Introducing Scotiabank's Nowcasting Model for the Canadian Economy

CONTACTS

Nikita Perevalov, Senior Economist
416.866.4205
Scotiabank Economics
nikita.perevalov@scotiabank.com

- **Following best practice in the short-term forecasting literature, we have developed a dynamic factor model intended for forecasting growth in Canadian GDP.**
- **The model, a variation on the Chernis and Sekkel (2017) of the Bank of Canada, uses a mix of Canadian and US variables at monthly and quarterly frequencies and is able to handle missing data and varying publication lags.**
- **The model provides a robust framework for assessing the impact of data releases on the forecast of Canadian real GDP growth in the current quarter, also referred to as nowcast. The model will be run after major data releases and the forecast published in a note to clients.**
- **Scotiabank Economics nowcasting model expected Canadian GDP to grow by an annualized 3.2% q/q SAAR in 2018Q2 just before the official Statistics Canada print of 2.9%.**
- **Relatively strong growth forecasted by the model was in line with strength in most major indicators: merchandise trade for April and June, GDP at basic prices in May and manufacturing shipments throughout the quarter.**

INTRODUCTION

Making informed policy decisions or forecasting the evolution of the economy over the medium term requires first an up-to-date assessment of the current state of the economy. While national statistical agencies provide the most comprehensive economic snapshot, this information is published with a significant lag. Hence, predicting the current state of the economy, in particular real GDP growth in the current quarter, in advance of national accounts publication, remains an important and difficult task.

The task is complicated in no small part by the vast number of indicators available for analysis in real time. While each economic indicator may contain some useful information, it may also be driven by idiosyncratic factors to some degree. Thus, the analyst needs to separate useful information from the noise.

To that end economists have been using various approaches to nowcast real GDP growth in a data-rich environment. Two common approaches are:

- The bottom-up approach, which requires pre-selecting among the available indicators, based on the analyst's judgement, those that are likely to be linked to specific components of GDP, estimating simple econometric relationships between them (so-called bridge models). The GDP components can then be aggregated up to arrive at total GDP (see for example Klein and Sojo, 1989 for a taxonomy of various approaches).

- Another approach combines multiple indicators into one dataset using statistical techniques to extract signals useful for nowcasting real GDP from the full panel of data. There are various techniques for doing that, including so-called factor models.

Often a combination of various methods is used. For example, the nowcast of the Federal Reserve Bank of Atlanta, GDPNow, uses both approaches for US GDP (see Higgins, 2014). For Canada, the Institute of Fiscal Studies and Democracy (IFSD) publishes a nowcast of Canadian real GDP growth, using an approach similar to that underlying GDPNow. In this paper we estimate a dynamic factor model, which is an example of a signal extraction approach because it is standard in the literature, is relatively simple to use, easy to maintain and it produces accurate forecasts. The model estimated here will supplement simple indicator models used in Scotiabank Economics for forecasting GDP components.

FACTOR MODELS IN ECONOMICS

The forecasting literature has long found that the dynamic behaviour of large panels of macroeconomic data can be summarized by a few unobserved drivers. These can be extracted using factor models, developed specifically with that problem in mind. In the context of nowcasting real GDP growth, dynamic factor models have been widely and successfully used in order to systematically extract relevant information from large panels of data for various economies (see Giannone et al, 2008, and Maier and Perevalov, 2010, for the US; Banbura and Modugno, 2010, for the euro area; and the Bank of Canada's (BoC) Chernis and Sekkel, 2017, for Canada). These and other previous studies found that:

- Factor models succeed in extracting useful information from multiple economic series to arrive at accurate forecasts of GDP growth.

- Factor models easily combine data released at various frequencies, for example monthly and quarterly, as well as handle missing data in general, including due to different publication lags and the timing of publication.

- Finally, factor models have been used as a rigorous framework for quantifying the impact of economic "news" on the nowcast of GDP.

Following the best practice in the forecasting literature, we have developed a dynamic factor model intended to be used for nowcasting quarterly real GDP growth in Canada in the current quarter. The model outperforms simpler alternative benchmark models we consider – in particular simple statistical models based exclusively on quarterly or monthly GDP data for Canada.

MODEL DESCRIPTION, DATA AND ESTIMATION

The Scotiabank nowcasting model presented in this note uses 26 monthly and 1 quarterly indicators, including 16 monthly economic indicators and quarterly GDP growth for Canada, 9 series covering the US economy and a global Purchasing Managers Index (PMI). The variables we use, as well as any transformations, are listed in Table 1.

Note that in Canada two measures of GDP are released by Statistics Canada (Statcan): quarterly GDP at market prices and monthly GDP at basic prices. These measures usually move very closely together, but sometimes can have persistent differences as well. Both are included in the data used in estimation, with the monthly indicator serving as one of the best predictors of GDP in the quarter.

The selection of variables is motivated by previous studies, e.g. BoC's Chernis and Sekkel, 2017, and includes most commonly analyzed indicators, such as monthly Canadian merchandise trade and manufacturing sales and others for Canada, as well as influential releases for the US, including industrial production and nonfarm employment.

As in the existing literature, we assume that the panel of indicators above summarizes information on three common factors that, to a greater or lesser extent, ultimately drive the variation of the observed series:

- One common factor is associated with the US and global economic developments (see Box 1 for an explanation of restrictions we assume).

- One factor explains the additional variation in Canadian variables that is not already explained by the US factor.

- The last factor captures the marginal information contained in Canadian inventory investment series that is not explained by either of the first two factors.

The intuition for the factor identification scheme is that for Canada – being a small open economy – developments in the US and global economy are, on the one hand, highly consequential and, on the other hand, exogenous, meaning Canada's domestic economic developments have little impact on the US and global economies. Hence, the purely Canadian factors are not assumed to drive the dynamics of the US indicators. In contrast, Canadian indicators are highly sensitive to the US-specific factor.

The factors themselves are assumed to evolve according to a vector autoregression (VAR) with one lag, which is sufficient to capture their autoregressive structure. The estimation procedure is a simple and robust implementation of the Expectations-Maximization Algorithm (see Banburra, Giannone, Reichlin, 2010), consisting of a two-step iterative process that also allows us to impose the above structure at the estimation stage (more details on the estimation are shown in Box 1).

NOWCASTING CANADIAN REAL GDP: HOW GOOD IS THE MODEL?

In this section we examine the forecasting accuracy of the model. Although the model can forecast at any horizon, we focus on predicting GDP growth in the quarter immediately following the latest data release, since economic literature finds that factor models are most useful in forecasting at short horizons.

To compare the accuracy of the model's out-of-sample forecasts of quarterly real Canadian GDP, we choose the following benchmarks:

- The consensus forecasts of quarterly real GDP, which are available from Bloomberg at a monthly frequency for the period of 2007–18. Since consensus forecasts are based on vintages of data available at the time of the forecast, we use Statcan's first release of real GDP from the real-time table to construct forecast errors for the consensus forecast.

Table 1: The impact of data releases on RMSEs in the Scotiabank Nowcasting Model

Series in order of publication	Data Transformation*	First month of the quarter	Second month of the quarter	Third month of the quarter	First month of the following quarter	Second month of the following quarter
US: All Employees: Total Nonfarm (SA, Thous)	DL1	-3.9	7.4	5.9	6.7	0.4
US: ISM Mfg: PMI Composite Index (SA, 50+ = Econ Expand)	DL1	-1.4	-1.3	0.0	0.2	0.0
Global Manufacturing PMI Using Markit Mfg for U.S. (SA, 50+=Expansion)	DL1	0.0	0.0	0.0	0.0	0.0
US: Manufacturers New Orders: Nondefense Capital Goods (SA, Mil.$)	DL1	-0.5	0.3	-0.6	0.1	0.0
US: Manufacturers Shipments: Nondefense Capital Goods (SA, Mil.$)	DL1	-1.1	-1.0	-0.5	-0.8	-0.8
Canada: Carloads Originated: Intermodal + Carload (Units)	DL12	1.5	0.0	-0.3	-2.1	-0.8
US: Light Weight Vehicle Sales [Autos+Light Trucks] (SAAR, Mil.Units)	DL1	1.1	-0.9	-0.4	-0.3	0.0
Canada: Chain Fisher BoC Commodity Price Index (NSA, Jan-72=100)	DL1	-4.9	-0.8	1.6	0.5	-0.3
Canada: Export, BOP Basis: All Merchandise (SA, Mil.Chn.2007.C$)	DL1	5.4	-7.8	-16.8	-17.1	0.8
Canada: Import, BOP Basis: All Merchandise (SA, Mil.Chn.2007.C$)	DL1	-3.8	4.2	-2.2	2.3	-3.2
Canada: Residential Building Permits, Units Created(SA, Units)	DL1	-0.2	0.5	-0.1	0.8	-0.4
Canada: Actual Hours Worked During Reference Week: All Sectors (SA, Thous.Hrs)	DL1	-6.9	-5.7	-4.3	-0.6	0.3
Canada: Dwelling Starts: All Areas (SAAR, Thous.Units)	DL1	-1.8	1.0	0.8	-1.1	-0.2
US: Retail Sales & Food Services (SA, Mil.$)	DL1	-0.8	-0.5	0.2	0.0	0.0
Canada: New Motor Vehicle Sales (NSA, Units)	DL12	-0.1	-0.7	0.0	0.1	-0.9
US: Industrial Production Index (SA, 2012=100)	DL1	-14.2	-7.5	-1.1	1.4	-0.1
US: IP: Motor Vehicle Assemblies (SAAR, Mil.Units)	DL1	-0.8	-1.5	-0.4	-0.1	-0.1
Canada: Shipments: All Manufacturing Industries [NAICS] (SA, Thous.C$)	DL1	-1.7	-2.9	-5.3	-0.6	2.7
Canada: Inventories: All Manufacturing Industries (SA, EOP, Mil.2012.C$)	DDL1	0.2	-0.4	-0.3	0.0	-0.1
US: Housing Starts (SAAR, Thous.Units)	DL1	0.0	-0.1	0.0	0.0	0.0
Canada: Wholesale Trade (SA, Thous.C$)	DL1	-2.2	0.4	0.2	-2.3	0.3
Canada: Wholesale Inventories (SA, Thous.C$)	DDL1	-0.8	-0.1	5.2	4.1	0.8
Canada: Retail Sales in Chained 2012 Prices (SA, Mil.Chn.2012.C$)	DL1	0.3	-0.1	-1.9	-0.2	0.3
Canada: Gross Domestic Product at Market Prices (SAAR, Mil.Chn.2007.C$)	DL3	0.0	1.3	0.0	0.0	
Canada: GDP: All Industries (SAAR, Mil.Chn.2007.C$)	DL1	-3.2	-9.2	-28.8	-9.8	
Canada: Business Outlook: Future Sales Growth: Balance of Opinion (%)	D3	0.0	0.0	0.2	0.0	
Canada: U.S. Dollar Exchange Rate (Avg, C$/US$)	DL1	0.4	0.0	-0.1	-0.1	
Total reduction in RMSE		-40	-25	-49	-19	-1

Sources: Scotiabank Economics, Haver Analytics

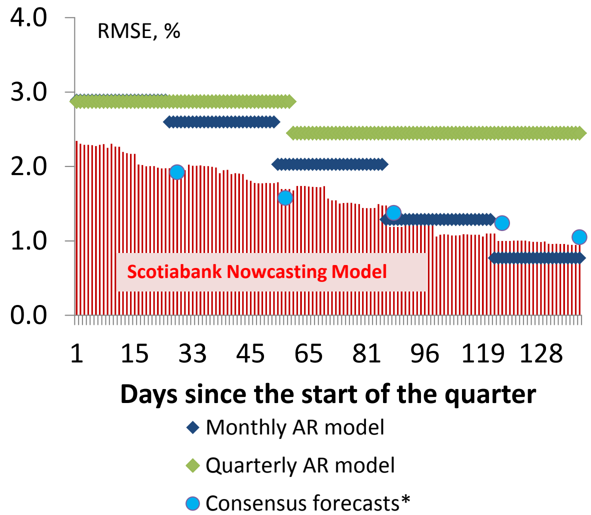
*Note: D stands for first difference, DD for first difference of first difference, L for natural logarithm, number stands for the lag in months.
E.g. DL12 is the difference in logs over 12 months.

- Forecasts from two simple statistical models: i) a model based on lags of quarterly real GDP (quarterly AR), and ii) a model based on monthly GDP at basic prices producing monthly GDP forecasts that are aggregated to quarterly frequency (monthly AR). Forecast errors for these models are computed based on the most recent release of Statcan real GDP data. The nowcasting model and the simpler statistical models are estimated over 1980–06 in order to make the comparison with consensus forecasts meaningful.

Chart 1 shows the evolution of root mean squared errors (RMSEs) from the nowcasting model and the three alternatives. The errors are defined as the difference between the realised Q/Q annualized growth in GDP and the forecast, over 2007Q1–18Q1.

- On most days in the quarter (the horizontal axis in the chart shows the number of days which elapsed since the start of the quarter), there is at least one data release from Table 1, prompting an update of the nowcast and hence a change in the RMSE from the nowcasting model (red bars) . The quarterly AR only uses quarterly data (green), which leaves its forecast constant until new quarterly GDP data is released. The forecast from the monthly AR is constant between monthly releases (blue), while the consensus forecast is published by Bloomberg at the end of the month (light blue circles).

Chart 1: Scotiabank Nowcasting Model Consistently Outperforms Simple Models and is Competitive with Consensus Forecasts



Sources: Scotiabank Economics, Haver analytics, Bloomberg, Statcan

**Note: Consensus forecasts use real-time Statcan Canadian real GDP data to compute forecast errors.*

Several important points emerge from Chart 1:

- As the time passes, more information is revealed about the economy in a given quarter, and the accuracy of forecasts from the nowcasting model and all the considered alternatives improves, as shown by declining RMSEs.
- The nowcasting model surpasses the accuracy of simpler alternative models, including an autoregressive model that uses monthly GDP data to forecast quarterly GDP growth. Only after the first two months of GDP at basic prices become available in a given quarter – around 120 days since its start – does the accuracy of the monthly AR model improve to slightly surpass the nowcasting model.
- The nowcasting model is competitive with consensus forecasts. In particular, the nowcasting model surpasses consensus forecasts in accuracy once the first month of GDP for the target quarter is released, 90 days since its start.
- The level of RMSEs from the nowcasting model is comparable to that of Chernis and Sekkel, 2017.
- Finally, the chart shows large declines in RMSEs for the nowcasting model and the monthly AR at the time of monthly GDP releases. The significant impact of monthly GDP releases on the accuracy of the nowcasting model is confirmed by the analysis presented in the next section.

WHICH SERIES HELP IMPROVE THE NOWCAST?

To investigate the impact of each data release on the model's forecasting performance, we document changes in RMSEs after each data release within a given month, respecting the usual order of publication and the lag. Table 1 shows the average reduction of RMSEs after each data release depending on the month of the quarter. The following key facts stand out:

- Monthly GDP at basic prices is overall the best economic indicator for quarterly GDP, judging by the reduction in RMSEs after the release, especially once GDP for the first month of the quarter is available (see the column "Third Month of the Quarter" in Table 1).

- Beyond monthly GDP, other variables that consistently improve the forecasting performance are Canadian merchandise trade, railroad traffic from Canada to the US and US industrial production.

- The remaining indicators provide important information at particular points in the quarter, but the improvement in RMSEs is not uniform.

Note that it is the average correlation between various series and Canadian GDP that is being exploited by the factor model. The correlation structure is something quite different from the relationship implied by the national accounting framework. As an illustration, while higher imports imply lower GDP (everything else held constant) , it is statistically likely that higher imports reflect higher domestic demand, including inventories, which may imply stronger GDP.

Furthermore, even though there are series that dominate the nowcast in terms of improving RMSEs, the information contained in all series is important, given that model errors are uniformly lower than those of simpler alternative models throughout the quarter, not only when a given influential indicator is released.

Chart 2: The evolution of the 2018Q2 Nowcast since June 28, 2018

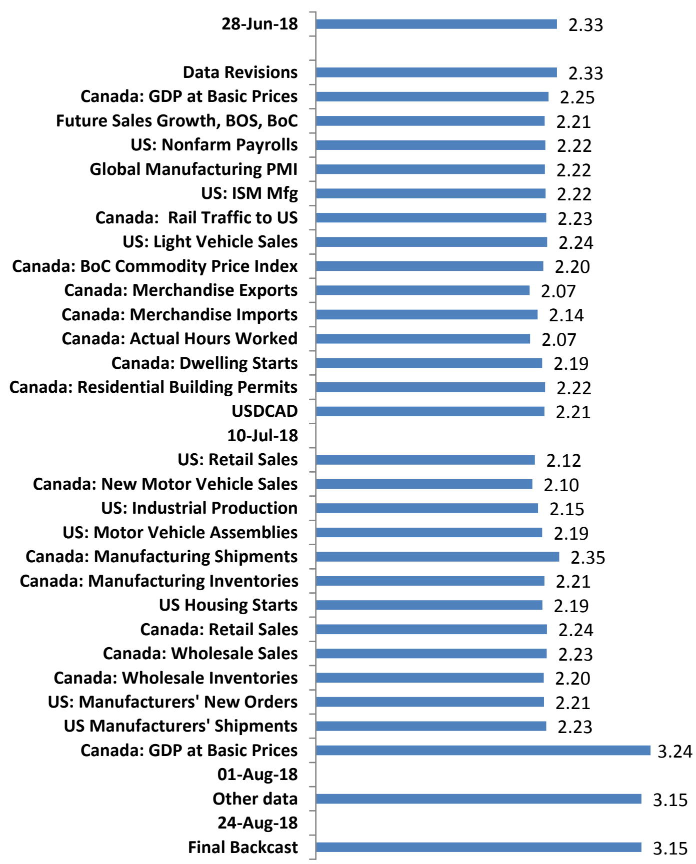


ILLUSTRATION OF THE NOWCAST FOR 2018Q2

In this section, we illustrate the use of the model in the current cycle using data releases between June 28 and the end of August, because it shows the evolution of the nowcast when important data prints are released (see Chart 2). The latest quarterly GDP data available was for 2018Q1.

- The time interval contained such important releases as monthly GDP for April on June 29th and May on July 31st, as well as international trade for May/June and hours worked for June.

- On July 10 the model expected GDP growth to reach 2.1% Q/Q SAAR on in 2018Q2, slightly lower than the nowcast of June 28th. The downgrade was explained by several data releases that surprised on the downside.

- The first negative revision came from lower-than-expected April GDP data, which expanded by just under 0.1% m/m.

- After the monthly GDP release, the largest negative revisions to July 10th came from the merchandise trade release, where exports declined in May.

- Nowcast drifted slightly higher on imports data, with a rebound in imports suggesting stronger domestic demand.

- Actual hours worked in June also came in weaker than expected, reducing the nowcast again, to 2.1.

- At the end of the week, a very strong preliminary release on the Canadian housing market, with housing starts reaching the level of 248K units, lifted the nowcast by 0.1ppt again, ending at 2.2%.

- Since July 10th, and before July 31st, the model's nowcast for 2018Q2 fluctuated around 2.2, with positive signals from manufacturing sales, retail trade and other indicators being offset by weaker US releases.
- Once monthly GDP for May was released on July 31st, the nowcast jumped by a full percentage point, reaching 3.2%. For the rest of August, other data releases came in slightly lower than expected, but the nowcast remained at 3.2% until Statcan released its estimate of 2.9% at the end of August.

CONCLUSION

The model presented in the note is of type used widely for nowcasting GDP in various countries.

- The model is shown to outperform simpler statistical alternatives and is competitive with Bloomberg consensus forecasts, incorporating various indicators that contain information about real GDP growth in a given quarter.
- The model provides a robust framework for assessing the information content in various Canadian data releases, and their impact on GDP growth in the quarter.
- The model is meant to be run regularly and is going to inform Scotiabank Economics in its short-run forecast of the Canadian GDP growth, complementing simpler forecasting models.

REFERENCES

Banbura, M., D. Giannone, and L. Reichlin (2011): "Nowcasting," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, Oxford University Press, 63–90.

Banbura, M., M. Modugno (2010): "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data", *ECB Working Paper Series*, May 2010, No 1189.

Bork, L., Dewachter, H., and R. Houssa (2009): "Identification of Macroeconomic Factors in Large Panels," *CREATES Research Papers* 2009–43, School of Economics and Management, University of Aarhus.

Chernis, T., and R. Sekkel (2017): "A dynamic factor model for nowcasting Canadian GDP growth", *Bank of Canada Staff Working Paper*, 2017–2.

Engle, R., and M. Watson (1983): "Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models," *Journal of Econometrics*, 23 (1983), 385–400.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55, 665–676.

Higgins, P. (2014): "GDPNow: a model for GDP "nowcasting"", *FRB Atlanta Working Paper* 2014–7, Federal Reserve Bank of Atlanta.

Klein, L.R. and E. Sojo (1989): "Combinations of High and Low Frequency Data in Macroeconometric Models", Klein and Marquez (eds), *Economics in Theory and Practice: An Eclectic Approach*. Advanced Studies in Theoretical and Applied Econometrics, vol 17. Springer, Dordrecht

Mariano, R. S. and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

Maier, P., and N. Perevalov (2010): "On the advantages of disaggregated data: insights from forecasting the US economy in a data-rich environment," *Bank of Canada Staff Working Paper*, 2010–10.

Box 1. Details of the Scotiabank Nowcasting Model

We assume that the data is described by the following structure (see Banbura, Giannone and Reichlin, 2010):

$$y_t = \Lambda f_t + \varepsilon_t$$
$$f_t = \mathrm{A} f_{t-1} + u_t$$
$$\varepsilon_t \sim IID\ N(0, R)$$
$$u_t \sim IID\ N(0, Q)$$

Using Mariano and Murasawa (2003), the approximation for aggregating monthly GDP growth to q/q GDP growth is as follows:

$$y_t^q \approx y_t^m + 2y_{t-1}^m + 3y_{t-2}^m + 2y_{t-3}^m + y_{t-4}^m$$

$$\begin{bmatrix} y_t^q \\ y_t^m \\ us\ and\ global\ valiables \\ Canadian\ variables \\ Inventories \end{bmatrix} = \begin{bmatrix} \lambda^q & 2\lambda^q & 3\lambda^q & 2\lambda^q & \lambda^q \\ \lambda & 0 & 0 & 0 & 0 \\ \lambda^{us} & 0 & 0 & 0 & 0 \\ \lambda^{can} & 0 & 0 & 0 & 0 \\ \lambda^{inv} & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \end{bmatrix} + \varepsilon_t$$

$$\begin{bmatrix} \lambda \\ \lambda^{us} \\ \lambda^{can} \\ \lambda^{inv} \end{bmatrix} = \begin{bmatrix} \lambda^1 & \lambda^2 & \lambda^3 \\ \lambda^{us} & 0 & 0 \\ \lambda^{can} & \lambda^{can} & 0 \\ \lambda^{inv} & \lambda^{inv} & \lambda^{inv} \end{bmatrix}$$

The parameters of the model $(\Lambda, \mathrm{A}, \mathrm{R}, \mathrm{Q})$ are estimated using the Expectations-Maximization Algorithm (see Engle, Watson, 1983). First an initial guess for the parameters is made. Having an estimate of parameters from iteration *j*, a run of the Kalman filter produces estimates of the expectation terms, $E_{\theta(j)}(f_t f_t'|\Omega_T), E_{\theta(j)}(f_t'|\Omega_T), E_{\theta(j)}(f_t f_{t-1}'|\Omega_T)$.

The expectation terms are used to compute parameter estimates at iteration *j+1*:

$$vec(\Lambda(j+1)) = \left(\sum_{t=1}^{T} E_{\theta(j)}(f_t f_t'|\Omega_T) \otimes W_t\right)^{-1} vec\left(\sum_{t=1}^{T} W_t y_t E_{\theta(j)}(f_t'|\Omega_T)\right)$$

$$A(j+1) = \left(\sum_{t=1}^{T} E_{\theta(j)}(f_t f_{t-1}'|\Omega_T)\right)\left(\sum_{t=1}^{T} E_{\theta(j)}(f_{t-1} f_{t-1}'|\Omega_T)\right)^{-1}$$

$$R(j+1) = diag\left(\frac{1}{T}\sum_{t=1}^{T}\left(W_t\left(y_t - \Lambda(j+1)E_{\theta(j)}(f_t|\Omega_T)\right)\left(y_t - \Lambda(j+1)E_{\theta(j)}(f_t|\Omega_T)\right)' W_t' + (I - W_t)R(j)(I - W_t)'\right)\right)$$

$$Q(j+1) = diag\left(\frac{1}{T}\sum_{t=1}^{T}\left(E_{\theta(j)}(f_t f_t'|\Omega_T) - A(j+1)\sum_{t=1}^{T} E_{\theta(j)}(f_{t-1} f_t'|\Omega_T)\right)\right)$$

, where W_t is a matrix that selects non-missing observations in the observable data, y_t.

Finally, at iteration *j+1* we impose restrictions embedded in the loading matrix, using the formula for constrained least squares estimation (see for example Bork, 2009 or Banbura, Modugno 2010). We consider restrictions of the following form:

$$H_\Lambda vec(\Lambda^{restr}(j+1)) = \kappa_\Lambda$$

This estimation subject to the type of restrictions above implies the following step is added at the end of the iteration:

$$vec\left(\Lambda^{restr}(j+1)\right) = vec\left(\Lambda(j+1)\right) + \\ + \left(\sum_{t=1}^{T} E_{\theta(j)}(f_t f_t' | \Omega_T) \otimes W_t\right)^{-1} \cdot H_\Lambda' \left(H_\Lambda \left(\sum_{t=1}^{T} E_{\theta(j)}(f_t f_t' | \Omega_T) \otimes W_t\right)^{-1} H_\Lambda'\right) \\ \cdot \left(\kappa_\Lambda - H_\Lambda vec\left(\Lambda(j+1)\right)\right)$$

The iterations terminate once the likelihood reaches a maximum:

$$l(y, f; \Lambda, A, R, Q) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2} f_0' \Sigma^{-1} f_0 - \frac{T}{2}\log|Q| \\ -\frac{1}{2} tr\left[Q^{-1} \sum_{t=1}^{T} (f_t - A f_{t-1})(f_t - A f_{t-1})'\right] \\ -\frac{T}{2}\log|R| - \frac{1}{2} tr\left[R^{-1} \sum_{t=1}^{T} (y_t - \Lambda f_{t-1})(y_t - \Lambda f_{t-1})'\right]$$

The model is estimated over 1980–2018 period.