-i} + U_t, \tag{2}$$

where:

- $t = 1, ..., n$.
- X_t $(N \times 1)$ contains the observed features at time t; F_t $(r \times 1)$ contains the unobserved factors at time t.
- Λ is a $N \times r$ deterministic matrix; A_i, $i = 1, ..., p$ are $r \times r$ deterministic matrices.
- ξ_t $(N \times 1)$ and U_t $(r \times 1)$ contain the error terms for X_t and F_t respectively.
- $\xi_t \sim iid\ N(0, \Psi)$, with $\Psi = Diag(\psi_1, ..., \psi_N)$ being a $N \times N$ covariance matrix. It is assumed that $\xi_t, t = 1, ...n$ are independent in time and space.
- $U_t \sim iid\ N(0, Q)$, with Q being a $r \times r$ covariance matrix.
- ξ_t and U_t are independent processes: $E(\xi_t U_{t-s}^T) = 0$ for all s. It is assumed that $U_t, t = 1, .., n$ are independent in time (but not necessarily in space).
- We assume that X_t and F_t are stationary and zero mean.

The observed features $X_1, ..., X_N$ are reduced into a small number (r) of factors $F_1, ..., F_r$ which are then used to predict Y (Section 1.3).

The matrix form of the model is:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \\ . \\ . \\ . \\ X_{Nt} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & . & . & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & . & . & \lambda_{2r} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \lambda_{N1} & \lambda_{N2} & . & . & \lambda_{Nr} \end{bmatrix} \begin{bmatrix} F_{1t} \\ F_{2t} \\ . \\ . \\ F_{rt} \end{bmatrix} + \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ . \\ . \\ . \\ \xi_{Nt} \end{bmatrix},$$

$$\begin{bmatrix} F_{1t} \\ F_{2t} \\ . \\ . \\ F_{rt} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} A_{i,11} & A_{i,12} & . & . & A_{i,1r} \\ A_{i,21} & A_{i,22} & . & . & A_{i,2r} \\ . & . & . & . & . \\ . & . & . & . & . \\ A_{i,r1} & A_{i,r2} & . & . & A_{i,rr} \end{bmatrix} \begin{bmatrix} F_{1t-i} \\ F_{2t-i} \\ . \\ . \\ F_{rt-i} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ . \\ . \\ U_{rt} \end{bmatrix}.$$

1.2 Model Estimation

There are many estimation methods proposed in the literature for dynamic factor models. The method we'll use is the two-stage approach described in [3]. The first step of this approach uses Principal Component Analysis (PCA) to estimate the factors. A Vector Autoregressive (VAR) model is then fitted to those estimated factors. In the second step, the estimated parameters from the VAR on the principals components are considered as true parameters of the model. An algorithm called State Smoothing is then used to infer the values of the factors from the estimated parameters and the observed values X_t.

This estimation method is implemented in two R packages that we'll use interchangeably in our analysis: the dynfactoR package [1] and the nowcasting package [2]. While the former is faster, the latter has more built-in methods. The implementation of the two-stage estimation is exactly the same in both packages.

Step 1: PCA (Principal Component Analysis)

In this step, the factors $F_{1t}, ..., F_{rt}$ are estimated as the first r principal components of X_t.

- Let's rewrite the model as $X = F\Lambda^T + \xi$, where:

- $X = \begin{bmatrix} X_1 & X_2 & . & . & . & X_N \end{bmatrix} \quad (X \; is \; n \times N)$,
- $F = \begin{bmatrix} F_1 & F_2 & . & . & . & F_r \end{bmatrix} = \begin{bmatrix} F_{11} & F_{11} & . & . & F_{r1} \\ F_{12} & F_{22} & . & . & F_{r2} \\ . & . & . & . & . \\ . & . & . & . & . \\ F_{1n} & F_{2n} & . & . & F_{rn} \end{bmatrix}$,
- $\xi = \begin{bmatrix} \xi_1 & \xi_2 & . & . & . & \xi_N \end{bmatrix}$.

- The matrix X has to be standardized so that each column is zero mean and have a variance of 1.F or the remaining, X will refer to the standardized X.

- In PCA, X is mapped to a low dimensional representation X_{proj} that captures most of its variance. More precisely,

$$X_{proj} = \begin{bmatrix} X_1^T v_1 & X_1^T v_2 & . & . & X_1^T v_r \\ X_2^T v_1 & X_2^T v_2 & . & . & X_2^T v_r \\ . & . & . & . & . \\ . & . & . & . & . \\ X_n^T v_1 & X_n^T v_2 & . & . & X_n^T v_r \end{bmatrix},$$

where:

- v_i is $N \times 1$,

- $\hat{var}(Xa) = \frac{1}{n}\sum_{i=1}^{n}(X_i^T a)^2$,

- $v_1 = \underset{\|a\|_2=1}{\text{argmax}}\ \hat{var}(Xa)$ (where $\|.\|_2$ is the Euclidian norm),

- $v_2 = \underset{\|a\|_2=1}{\text{argmax}}\ \hat{var}(Xa) \quad$ such that $a^T v_1 = 0$,

...

- $v_r = \underset{\|a\|_2=1}{\text{argmax}}\ \hat{var}(Xa) \quad$ such $a^T v_1 = 0, a^T v_2 = 0, ..., a^T v_{r-1} = 0$,

- $X_{proj} = \begin{bmatrix} PC_1 & PC_2 & . & . & . & PC_r \end{bmatrix}$ is a $n \times r$ matrix where each column is called a principal component.

A principal component is a projection of the original dataset on a 1 dimensional space. The i^{th} principal component is the projection, orthogonal to the others, which has the i^{th} largest variance (and is said to have the i^{th} largest contribution to the variability of the dataset).

- It can be shown that v_i is the eigenvector corresponding to the i^{th} largest eigenvalue of the variance-covariance matrix of X_t $(X^T X)$.

Step 2: Parameters estimation

-The estimate $\hat{\Lambda}^T$ of Λ^T in Eq. (1) is the Ordinary Least Squares (OLS) estimate of the parameters of the linear regression of X against X_{proj}. This estimate is

$$\hat{\Lambda}^T = (X_{proj}^T X_{proj})^{-1} X_{proj}^T X$$

.

- The estimate $\hat{\xi}$ of ξ in Eq. (1) corresponds to the residuals of the above regression, i.e

$$\hat{\xi} = X - X_{proj}\hat{\Lambda}^T$$

.

-Ψ is estimated using the sample covariance matrix of ξ. That is

$$\hat{\Psi} = diag(\frac{1}{n}\sum_{t=1}^{n}(X_{t\cdot} - \hat{\Lambda}X_{proj}[t,])(X_{t\cdot} - \hat{\Lambda}X_{proj}[t,])^T),$$

where $X_{proj}[t,]$ is the t^{th} row of X_{proj}. All off-diagonal elements are set to 0 in order to satisfy the assumption of independence in space.

-The estimates of A_i, i=1,...,p in Eq. (2) are obtained from the OLS estimate of the linear regression of $X_{proj}[t,]$ against $X_{proj}[t-1,], ..., X_{proj}[t-p,]$. Let

$$\beta = [A_1[1,].., A_1[r,], A_2[1,].., A_2[r,], ..., A_p[1,].., A_p[r,]],$$

$$f_{(t)} = \begin{bmatrix} X_{proj}[n,] \\ X_{proj}[n-1,] \\ . \\ . \\ X_{proj}[p+1,] \end{bmatrix},$$

and let $f_{(t-1)}$ be the matrix which columns are

$$\begin{bmatrix} X_{proj}[n-1,] \\ X_{proj}[n-2,] \\ . \\ . \\ X_{proj}[p,] \end{bmatrix}, \begin{bmatrix} X_{proj}[n-2,] \\ X_{proj}[n-3,] \\ . \\ . \\ X_{proj}[p-1,] \end{bmatrix}, ..., \begin{bmatrix} X_{proj}[n-p,] \\ X_{proj}[n-(p+1),] \\ . \\ . \\ X_{proj}[1,] \end{bmatrix}.$$

.

Then the OLS estimate of β is

$$\hat{\beta} = (f_{(t-1)}^T f_{(t-1)})^{-1} f_{(t-1)}^T f_{(t)}$$

and $\hat{A}_i$ is obtained from the elements of $\hat{\beta}$.

- The estimate $\hat{Q}$ of Q is the sample covariance matrix of the residuals of the above regression.

Step 3: State smoother

In this step, the parameters estimated at the previous step are considered as the true parameters of the model, and the factors are reestimated using the state smoother algorithm. The model can be written in the following form:

$$X_t = Z\alpha_t + \xi_t$$

$$\alpha_{t+1} = T\alpha_t + \eta_t$$

where

- $t = 1, ..., n$
- $\xi_t \sim N(0, \hat{\Psi})$, $\eta_t \sim N(0, \hat{Q})$
- $X_t = [X_{1t}, ..., X_{Nt}]^T$
- $Z = [\hat{\Lambda}\ 0_{N\times r}\ ...\ 0_{N\times r}]$ is $N \times (r \times p)$
- $\alpha_t = [F_{1t}, F_{2t}, ..., F_{rt}, F_{1t-1}, ..., F_{rt-1}, ..., F_{1t-p+1}, ..., F_{rt-p+1}]^T$ is $r \times p$
- $\xi_t = \begin{bmatrix} \xi_{1t} & \xi_{2t} & . & . & . & \xi_{Nt} \end{bmatrix}^T$
- $T = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 & . & . & . & \hat{A}_p \\ I_r & 0_{r\times r} & . & . & . & 0_{r\times r} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0_{r\times r} & 0_{r\times r} & . & . & I_r & 0_{r\times r} \end{bmatrix}$ is $(r \times p) \times (r \times p)$

- $\eta_t = [U_{1t}, ..., U_{rt}, 0, ..., 0]^T$ is $r \times p$
- $\Sigma = \begin{bmatrix} \hat{Q} & 0_{r\times r} & . & . & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} & . & . & 0_{r\times r} \\ . & . & . & . & . \\ . & . & . & . & . \\ 0_{r\times r} & 0_{r\times r} & . & . & 0_{r\times r} \end{bmatrix}$ is $(r \times p) \times (r \times p)$

The estimate $\hat{\alpha}_t$ of α_t is given by

$$\hat{\alpha}_t = E[\alpha_t | X_1, ..., X_n],$$

where the expectation is taken under the above normality assumptions.

In [4], the authors show that under the assumption that $\alpha_1 \sim N(a_1, P_1)$, these expected values can be computed recursively using the following algorithm:

Step 1: forward pass

Compute $v_t, E_t, K_t, L_t, a_t, P_t$ for $t = 1, ..., n$ using the following process:

For $t = 1, .., n$:

- $v_t = X_t - Za_t$,
- $E_t = ZP_tZ^T + \hat{\Psi}$,
- $K_t = TP_tZ^TE_t^{-1}$,
- $L_t = T - K_tZ$,
- $a_{t+1} = Ta_t + K_tv_t$,
- $P_{t+1} = TP_t(T - K_tZ)^T + \Sigma$.

Step 2: Backward pass

Compute $\hat{\alpha}_t, t = 1, ..., n$ using the following process:

Initialize $r_n = 0$

For $t = n, n-1, ..., 1$ (we go backward starting at $t = n$)

- $r_{t-1} = Z^T E_t^{-1} v_t + L_t^T E_t$,

- $\hat{\alpha}_t = a_t + P_t r_{t-1}$.

a_1 is initialized as

$$[X_{proj}[1,t], X_{proj}[2,t], ..., X_{proj}[r,t], X_{proj}[1,t-1], ..., X_{proj}[r,t-p+1]]^T,$$

and P_1 is initialized as the unconditional covariance matrix of α_t (which is a VAR(1) process) under the assumption that α_t is zero mean and weakly stationary [16], that is P_1 is obtained by solving

$$vec(P_1) = (I - T \otimes T)^{-1} vec(\Sigma),$$

where $vec()$ is the vectorization operator and $\otimes$ is the Kronecker product.

1.3 Forecasting approaches

Let V_t be the observed features, the factors or the principal components at time t. Let Y_t be the dependent variable at time t. We consider two models to describe the relationship of Y_t to V_t and we forecast the value of the dependent variable according to these models.

Approach 1: Direct approach

The data generating process for Y_t is assumed to be the following:

$$Y_t = a_1^T V_{t-1} + a_2^T V_{t-2} + ... + a_{p_y}^T V_{t-p_y} + e_t. \tag{3}$$

- Y_t is modeled as a multiple linear regression against $V_{t-1}, V_{t-2}, ..., V_{t-p_y}$.

- $V_t = [V_{1t}, ..., V_{dt}]^T$ is a d dimensional vector. If V_t corresponds to the features, then $d = N$. If we consider r factors or principal components, then $d = r$.

- $a_i = [a_{i1}, ..., a_{id}]^T$, $i = 1, ..., p_y$ are deterministic d dimensional vectors.

- The error terms e_t are iid with $E(e_t) = 0$. We assume that e_t and V_{t-i} are independent $\forall i \in \mathbb{Z}$.

- $Var(e_t) = \sigma_e^2$.

Let $\hat{V}_t, t = 1, ..., n$ be the estimated factors, the estimated principal components, or the observed features. Let $Y_t, t = 1, ..., n$ be the observed values of the dependent variable. Assume that p_y is known. Then the estimates $\hat{a}_i^T$ of $a_i^T, i = 1, ..., p_y$ are obtained from the OLS regression of Y_t against $\hat{V}_{t-1}, \hat{V}_{t-2}, ..., \hat{V}_{t-p_y}$. The predicted value $\hat{Y}_{n+1}$ of Y_t is given by

$$\hat{Y}_{n+1} = \hat{a}_1^T \hat{V}_{t-1} + \hat{a}_2^T \hat{V}_{t-2} + ... + \hat{a}_{p_y}^T \hat{V}_{t-p_y}.$$

Approach 2: Indirect approach

We consider the following data generating process for Y_t:

$$\begin{cases} V_t = B_1 V_{t-1} + B_2 V_{t-2} + ... + B_{p_v} V_{t-p_v} + \varepsilon_t, \\ Y_t = a^T V_t + e_t. \end{cases} \tag{4}$$

- V_t is modeled as a vector autoregressive process of order p_v (VAR(p_v)) and Y_t as a multiple linear regression against V_t.

- $V_t = [V_{1t}, ..., V_{dt}]^T$ is a d dimensional vector. If V_t corresponds to the features, then $d = N$. If we consider r factors or principal components, then $d = r$.

- $a = [a_1, ..., a_d]^T$ is a d dimensional deterministic vector. $B_1, B_2, ..., B_{p_v}$ are $d \times d$ deterministic matrices.

- $\varepsilon_t = [\varepsilon_{1t}, ..., \varepsilon_{dt}]^T$ and e_t are error terms for V_t and Y_t respectively.

- e_t and ε_t are independent; e_t are iid with $E(e_t) = 0$; ε_t are iid with $E(\varepsilon_t) = 0$.

- e_t and V_{t-i} are independent $\forall i \in \mathbb{Z}$. ε_t and V_{t-j} are independent for $j = 1, 2, 3, ...$

- $Var(\varepsilon_{it}) = \sigma^2_{\varepsilon_i}$; $Var(e_t) = \sigma^2_e$.

Let $\hat{V}_t, t = 1, ..., n$ be the estimated factors, the estimated principal components, or the observed features. Let $Y_t, t = 1, ..., n$ be the observed values of the dependent variable. Assume that p_v is known. The estimate $\hat{a}^T$ of a^T is obtained from the OLS regression of Y_t against $\hat{V}_t$. The estimates $\hat{B}_i$ of $B_i, i = 1, ..., p_v$ are obtained from the OLS regression of $\hat{V}_t$ against $\hat{V}_{t-1}, \hat{V}_{t-2}, ..., \hat{V}_{t-p_v}$.

The forecasting of the dependent variable is done in two steps. First, we obtain a forecast for V_t:

$$\hat{V}_{n+1} = \hat{B}_1^T \hat{V}_{t-1} + \hat{B}_2^T \hat{V}_{t-2} + ... + \hat{B}_{p_v}^T \hat{V}_{t-p_v}.$$

The next value of Y_t is then predicted as

$$\hat{Y}_{n+1} = \hat{a}^T \hat{V}_{n+1}.$$

For each method (DFM, PCA, Time Series Regression), we'll evaluate the forecasting with these two approaches (by replacing $\hat{V}_t$ by the estimated factors, the estimated principal components and the observed features respectively). We'll only focus on one-step ahead forecasting.

2 Exploring the dataset

- Statisitcis Canada provided the monthly Canadian GDP from January 2005 to February 2020. They also provided 17 additional monthly time series as potential predictors of the GDP. These series are labeled Y1_HOURS, X1_HOURSMAN, X2_HOURSWSRT, Y2_CPI, Y3_EMP, Y4_USCAN, Y5_ROWCAN, Y6_CANUS, Y7_CANROW, Y8_RAIL, Y9_DOM, Y10_TRAN, Y11_INT, Y12_MSM, Y13_MRTS, Y14_IMP, Y15_EXP. They range from January 2005 to February 2020. What each series represents was not disclosed by Statistics Canada.

2.1 Transforming and making the series stationary

- After visualizing the dataset, we noticed that the series Y3_EMP, Y9_DOM, Y10_TRAN and Y11_INT displayed so strong seasonalities that the stationary part was quasi non-existent as displayed on Figure 1 below. We want to focus only on stationary time series, so we removed these series. We're now left with 13 predictor variables.

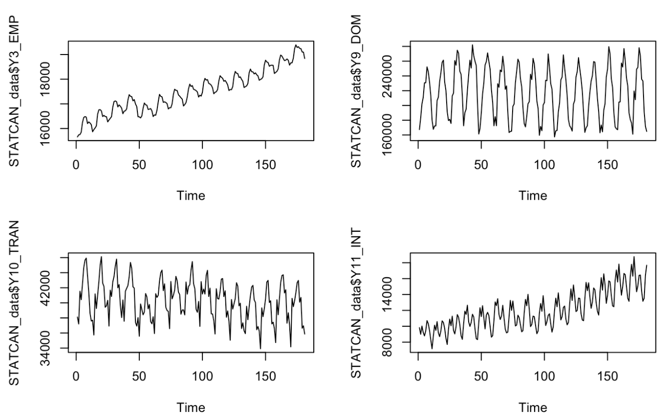

Figure 1: Seasonal time series removed from the dataset

- For each of the remaining predictors and the GDP, we apply the following transformation $x_t \sim (x_t - x_{t-1})/x_{t-1}$. In other words, we now focus on the growth rate of the series rather than the actual series. The advantage of this transformation is that the time series become stationary and the values of all our series are in the same range. It is also easy to come back to the original series: by multiplying the predicted growth rate by the current value of GDP and adding the result back to the GDP, we get a forecast. After this transformation, we subtract the mean for each of our series in order for them to be zero mean. The mean can be added back later when needed.

Below on Figure 2, we plot some of the transformed series.

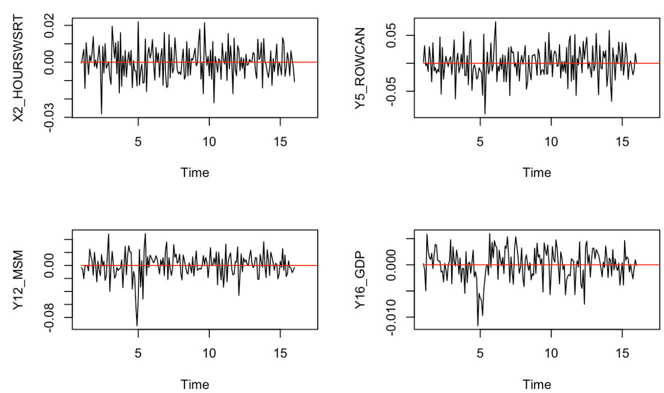

Figure 2: Some transformed series

From now on, whenever we talk about the Canadian dataset, we refer to this transformed dataset.

2.2 Verifying vector autoregressive assumptions

- When describing the indirect approach to forecasting (see Section 1.3), we modelled the features as a vector autoregressive process. There are many reasons why we made this choice: it takes into account the dependence in time and space of the features, it is relatively easy to simulate, it offers a possibility for comparison with the dynamic factor model in which the factors are also modelled as a vector autoregressive (VAR).

To make sure that this assumption is reasonable, we fit a VAR to our predictors.

To that end, we first select an autoregressive order. We use 4 information criteria: the Akaike information criterion (AIC), the Hannan–Quinn information criterion (HQ), the Schwartz information criterion (SC) and the Final Prediction Error criterion (FPE). For each of these information criterions, we display below in Figure 3 their values for different autoregressive orders as well as the order for which the criterion is minimized.

```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
    10      1      1      1

$criteria
                   1             2             3             4             5             6             7             8             9            10
AIC(n) -1.097679e+02 -1.093516e+02 -1.093917e+02 -1.089058e+02 -1.090054e+02 -1.091150e+02 -1.097305e+02 -1.109723e+02 -1.121653e+02 -1.145224e+02
HQ(n)  -1.084112e+02 -1.067350e+02 -1.055153e+02 -1.037695e+02 -1.026093e+02 -1.014590e+02 -1.008147e+02 -1.007967e+02 -1.007298e+02 -1.018270e+02
SC(n)  -1.064242e+02 -1.029029e+02 -9.983810e+01 -9.624732e+01 -9.324197e+01 -9.024662e+01 -8.775721e+01 -8.589415e+01 -8.398223e+01 -8.323433e+01
FPE(n)  2.140267e-48  3.343955e-48  3.480254e-48  6.634414e-48  7.900226e-48  1.094797e-47  1.147858e-47  8.904254e-48  1.178670e-47  1.063036e-47
```

Figure 3: Some information criterions and their values for the Canadian dataset

For 3 of the criterions, the number of lags selected is 1, therefore, we fitted a VAR(1) to the predictors. We'll use a VAR(1) as well when modelling the dynamic factors and the principal components.

- We made three assumptions about the innovations in the VAR representation: they are zero mean, independent in time, and independent of the previous values of the features.

To evaluate whether or not a VAR(1) is a reasonable model for the observed features, we verified the above assumptions using the residuals of the VAR we fitted.

For the first assumption (zero mean), we plot the residuals of the fitted VAR for each predictor along with a red line at the level of their mean. In Figure 4 we display some of these series, we have similar pictures for the remaining features.

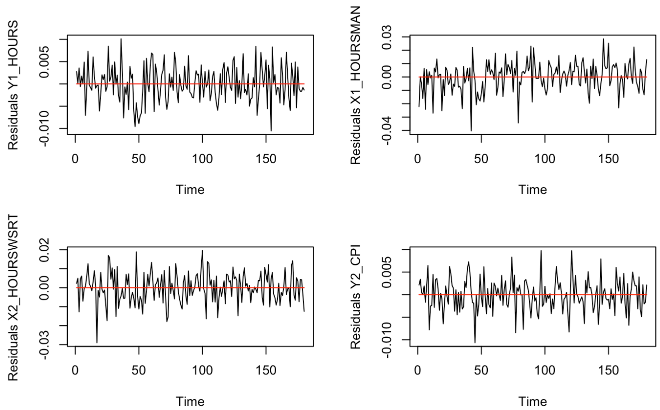

Figure 4: Residuals of fitted VAR for some features

The means are very close to zero and the residuals seem equally distributed around zero. It is therefore reasonable to assume that the innovations have a mean of zero.

To evaluate the second assumption (independence of innovations), we plot the sample autocorrelation and cross-correlation functions of the residuals. Below in Figure 5, we display some of these plots. The observations are similar for the remaining plots.

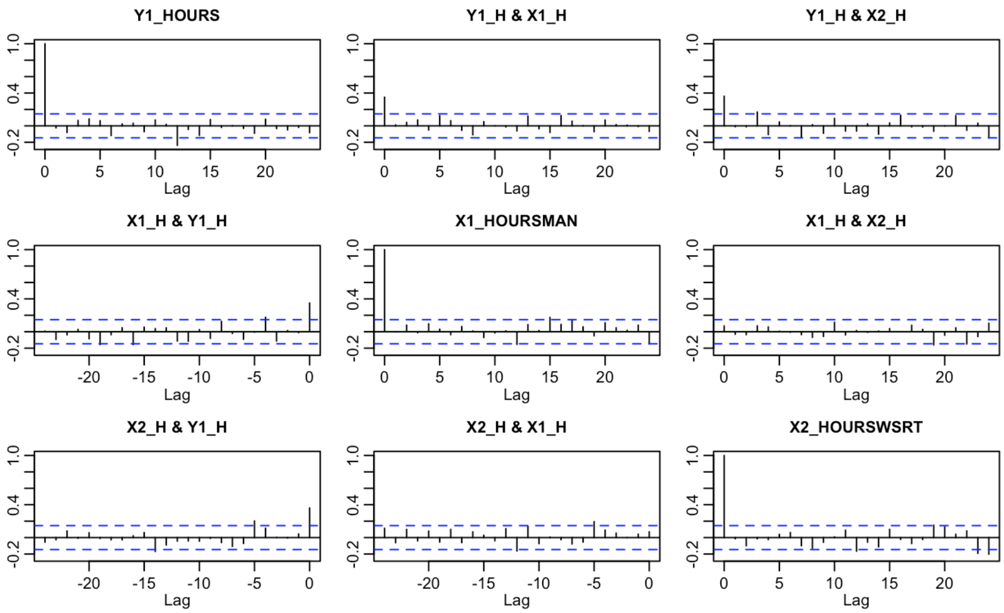

Figure 5: Sample autocorrelation and cross-correlation for the residuals of the fitted VAR

Starting at lag 1, the sample autocorrelations and cross-correlations of the residuals are very close to zero and they are not significant. Therefore, it is reasonable to assume that the innovations are independent in time.

For the third assumption (independence with previous values of the features), we plot in Figure 6 the sample cross-correlation function between the residuals at time t and the observed features at time $t, t-1, t-2, \ldots$. We only display part of the results (the remaining plots are similar).

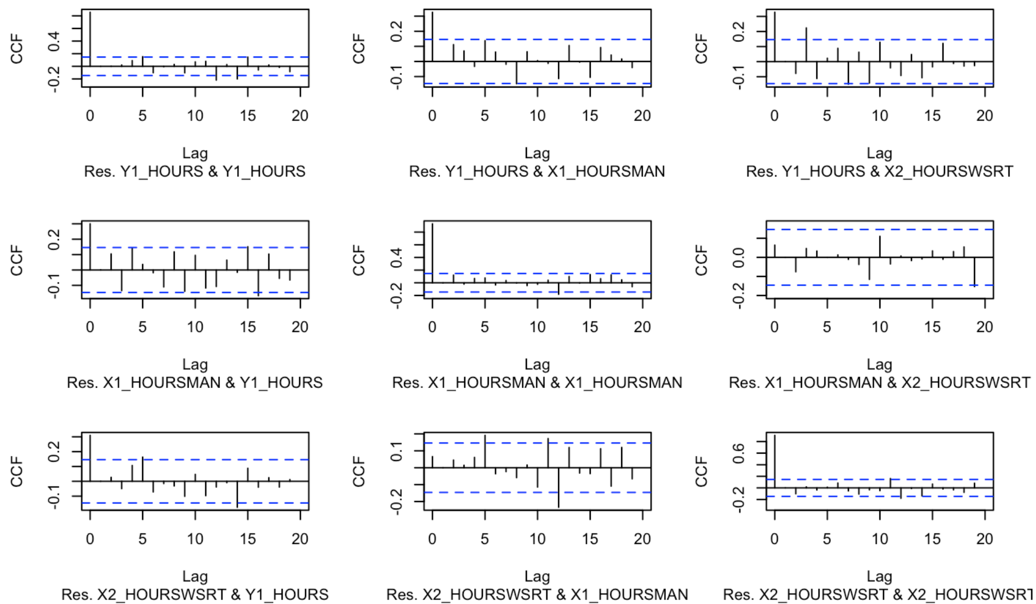

Figure 6: Cross-correlation between residuals and previous values of observed features

The general tendency is that from lag 1, the sample cross-correlation is not significant, which is agreement with our assumption.

- The three assumptions of the VAR on the features seem to be satisfied, so we can use this process to model the observed variables.

2.3 Selecting the number of lags for direct approach

The number of lags for prediction is the value of p_y (Eq. (3)) in the description of the direct approach to forecasting. Note that this is different from the autoregressive order p_v (Eq. (4)) of the VAR in the direct approach.

To select p_y, we regress the dependent variable Y_t against the features $X_{t-1}, X_{t-2}, ..., X_{t-p_y}$ for $p_y = 1, 2, 3, 4, 5$. We then get the value of the adjusted R^2 of the regression for each of these values of p_y (see Figure 7).

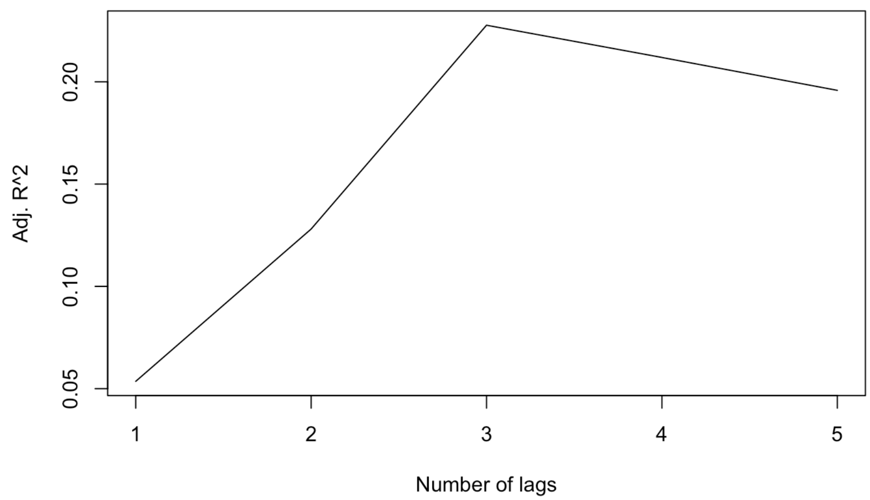

Figure 7: Adjusted R^2 of regression for different numbers of lags

The maximum adjusted R^2 is obtained with 3 lags (the values is 80% bigger than what is obtained with 2 lags). Therefore we will use $p_y = 3$ for the direct approach to forecasting.

3 Goodness of fit

Our objective in this section is to assess how the dynamic factors describe the movements of the dependent variable Y (in our case the GDP).

We use Time Series Regression and Principal Component Analysis as benchmarks since those are simple and widely used approaches. We first start the comparison on the Canadian data and then we use some simulated data to evaluate the goodness of fit when we vary certain aspects of the dataset like the dependence in time of the predictors and the variance of the error term.

We use the Adjusted R^2 as a measure of goodness of fit to account for the number of features used in the regressions.

3.1 Goodness of fit on Canadian data

To evaluate the goodness of fit, we need to consider the forecasting approaches we mentioned before:

- For the direct approach to forecasting (Eq. (3)), we evaluate how V_{t-1},V_{t-2},V_{t-3} explain the movements of Y_t . For that, we use the adjusted R^2 of the linear regression of Y_t against V_{t-1}, V_{t-2} and V_{t-3} (we selected $p_y = 3$ in section 2.3).

- For the second step of the indirect approach , we're interested on how V_t explains the value of Y_t. To evaluate this, we use the adjusted R^2 of the regression of Y_t against V_t. (The relation of V_{t-1} to V_t is also of importance for the first step).

For each forecasting approach, we do a comparison between the 3 models (Time Series Regression, PCA and DFM). For DFM and PCA, we consider different numbers of factors and principal components.

Let's start by the indirect approach. Below (Figure 8), we plot adjusted R^2 of the regression of Y_t against: the original features at time t (X_t), the principal components at time t, and the factors at time t (F_t).

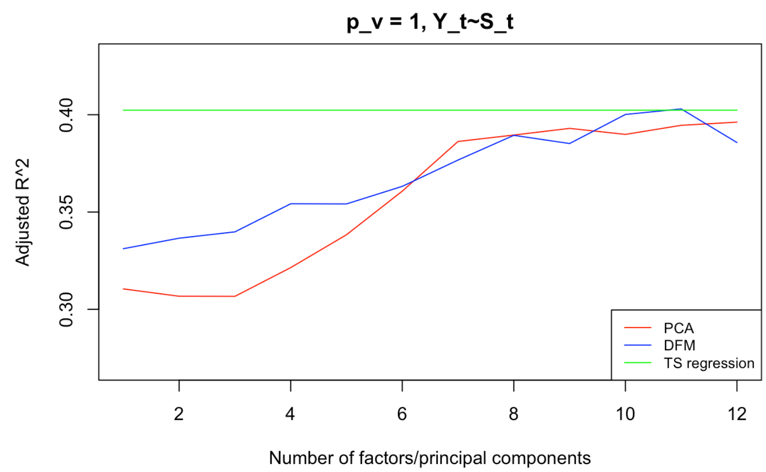

Figure 8: Adjusted R^2 of indirect approach for DFM, PCA and TS Regression

The first thing we observe is that the value of the adjusted R^2 using DFM and PCA is generally lower than the one obtained using the original features (13 variables). Starting at 7 principal components/factors the difference gets smaller and DFM achieves a similar value as TS Regression when 11 factors are used. Even at 8 factors/principal components, the adjusted R^2 for Time Series Regression is just 4% bigger than the value obtained for DFM and PCA, which we think is reasonable since we use less features for the regression.

We also observe that for low numbers of factors/ principal components DFM seem to outperform PCA when it comes to goodness of fit (the difference is relatively small), but for high numbers of factors/principal components, the two methods seem to yield similar results.

The scale of the plot can be misleading and further tests are needed if we want to assess the significance of these relative differences. But it is evident from the plot that with the right number of factors, dynamic factor models can achieve the same goodness of fit (as measured by the adjusted R^2) as PCA and Time Series Regression.

We do a similar analysis for the direct approach by plotting the adjusted R^2 of the regression of Y_t against: the original features at time $t-1$,$t-2$,$t-3$ (X_{t-1},X_{t-2},X_{t-3}), the principal components at time $t-1$,$t-2$,$t-3$,and the factors at time $t-1$,$t-2$,$t-3$ (F_{t-1},F_{t-2},F_{t-3}). We get the following plot:

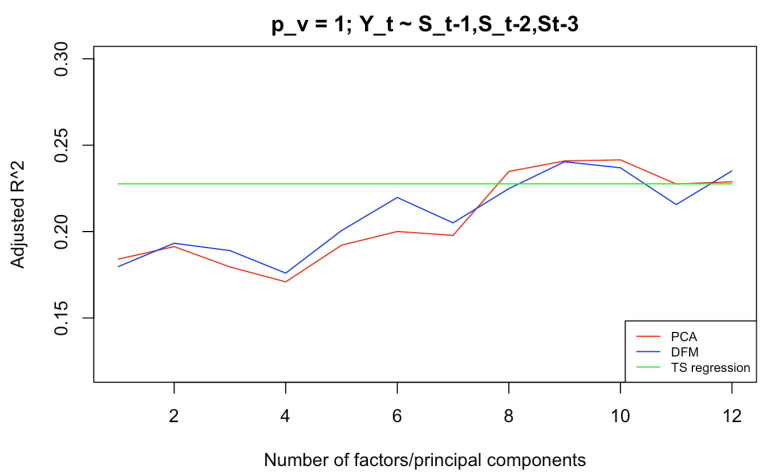

Figure 9: Adjusted R^2 of indirect approach for DFM, PCA and TS Regression

The conclusion is the same as the previous one: dynamic factor models achieve at least the same goodness fit as PCA and TS Regression when the number of factors is correctly chosen.

For this approach (when we take into account the scale of the plot), PCA and DFM seem to have almost the same result for each number of factors/principal components.

We can not conclude on whether we should use DFM or not solely on the basis of goodness of fit. A good fit doesn't necessarily translate in a good forecasting performance since there is a possibility of overfitting. On the other hand, a model with a relatively good forecasting accuracy but a bad fit for the data is not reliable, as those good results wouldn't be explainable and could just be due to luck.

So far, the results are encouraging for the dynamic factor model, in the sense that if we observe that the DFM has a better forecasting performance than the benchmarks, we can be confident in this observation as the model is a relatively good fit for the data.

Using simulations, we did some additional analysis to assess if the model is still a good fit (or is a better fit) when we modify some characteristics of the dataset (dependence in time of the features, variance of the noise). This could provide some useful information about the type of data that works best with DFM.

3.2 Simulation setup

The following process describes how we simulated the datasets we used for our analysis.

For each standard deviation σ in the set $\{0.0005; 0.001; 0.002; 0.005; 0.007; 0.01\}$:

- For each interval I in the set { (-0.1,0.1); (-0.2,0.2); (-0.3,0.3); (-0.4,0.4)}:
  - We do a Monte Carlo simulation with 20 replications. (The number of replications is relatively low because within the replications we generate 13×13 random coefficient matrices, 13×180 multivariate time series, and we fit the DFM (and PCA) to the simulated data, which involves two loops (See section 1.2). In addition, the outer loops constitute 20 combinations of coefficient matrices and standard deviations, so that we

have in total 400 replications. The computational burden of the simulation is therefore very high.)

For each replication:

* We randomly generate 169 values in the interval I and use these values to create a 13×13 coefficient matrix (same size as the coefficient matrix of the Canadian dataset).

* We then generate 180 observations (Canadian data sample size) from a VAR(1) using the above coefficient matrix and the same covariance matrix of innovations as the Canadian dataset. Let's call the generated series $TS_1, TS_2, TS_3, TS_4, TS_5, TS_6, TS_7, TS_8, TS_9, TS_{10}, TS_{11}, TS_{12}, TS_{13}$.

* Next, we generate 180 error terms $e_t, t = 1, ..., 180$ from a normal distribution with mean 0 and standard deviation σ.

* We use these error terms to construct the values of the dependent variable Y:

  · For the <u>indirect approach</u> to forecasting, we get Y_t from the following formula:
  $Y_t = a_1 \times TS_{1t} + a_2 \times TS_{2t} + a_3 \times TS_{3t} + a_4 \times TS_{4t} + a_5 \times TS_{5t} + a_6 \times TS_{6t} + a_7 \times TS_{7t} + a_8 \times TS_{8t} + a_9 \times TS_{9t} + a_{10} \times TS_{10t} + a_{11} \times TS_{11t} + a_{12} \times TS_{12t} + a_{13} \times TS_{13t} + e_t$,
  where $a_i, i = 1, ..., 13$ are obtained from regressing the dependent variable at time t against the predictors at time t in the transformed Canadian dataset.

  · For the <u>direct approach</u>, Y_t is obtained from the following formula:
  $Y_t = b_1^T TS_{t-1} + b_2^T TS_{t-2} + b_3^T TS_{t-3} + e_t$, where $b_i, i = 1, ..., 3$ are

obtained from regressing the dependent variable at time t against the predictors at time $t-1, t-2, t-3$ in the transformed Canadian dataset.

* We finally fit DFM and PCA to the simulated dataset and compute the following metrics: The adjusted R^2 for PCA and DFM, The ratio of DFM's adjusted R^2 to PCA's adjusted R^2. We compute these metrics for different numbers of factors/principal components.

For each standard deviation σ and coefficient matrix interval I, we average these metrics across all replications.

Note that we don't do the comparison with Time Series Regression. Since we specifically generate the values of Y_t from a linear regression against the simulated time series, the results could be biased toward TS Regression.

3.3 Goodness of fit and dependence in time

To simulate different levels of dependence in time, we randomly generate coefficient matrices for the VAR, with values in the following ranges: (-0.1,0.1); (-0.2,0.2); (-0.3,0.3); (-0.4,0.4). The larger the bounds (in absolute value) of the interval, the higher the level of dependence in time between the features. We do not go above 0.4, nor do we consider fixed values (let's say 0.3 for all the coefficients) because in these cases we would get some non-stationary series.

For illustration, we will only display the Adjusted R^2 results for a standard deviation of 0.002 of the error term. The conclusions obtained for this standard deviation are similar to what we obtained from the other values of the standard deviation. The results are displayed in Figures 10 to 13 at the end of section 3.

Indirect approach

Firt let's look at at the relative goodness of fit of PCA to DFM as measured by the ratio of DFM's adjusted R^2 to PCA's adjusted R^2 (Figure 10). All the ratios have values very close to 1, one model doesn't seem to stand out. The ratio doesn't seem to increase or decrease with the number of factors/principal components or the value range of the coefficient matrix. So the dynamic factor model doesn't seem to do better than PCA (in term of goodness of fit).

Now let's look at the goodness of fit of DFM only (Figure 11). We notice that the goodness of fit increases when the range of the coefficients increases. From that, we could hypothesize that DFM is better fit when we have strong dependence in time between the variables (and so is PCA since the results for the two methods are similar).

There's also a tendency of the adjusted R^2 to increase when the number of factors increases. Since the number of features increases, it makes sense that this is the case. This observation won't necessarily hold for forecasting since there's a risk of overfitting.

Direct approach

The ratios of DFM's adjusted R^2 to PCA's adjusted R^2 (Figure 12) are all close to 1, with no apparent tendency to increase or decrease with the range of the coefficient matrix or the number of factors. So both methods are equally matched (for goodness of fit).

Similarly to the indirect approach, the adjusted R^2 of the dynamic factor model (Figure 13) increases when the range of the coefficient matrix increases, suggesting that more dependence in time of the features could increase the goodness of fit DFM.

3.4 Goodness of fit and standard deviation of the error term

Now we look at the evolution of the goodness of fit when the standard deviation of the error term increases. By error term, we refer to the quantity e_t in equations (3) and (4) of section 1.3. We consider the following levels of standard deviation: 0.0005, 0.001, 0.002, 0.005, 0.007, 0.01.

For illustration, we will only display the adjusted R^2 results for coefficient matrices in the range (-0.3,0.3) . The conclusions obtained for this interval are similar to what we obtained for the others. The results are displayed in Figures 14 to 17 at the end of the section.

Indirect approach

When we look at only the goodness of fit of the dynamic factor model (Figure 15) , the adjusted R^2 decreases with the standard deviation of the error term. This result is obvious since more noise makes the dependent variable more difficult to predict.

What we're more interested in is whether DFM does better than PCA for high levels of noise. To that end, we look at the ratio of DFM's adjusted R^2 to PCA's adjusted R^2 (Figure 14). The observations are similar to what we had in section 3: no model seem to be better than the other (from the goodness of fit perspective).

Direct approach

We notice that the ratio of DFM's to PCA's adjusted R^2 (Figure 16) is almost constant, as the observed values are very close to 1. None of the two methods seem to be a better fit.

The adjusted R^2 of the dynamic factor model (Figure 17) decreases with the level of noise (as measured by the standard deviation of the error term), which is normal.

[1] "sd = 0.002"

Adj_R^2_DFM/Adj_R^2_PCA	coeff1 (-0.1,0.1)	coeff2 (-0.2,0.2)	coeff3 (-0.3,0.3)	coeff4 (-0.4,0.4)
1 PC/factor	1.05246420755263	0.216243705672287	1.05756668280855	1.0718616203013
2 PCs/factors	1.17077688830627	1.1191087029598	1.12055759051234	1.06059856503398
3 PCs/factors	1.07770457584191	1.0956812801754	1.11580131656071	1.02335169680036
4 PCs/factors	0.951764089753548	0.951135335980966	1.04650780231082	1.02180468693477
5 PCs/factors	0.943610763181161	0.926513290118673	0.982465917394529	1.00754122689025
6 PCs/factors	0.922896222466065	0.931892908134436	0.962583136936981	1.00334773828776
7 PCs/factors	0.905166730123653	0.934159645945686	0.955095342588692	0.979376053517802
8 PCs/factors	0.928789190012858	0.947388682763832	0.9418225606397	0.974582162404864
9 PCs/factors	0.996708410069887	0.947681389912006	0.961766091317266	0.982037507389665
10 PCs/factors	1.0278487697962	0.989661420278973	0.981157294860332	0.980842526371208
11 PCs/factors	1.04157058765625	1.00789595403083	0.984012291513216	0.995097492838155
12 PCs/factors	1.00581043149247	1.01025393415795	1.00325831431808	0.997312553068707

Figure 10: Ratio of adjusted R^2 (DFM/PCA) for different coefficient matrices (indirect approach)

[1] "sd = 0.002"

Adj_R^2_DFM	coeff1 (-0.1,0.1)	coeff2 (-0.2,0.2)	coeff3 (-0.3,0.3)	coeff4 (-0.4,0.4)
1 PC/factor	0.0828174553385422	0.116855125247126	0.185989901430907	0.300971765952486
2 PCs/factors	0.153915359696657	0.209124788000056	0.357030142711388	0.473792485241597
3 PCs/factors	0.257411926970787	0.28900408036682	0.397020021100992	0.607946310833743
4 PCs/factors	0.307546721588912	0.356415996114905	0.44933952607464	0.652472535456157
5 PCs/factors	0.344003492948903	0.41414016869022	0.52833542123868	0.702796016358539
6 PCs/factors	0.396745035967575	0.479320981073024	0.590808882609082	0.737763733943942
7 PCs/factors	0.435634683658525	0.516726404389751	0.626063754551101	0.764619953311908
8 PCs/factors	0.488490291573548	0.570729499338397	0.658991297548195	0.788209840723819
9 PCs/factors	0.554834010290546	0.625025387194408	0.701137762802058	0.844511184923642
10 PCs/factors	0.618672983481505	0.678444757293526	0.772751405668379	0.865623622139917
11 PCs/factors	0.682180718452198	0.749092112479989	0.804217803420233	0.905851498776188
12 PCs/factors	0.734761417823246	0.784972710729128	0.85160884672761	0.922613188130313

Figure 11: Adjusted R^2 of DFM for different coefficient matrices (indirect approach)

[1] "sd = 0.002"

Adj_R^2_DFM/Adj_R^2_PCA	coeff1 (-0.1,0.1)	coeff2 (-0.2,0.2)	coeff3 (-0.3,0.3)	coeff4 (-0.4,0.4)
1 PC/factor	1.0708687751087	1.15755368989497	1.00532403097556	1.0211860668264
2 PCs/factors	1.0402762307565	1.10822674885872	1.00265272222631	1.02982031298942
3 PCs/factors	1.03883354187522	1.04023679719874	1.03146909194283	1.00687705563337
4 PCs/factors	1.0504483828099	1.03311387580712	1.01649798292101	1.01523382890181
5 PCs/factors	1.04098989218414	1.02039328938051	1.01131858583803	1.00236452850367
6 PCs/factors	1.00572316830245	1.01384395079435	1.00604154357175	1.00023828558432
7 PCs/factors	1.00812012385627	0.996441307360737	0.999189750234895	1.00304207576188
8 PCs/factors	1.00341287997952	0.979876618306576	0.998326437503422	0.996786817084049
9 PCs/factors	0.989554629490792	0.965045088343647	0.993867300990482	0.997737783667299
10 PCs/factors	0.972520538095372	0.977151271563005	0.997193642028707	0.998299040518487
11 PCs/factors	0.987142061461063	0.986075609844049	0.997088203741668	0.996956105188052
12 PCs/factors	1.00189488842182	0.993107502947501	0.994801816609798	0.997778543283258

Figure 12: Ratio of adjusted R^2 (DFM/PCA) for different coefficient matrices (direct approach)

[1] "sd = 0.002"

Adj_R^2_DFM	coeff1 (-0.1,0.1)	coeff2 (-0.2,0.2)	coeff3 (-0.3,0.3)	coeff4 (-0.4,0.4)
1 PC/factor	0.170851182447667	0.183723671588303	0.380438951049382	0.563046918623491
2 PCs/factors	0.232147102810969	0.302967007717484	0.497419767489409	0.701864428297811
3 PCs/factors	0.279750560896466	0.395871325368932	0.566214480946563	0.783239448711042
4 PCs/factors	0.314008191636104	0.439217510461107	0.622888961438461	0.822837232032305
5 PCs/factors	0.344103738068718	0.473893669185044	0.663711361787613	0.854934151814644
6 PCs/factors	0.363100355429166	0.500867046230771	0.693494884961264	0.882075468148971
7 PCs/factors	0.395156820197461	0.52917226154544	0.716708042889877	0.897433376453752
8 PCs/factors	0.42568491167034	0.542241387869682	0.743358493801379	0.904933774056767
9 PCs/factors	0.455705093218701	0.561173836729889	0.757115053622983	0.9146594641661
10 PCs/factors	0.469044280256537	0.585831287534783	0.773468933058355	0.922403350185966
11 PCs/factors	0.49552313881038	0.616644936763141	0.787764537096852	0.930107306882775
12 PCs/factors	0.514164904551247	0.637901545664625	0.797228116084461	0.93368471738303

Figure 13: Adjusted R^2 of DFM for different coefficient matrices (direct approach)

[1] "coeff3 (-0.3,0.3)"

Adj_R^2_DFM/Adj_R^2_PCA	sd = 0.0005	sd = 0.001	sd = 0.002	sd = 0.005	sd = 0.007	sd = 0.01
1 PC/factor	1.43679414913546	0.549917805147013	1.05756668280855	2.52989579452378	3.56633840960767	1.04744152173665
2 PCs/factors	1.07203751643929	1.02135316672073	1.12055759051234	0.994729750928639	1.255490964231	1.24322107466953
3 PCs/factors	1.0606258886803	1.00657153728147	1.11580131656071	1.07455143393522	0.909569905186463	0.959493795770823
4 PCs/factors	0.966891234893584	1.00319927346228	1.04650780231082	1.00274131188201	0.978246866407632	1.02021277497582
5 PCs/factors	0.957326959494527	0.994171518705951	0.982465917394529	0.973857869853173	0.934823171469008	1.01227352334533
6 PCs/factors	0.985308685848164	0.974304505817401	0.962583136936981	0.946865029639027	0.953492646610936	1.00781970157315
7 PCs/factors	0.968827092464384	0.958115631171158	0.955095342588692	0.955361409337716	0.952008555081319	0.98706315374351
8 PCs/factors	0.976967229969193	0.946769291024096	0.94182256066397	0.949356755974142	0.954606315818951	0.99858221417369
9 PCs/factors	1.00328896896114	0.964656673765252	0.961766091317266	0.949372579024053	0.980925111809517	1.01393224516324
10 PCs/factors	1.00934465545941	0.988757245307045	0.981157294860332	0.98388366197417	1.00185068437125	1.03335642936006
11 PCs/factors	1.00358115210833	1.00432725836764	0.984012291513216	0.98513735920043	1.00754891082019	1.02419648808387
12 PCs/factors	1.00172814728175	1.00309444274617	1.00325831431808	1.00273852015914	1.00596204444129	1.01547113190128

Figure 14: Ratio of adjusted R^2 (DFM/PCA) for different standard deviations (indirect approach)

[1] "coeff3 (-0.3,0.3)"

Adj_R^2_DFM	sd = 0.0005	sd = 0.001	sd = 0.002	sd = 0.005	sd = 0.007	sd = 0.01
1 PC/factor	0.199062677188985	0.232931708561926	0.185989901430907	0.151623152433403	0.0922797992394812	0.0469905205346279
2 PCs/factors	0.317334242167502	0.361387807104666	0.357030142711388	0.225555279078475	0.15215109293901	0.0797251130531629
3 PCs/factors	0.452679384403547	0.438059290399724	0.397020021100992	0.277955595622943	0.205680183494563	0.103808993557386
4 PCs/factors	0.529517042737679	0.521831207008121	0.44933952607464	0.306695607315654	0.236951033195596	0.126518953357365
5 PCs/factors	0.623758588017335	0.602319203419721	0.52833542123868	0.328198941179293	0.259175303883681	0.139465286607463
6 PCs/factors	0.70801579661945	0.667275543630824	0.590808882609082	0.363069525402299	0.280020622030229	0.148840508195698
7 PCs/factors	0.733365712278213	0.718539260610499	0.626063754551101	0.376656932429303	0.288043624514264	0.157393858259334
8 PCs/factors	0.807568162601499	0.755461207593845	0.658991297548195	0.414834556399372	0.293512945741929	0.171529371510832
9 PCs/factors	0.863524333869032	0.809918991514226	0.701137762802058	0.445789844398391	0.319803012094951	0.183180116085258
10 PCs/factors	0.913866316433443	0.88887274911518	0.772751405668379	0.49044350574354	0.340189690698137	0.198198120146074
11 PCs/factors	0.936846282403135	0.928988128088691	0.804217803420233	0.504647681840057	0.343626343909967	0.206652115947757
12 PCs/factors	0.97854987452977	0.950349031333331	0.85160884672761	0.526916327543704	0.359954748160118	0.209882150308047

Figure 15: Adjusted R^2 of DFM for different standard deviations (indirect approach)

```
[1] "coeff3 (-0.3,0.3)"
                       sd = 0.0005       sd = 0.001        sd = 0.002        sd = 0.005        sd = 0.007        sd = 0.01
Adj_R^2_DFM/Adj_R^2_PCA
1 PC/factor            1.01588517990208  1.05841704275972  1.00532403097556  0.961914825800577 1.00306250725956  1.07952644297072
2 PCs/factors          1.00914754727805  1.00479127008866  1.00265272222631  0.9717817423423   1.03810853930541  0.963398818972184
3 PCs/factors          1.01711018415668  1.00042035855743  1.03146909194283  1.01573742481006  1.02892505590039  1.09232078229648
4 PCs/factors          1.02055668490999  1.01075996634846  1.01649798292101  1.05172485697212  1.02825100182654  1.06675380255231
5 PCs/factors          1.01878931111874  1.01168199713698  1.01131858583803  1.0543822501526   1.00133841192683  1.02342273257955
6 PCs/factors          1.00803318413736  1.00343247979935  1.00604154357175  1.03274742894134  1.02049127360298  1.0282202098747
7 PCs/factors          1.00115270284043  0.99657657772293  0.999189750234895 1.02263893625005  1.00069178894893  0.891221956564924
8 PCs/factors          0.993579713992956 1.00059068903691  0.998326437503422 1.02614314284671  0.996467164091917 0.9595846675109
9 PCs/factors          0.989560663507826 0.997604000521887 0.993867300990482 1.00694822249411  1.01576332569476  0.937301967219349
10 PCs/factors         0.990630912050796 0.997818111501023 0.997193642028707 0.996773531943322 1.00894874370026  1.03921787579573
11 PCs/factors         0.996640489617452 0.995892371683819 0.997088203741668 0.99688725414909  1.00178328219723  1.12016781588185
12 PCs/factors         0.998315817574759 0.997715860752491 0.994801816609798 1.00581489873695  0.999035076589505 0.91336064860698
```

Figure 16: Ratio of adjusted R^2 (DFM/PCA) for different standard deviations (direct approach)

```
[1] "coeff3 (-0.3,0.3)"
                sd = 0.0005       sd = 0.001        sd = 0.002        sd = 0.005        sd = 0.007        sd = 0.01
Adj_R^2_DFM
1 PC/factor     0.339785249902712 0.405263218799816 0.380438951049382 0.201726668270616 0.16783915409043  0.0662450835381195
2 PCs/factors   0.51991875502934  0.585273448142713 0.497419767489409 0.271044589075217 0.205675984231526 0.0917460080761996
3 PCs/factors   0.651448056837467 0.728071574839347 0.566214480946563 0.330504462308934 0.233282419544889 0.101230686604469
4 PCs/factors   0.785813826150296 0.800573917090108 0.622888961438461 0.364431168456633 0.247850300526001 0.105143389576584
5 PCs/factors   0.816843527637644 0.83557165003797  0.663711361787613 0.389872104728305 0.259185138423507 0.113338400302905
6 PCs/factors   0.852708075300957 0.847805276068676 0.693494884961264 0.404025371194578 0.270904798901773 0.112272778812974
7 PCs/factors   0.871889173599502 0.867695830278401 0.716708042889877 0.417698331191737 0.272486257906116 0.117713279307659
8 PCs/factors   0.888129531983112 0.887215249096145 0.743358493801379 0.425214905418643 0.282605718133603 0.118909600237989
9 PCs/factors   0.908833340730647 0.905404531559805 0.757115053622983 0.438064952001272 0.302005920517584 0.121104848223615
10 PCs/factors  0.936302371064146 0.918443208893871 0.773468933058355 0.447193967707993 0.308992689634007 0.133333799167224
11 PCs/factors  0.954985654684493 0.932076500382179 0.787764537096852 0.459911126082484 0.306339803209658 0.129712052124535
12 PCs/factors  0.971827012794879 0.941131437165661 0.797228116084461 0.469633071830499 0.305591825393637 0.13456757171492
```

Figure 17: Adjusted R^2 of DFM for different standard deviations (direct approach)

From our simulations, we observe that the dependence in time of the predictors impacts positively the goodness of fit of DFM while the level of noise has a negative impact. However, in all the cases we evaluated, DFM doesn't seem to be a better fit than PCA.

4 Forecasting accuracy and robustness

Now, we tackle the empirical forecasting performance of dynamic factor models (relatively to PCA and TS Regression). To evaluate this performance, we focus on two aspects:

- The forecasting accuracy, i.e how precise are the forecasts (in average).
- The robustness of the forecasts, i.e how variable is the forecasting performance, how confident are we in the forecasts.

Let $\hat{Y}_t$ be the predicted value of Y_t at time t, and let

$$e_t = Y_t - \hat{Y}_t,$$

be the forecast error at time t.

As measures of the forecasting accuracy, we use:

- The RMSE (Root Mean Squared Error):

$$\sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}},$$

- MAE (Mean Absolute Error):

$$\frac{\sum_{t=1}^{n} |e_t|}{n}.$$

As measure of the forecasting robustness, we used the sample variance of the forecast errors:

$$\frac{\sum_{t=1}^{n} (e_t - \bar{e})^2}{n-1},$$

where

$$\bar{e} = \frac{\sum_{t=1}^{n} e_t}{n}.$$

4.1 Evaluation method

We evaluate the forecasting performance of DFM (and the other models) using the expanding window approach. We first divided the Canadian dataset in two parts: the 145 first observations (80% of the dataset) served as the initial training set and the remaining observations served as the initial testing set. We then proceeded like this:

- First, ee trained the model (DFM, PCA or TS Regression) on the 145 first observations and used the trained model to forecast the 146^{th} observation. We then stored the forecast errors (predicted value - observed value).
- Next, we trained the model on the 146 first observations forecast the 147^{th} observation and stored the forecast error.
- We kept expanding the training set and predicting the next observation until we reached the last observation of the dataset and stored the corresponding forecast errors.

For each model (DFM, PCA and Time Series Regression), and for each number of factors/principal components (for PCA and DFM), we did this process for both the direct and indirect approaches to forecasting. We then used the stored forecast errors to compute the RMSE, the MAE and the variance of the forecast errors.

For each forecasting accuracy metric (RMSE and MAE) and each model, we selected the forecasting approach (direct or indirect) and number of factors/principal components with the 'best performance'. By 'best performance' for a given metric, let's say the RMSE, we refer to the approach (and number of factors/principal components) having the lowest RMSE and a relatively low error variance.

After selecting the best performance, we stored the following for the corresponding approach: RMSE, MAE, error variances corresponding to the best RMSE and MAE, error terms corresponding to the best RMSE and MAE.

Those are used later for comparisons between the models.

4.2 Selection of forecasting approach for Time Series Regression

Below (Figure 18), we plot the RMSE, MAE and Error variance for each forecasting approach.

As we can see on the plots, the best forecasting accuracy is obtained using the indirect approach to forecasting (lowest RMSE and MAE). Also, this approach has the lowest error variance, so is more robust.

We therefore selected the indirect approach for having the best forecasting performance for TS Regression.

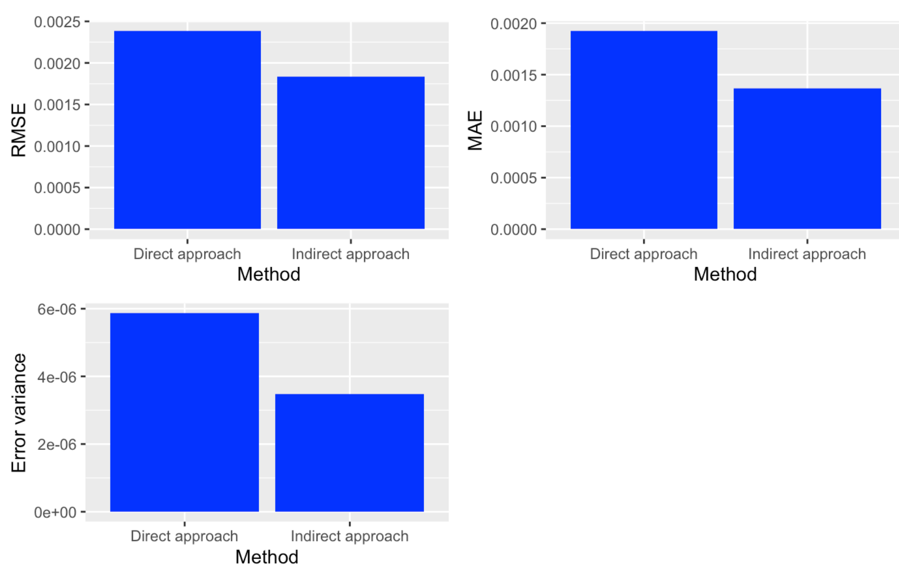

Figure 18: Forecasting performance measures for TS Regression

4.3 Selection of forecasting approach and number of factors for dynamic factor model

The results for the dynamic factor models are summarized in the following plots:

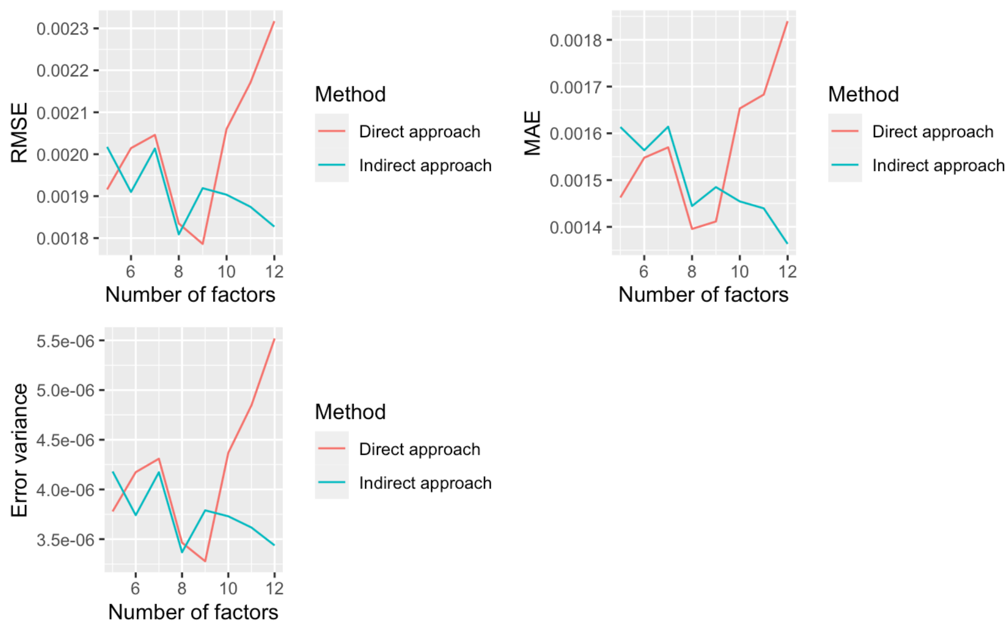

Figure 19: Forecasting performance measures for DFM

For the RMSE, the best forecasting accuracy is obtained using the direct approach and 9 factors. For the MAE, the best forecasting accuracy is obtained using the direct approach and 8 principal components. For each of these cases, the variance of the error terms is relatively low, therefore they can be selected as best forecasting performances for each metric.

4.4 Selection of forecasting approach and number of principal components for principal component analysis

The forecasting results for Principal Component analysis are summarized in the following plots:

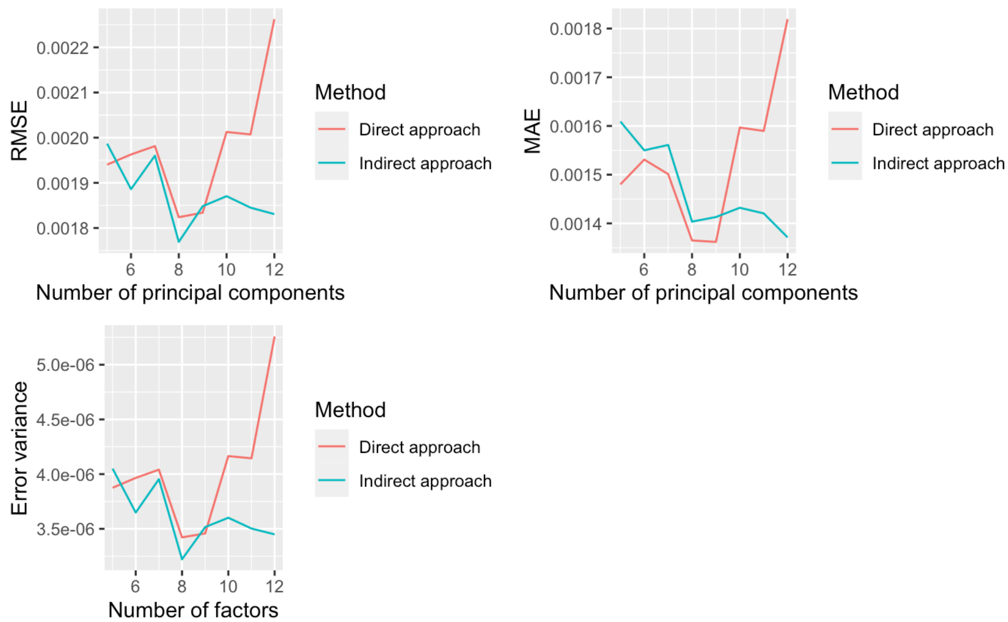

Figure 20: Forecasting performance measures for PCA

For the RMSE, the best forecasting accuracy is obtained with the indirect approach and 8 principal components. For the MAE, the best forecasting accuracy is obtained with the direct approach and 9 principal components. Both cases yield relatively low forecast error variances and are therefore relatively robust. We therefore selected them as best forecasting performances for PCA.

4.5 Comparing the 3 models

a) Visualizing the results

For each forecast accuracy metric (RMSE and MAE), we plot the best forecasting performance (selected in the previous sections) for each model (DFM, PCA, TS Regression). We also plot the forecast error variances corresponding to these best performances. We obtain the following:

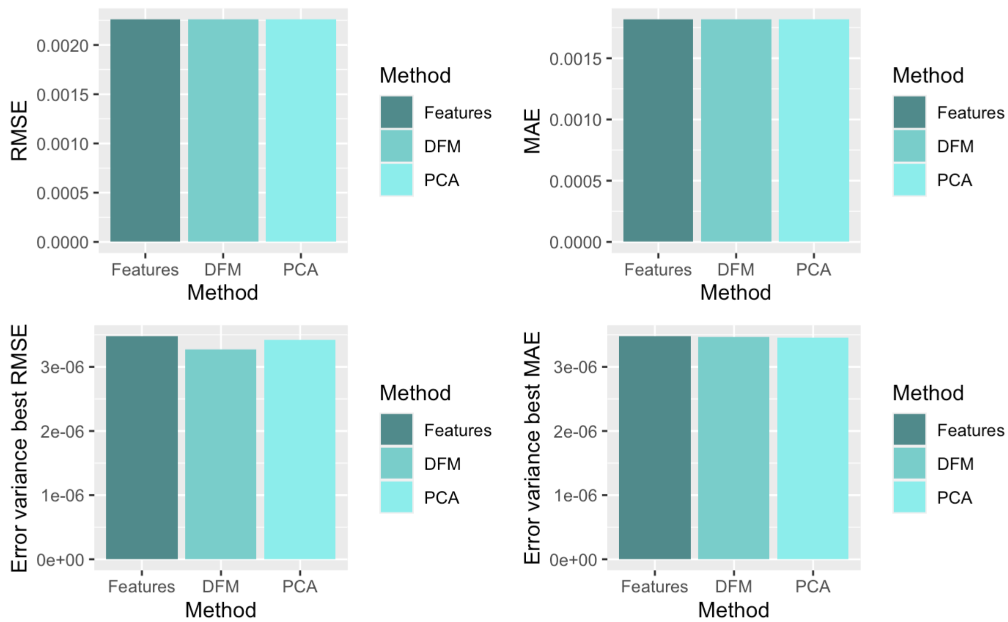

Figure 21: Best forecasting performances by accuracy metric for each model: TS Regression ("Features"), PCA and DFM

Visually, for both the forecast accuracy and the robustness, the differences between the 3 models seem very small.

In the following section, We do further tests to evaluate whether or not the 3 models have the same forecasting performance.

b) Testing equality of forecast accuracy

<u>A quick introduction to the Harvey, Leybourne and Newbold (HLN) test</u>

The Harvey, Leybourne and Newbold (HLN) test is a modification of the Diebold-Mariano test [6] to account for small samples. The test is used to assess whether the prediction accuracy of two forecasting methods is the same. The reason why we use this test is that its assumptions are only on the forecast errors and not on the models used to forecast. Therefore, it can be applied to a wide range of situations.

The test and its motivations are explained in detail in the original paper [5]. Bellow, we give a short description.

Let e_{1t} and e_{2t}, $t = 1, ..., n$ be the h-step ahead forecast errors of two competing forecasting methods. Let $d_t = g(e_{1t}) - g(e_{2t})$, where $g()$ is an arbitrary function. Suppose the following assumptions hold:

- d_t is covariance stationary,
- $\gamma(k) = 0$ for $k \geq h$, where $\gamma(k)$ is the autocovariance of d_t at lag k.

Consider the following hypothesis:

$$H_0 : E(d_t) = 0 \ \ vs \ \ H_1 : E(d_t) \neq 0.$$

Let

$$S_1^* = \left[\frac{n + 1 - 2h + n^{-1}h(h-1)}{n}\right]^{1/2} S_1,$$

where

$$S_1 = \left[n^{-1}(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k)\right]^{-1/2} \bar{d},$$

with

$$\bar{d} = n^{-1}\sum_{t=1}^{n} d_t,$$

and

$$\hat{\gamma}_k = n^{-1}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d}).$$

The HLN test rejects H_0 at level approximatively α if $|S_1^*| > t_{n-1,1-\alpha/2}$, where $t_{n-1,1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of a Student's t distribution with $n - 1$ degrees of freedom.

<u>Verifying the assumptions of the HLN test</u>

Our goal is to test the equality of:

- The squared errors (i.e $g(e_{it}) = e_{it}^2$),
- The absolute errors(i.e $g(e_{it}) = |e_{it}|$).

For each of the above functions of the forecast error, we verify the first assumption by plotting $d_t, t = 1, ..., n$ and the second assumption by plotting the autocorrelation function of d_t.

- For $g(e_{it}) = e_{it}^2$, We obtain the following:

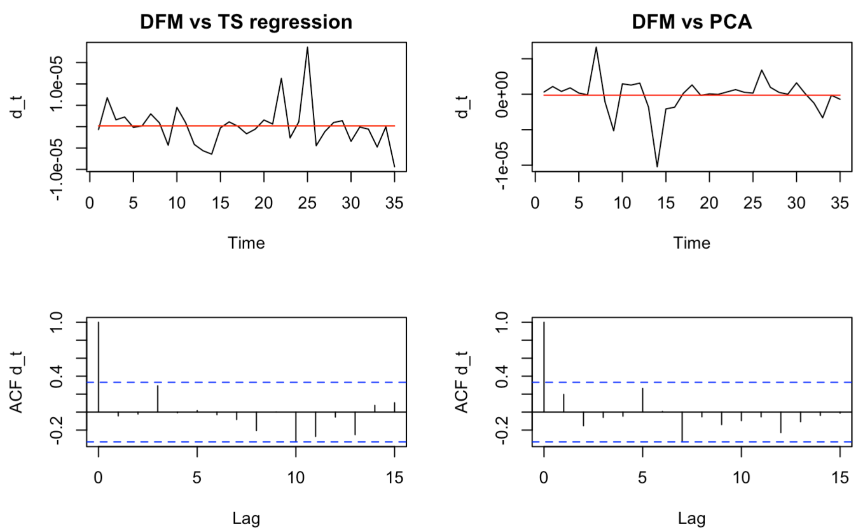

Figure 22: Difference of squared forecast errors and corresponding sample autocorrelation functions

For both comparisons (DFM vs TS Regression and DFM vs PCA), the process d_t look stationary and the autocorrelation at lag 1 is close to zero (the sample autocorrelation is within the confidence interval). It is therefore reasonable to assume that the assumptions of the HLN test are verified.

- For $g(e_{it}) = |e_{it}|$, We obtain the following:

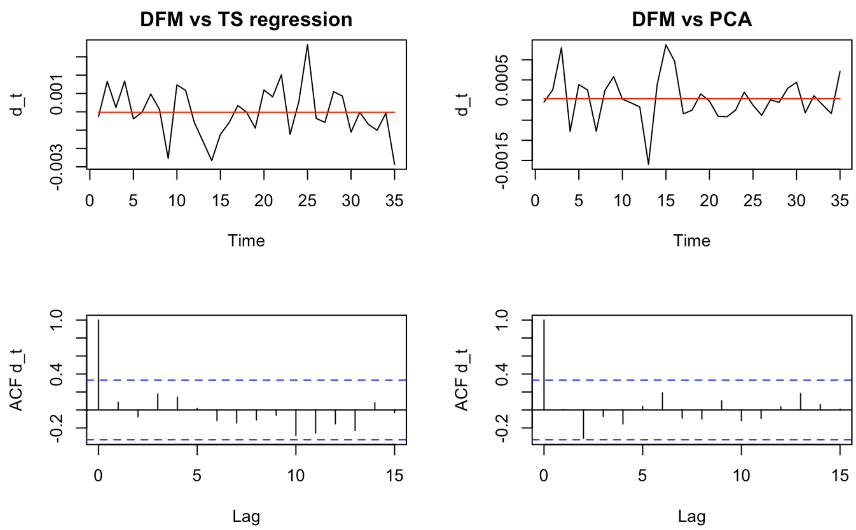

Figure 23: Difference of absolute forecast errors and corresponding sample autocorrelation functions

In this case, d_t also looks stationary for both comparions. The autocorrelation at lag 1 is small enough in both cases so that we can assume it is equal to 0.

<u>HLN test on squared errors</u>

For both comparisons, we performed the test on the forecast errors corresponding the best RMSEs (since we're dealing with squared errors). We obtained p-values of 0.8259 and 0.7562 for DFM vs TS Regression and DFM vs PCA respectively.

These values are very high, so we didn't reject the Null hypothesis in both cases.

<u>HLN test on absolute errors</u>

For both comparisons, we performed the HLN test on the forecast errors corresponding the best MAEs (since we're dealing with absolute errors). The obtained p-values were 0.9116 and 0.7278 for DFM vs TS Regression and DFM vs PCA respectively. Since these values are very high, we did not reject the Null hypothesis.

From the results of the HLN test, it is reasonable to conclude that DFM and TS have equal forecast accuracies using the provided dataset. We can also conclude that DFM and PCA have equal forecasting accuracies.

b) Testing equality of forecast error variance

We concluded in the previous section that the 3 models (PCA, DFM, TS Regression) have similar forecast accuracies as measured by both the squared errors and the absolute errors. Now we will also verify if they're equally matched in term of robustness. For that, we will test the equality of the forecast error variances.

To chose the right test, we first check if the forecast errors are normally distributed using normal Q-Q plots.

For the forecast errors corresponding to the best (lowest) RMSE, we get the following:

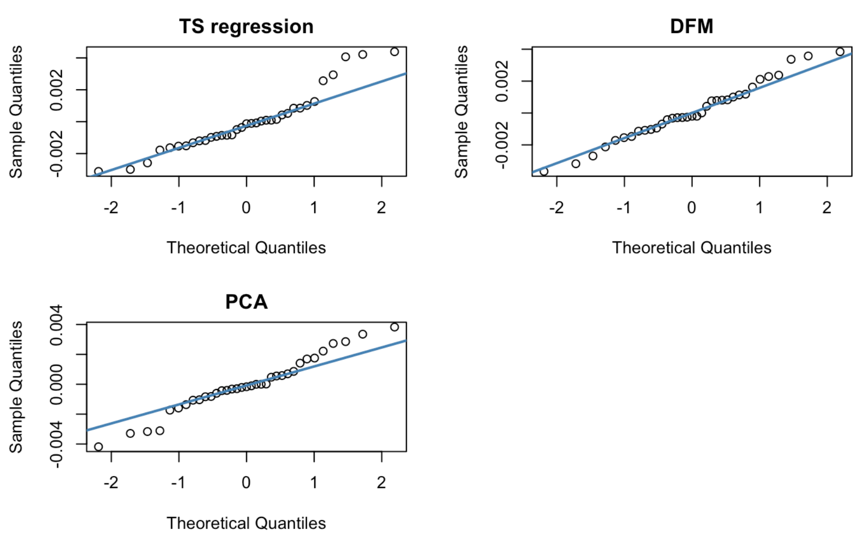

Figure 24: Q-Q plot of forecast errors corresponding to the lowest RMSE

Looking at the plots, the normality assumption doesn't seem reasonable for Time Series Regression and PCA. We do similar plots for the forecast errors corresponding to the best (lowest) MAE:

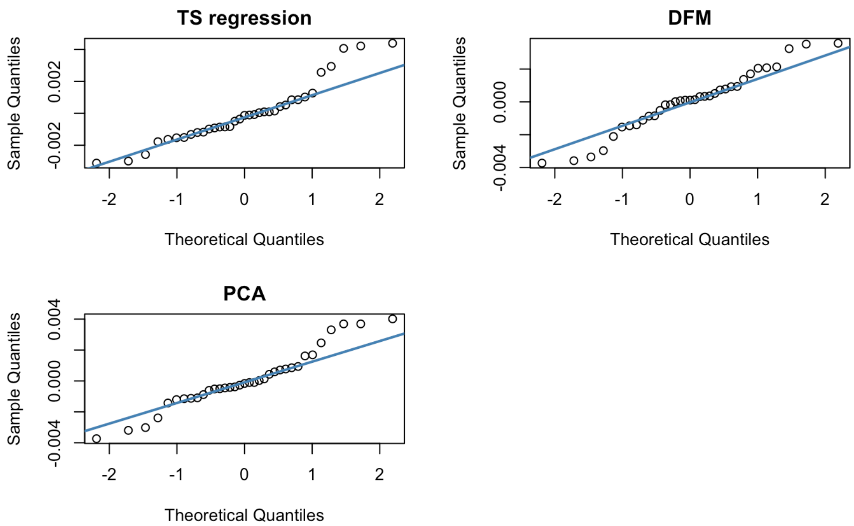

Figure 25: Q-Q plot of forecast errors corresponding to the lowest MAE

Here as well, the normality assumption doesn't seem reasonable. Therefore, we used the Brown–Forsythe test [7] for homogeneity of variances, with the median as estimate of central location. This test is more robust for non normal data.

- For the forecast errors corresponding to the best (lowest) RMSE, we ran the Brown–Forsythe test to compare forecast error variances of DFM and TS Regression, as well as the forecast error variances of DFM and PCA. The obtained p-values are 0.8941 and 0.8893 respectively. Both are high values, therefore we came to the conclusion that the forecast errors have similar variances for both comparisons

- For the forecast errors corresponding to the best (lowest) MAE, by running similar tests, the obtained p-values were 0.9296 and 0.9052 for DFM vs TS Regression and DFM vs PCA respectively. In this case, we can also conclude that the compared models have similar forecast error variances.

From the analysis we did in this section, the observation is that the dynamic factor model doesn't have an advantage on PCA and Time Series Regression when it comes to forecasting accuracy and robustness.

5 Variable selection

When using dynamic factor models for forecasting, careful variable selection is advised beforehand.

In the steps of the factor estimation, the dependent variable is not involved, so the factors are calculated without regard to how they affect the GDP. This is a disadvantage of the method.

The variables that 'contribute' the most to the factors are not necessarily those that are good predictors for the dependent variable.

In an attempt to visualize this fact, We plot for each predictor in the Canadian dataset, the R^2 of the regression of the dependent variable against the predictor. We use the R^2 as a measure of how well a given predictor explains the dependent variable.

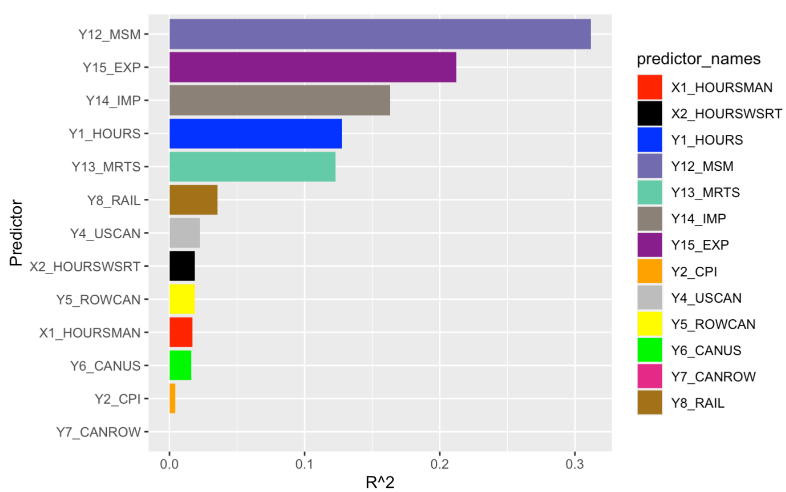

Figure 26: R^2 of the observed predictors (with respect to the dependent variable)

Next, in Figure 27, we plot for each factor the R^2 of the regression of the factor against each predictor, to measure the 'contribution' of each variable to the factor.

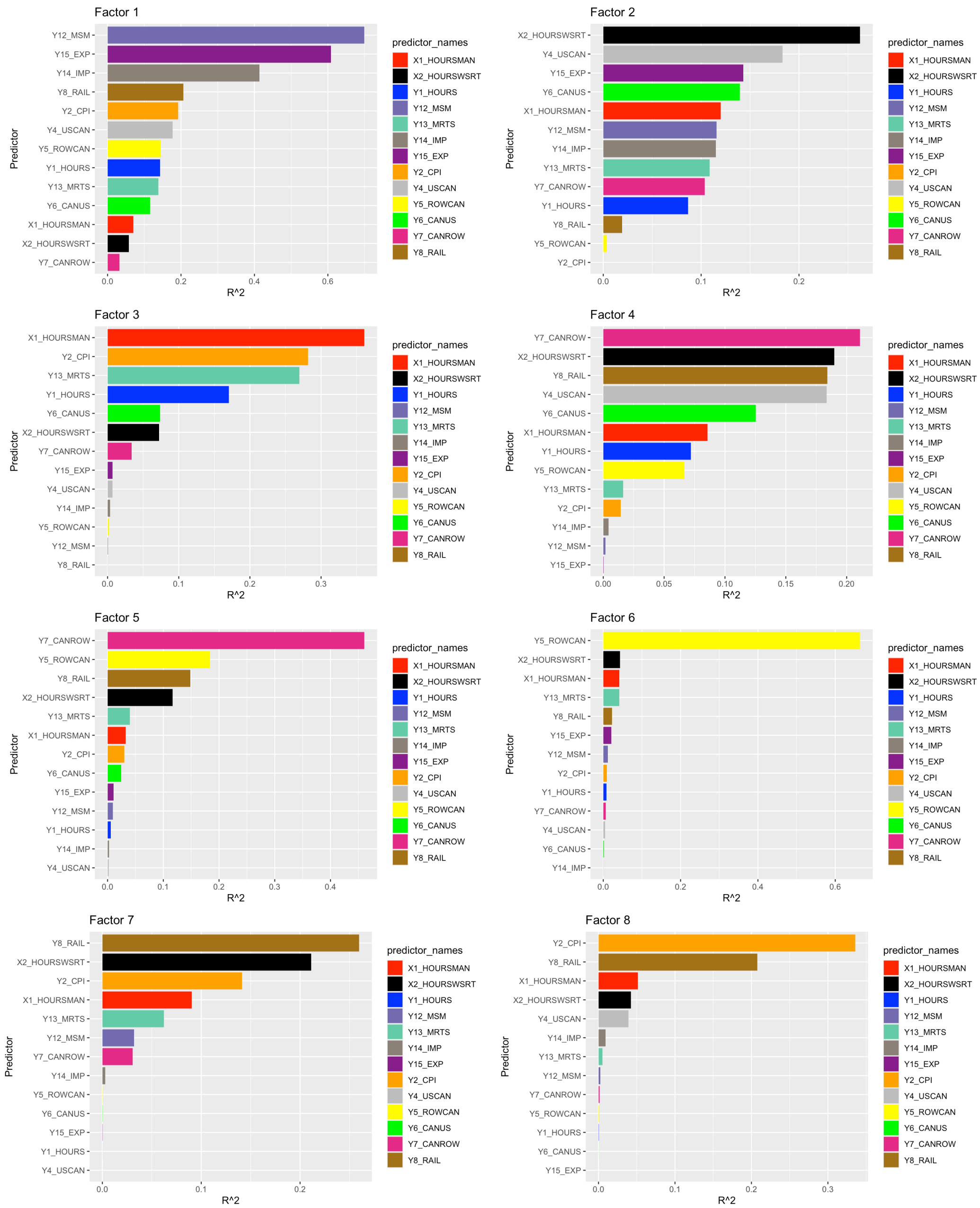

Figure 27: Variable contributions for each factor

We observe that the repartition of the contributions is very different for each factor. There doesn't seem to be variables with predominant contributions across all factors. It looks more like the factors average the information from distinct sets of variables. One could argue that the 3 predominant predictors ($Y12\_MSM$, $Y15\_EXP$ and $Y14\_IMP$) have the biggest contributions in the first factor (we didn't investigate why this is the case), but the variable $X2\_HOURSWRT$, which has a very low predictive power, has a much more important contribution to the remaining factors than $Y12\_MSM$ for example.

If we were to add variables with no relation at all to the GDP, the model wouldn't discard them, but would rather try to incorporate them.

As such, the factors are not suited to infer about the usefulness of a predictor and to select good predictors.

This can be an issue. If we want to select the variables most suited for the model with regard to forecasting, we'd have to try all combinations of variables and evaluate the forecasting performance empirically, which would be very time consuming.

Conclusions

From our analysis on the dataset provided by Statistics Canada, we observed that empirically, the dynamic factor model described in section 1.1 has the same forecasting performance as Time Series Regression and principal component analysis.

To conclude on whether or not to use dynamic factor models, we have to consider different aspects:

- If the goal is to obtain an improved forecasting performance, DFM should be discarded for this dataset, as TS Regression and PCA perform similarly.
- If the goal is to have a relatively good forecasting performance with less variables, DFM works well. However, PCA, which is simpler than DFM and computationally more efficient, performs almost as well with a similar number of factors/principal components.
- We also noticed that DFM is not appropriate for variable selection (and PCA as well). With regard to this issue, other methods like the LASSO could be preferable.

For all these reasons, the considered dynamic factor model may not be particularly useful for the provided dataset. However we shouldn't completely discard dynamic factor models, as our analysis is not exhaustive.

We only considered a simple dynamic factor model with very restrictive normality assumptions. If we relax these assumptions and the estimation method accordingly, there's a possibility to see DFM perform better.

We noticed from our simulations that DFM could be a better fit for the data when the predictors are highly dependent in time. This shows that even if we can't select variables based on the factors, some characteristics of the original features can help guide the selection. We only considered a specific set of predictors, we could obtain improved results with a different dataset.

References

[1] Rytis Bagdziunas (2015), dynfactoR, GitHub repository, https://github.com/rbagd/dynfactoR.

[2] Serge de Valk, Daiane de Mattos, Pedro Ferreira (2019). Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. The R Journal 11(1), 230-244.

[3] Domenico Giannone, Lucrezia Reichlin, David Small (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55(4), 665-676.

[4] James Durbin, Siem Jan Koopman (2012). Time Series Analysis by State Space Methods (2nd ed), 44-122. Oxford University Press.

[5] David Harvey, Stephen Leybourne, Paul Newbold (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting 13(2), 281-291.

[6] Francis X. Diebold, Robert S. Mariano (2002). Comparing Predictive Accuracy. Journal of business and economic statistics 20(1), 134-144.

[7] Morton B. Brown, Alan B. Forsythe (1974).Robust Tests for the Equality of Variances. Journal of the American Statistical Association 69(346), 364-367.

[8] Ruey S. Tsay (2012). An Introduction to Analysis of Financial Data with R. Wiley.

[9] Ruey S. Tsay (2013). Multivatiate Time Series Analysis with R. Wiley.

[10] Robert H. Shumway, David S. Stoffer (2017). Time Series Analysis and its Applications (4th ed). Springer.

[11] Walter Enders (2014). Applied Econometric Time Series (4th ed). Wiley.

[12] Jeffrey M. Wooldridge (2012). Introductory Econometrics: A Modern Approach (5th ed). South-Western College Pub.

[13] Trevor Hastie, Robert Tibshirani, Martin Wainwright (2015). Statistical Learning with Sparsity. Chapman and Hall/CRC.

[14] Vitor Cerqueira, Luis Torgo, Igor Mozetič (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. Machine Learning 109, 1997–2028.

[15] Ben Norwood (2020). Detecting patterns in Time Series Data with applications in Official Statistics (Order No. 28277942) [Doctoral dissertation, Lancaster University]. ProQuest Dissertations and Theses Global.

[16] Stefan Mittnik (1991). Derivation of the unconditional state-covariance matrix for exact maximum-likelihood estimation of ARMA models. Journal of economic dynamics and control 15(4), 731-740.

[17] Tony Chernis, Rodrigo Sekkel (2017). A dynamic factor model for nowcasting Canadian GDP growth. Empirical economics 53(1), 217-234.

[18] Tatjana Dahlhaus, Justin-Damien Guénette, Garima Vasishtha (2017). Nowcasting BRIC+M in real time. International Journal of Forecasting 33(4), 915-935.