Nowcasting and monitoring Israeli real economic activity*

Tim Ginker Tanya Suhoy

Bank of Israel

Abstract

We employ an extension of the collapsed dynamic factor model introduced by Bräuning and Koopman (2014) for GDP nowcasting. Compared with other popular benchmarks, our model is able to improve on the timelines and accuracy of the quarterly forecasts. Our real-time experiment during the COVID-19 crisis underlines the importance of using more timely released survey-based indicators for endpoints imputation of the "hard" data. The same framework allows the construction of a monthly index of real economic activity which is consistent with the nowcast. In contrast with the currently published by the Bank of Israel Composite State of the Economy Index, it utilizes a much broader data set and thus is likely to provide a more timely and precise picture of the course of economic activity.

Keywords: Nowcasting; Dynamic factor model; Partial Least Squares; COVID-19.

1 Introduction

In this paper, we present a nowcasting model for tracking Israeli real activity in terms of GDP growth. Using the mixed-frequency dynamic factor model, we address a number of important practical issues. Namely, the modeling framework allows us to combine monthly series with different historical lengths. In addition, it accommodates data with "jagged-edges" that arise from asynchronous data releases which lead to missing observations at the end of the sample. Along with traditional macroeconomic series with long history, such as the industrial production index, we incorporate some newly launched data sources, such as the Business Tendency Survey (hereafter - BTS) or daily volumes of credit cards purchases. In real-time applications, BTS based indices can be used as a timely proxy for important macroeconomic series that are released with significant delay. Our findings suggest that surveys can significantly reduce the nowcasting error when they are used for end points imputation.

Since the seminal research led by Stock and Watson (1989, 1991, 1993) showing that the co-movement across many macroeconomic indicators can be summarized in a few latent factors that can be used for tracking the course of economic activity, dynamic factor models have become popular in the analysis of large macroeconomic data sets and as a nowcasting tool at central banks and other institutions. Despite the existence of appropriate estimation routines (Banbura and Modugno, 2014) direct joint modeling of the quarterly GDP growth with a large panel of the available monthly indicators still leaves a model with a high number of parameters and hence higher forecast variance. To address this issue, Bräuning and Koopman (2014) introduced a collapsed dynamic factor model (hereinafter - CDFM) which allows the combination of a large number of monthly series in a relatively parsimonious model. This is a two-step procedure, where in the first step, the information contained in the monthly variables is summarized in a small number of factors using the principal component analysis (PCA). In the second step, the factors are modeled jointly with the GDP growth in a dynamic factor model.

We employ an alternative extension of the CDFM, aiming to increase the amount of relevant variation extracted in the first step. The idea is to adjust the collapsing scheme in such a way that it would put more weight on the variables that are more strongly related to economic growth. This can be achieved by finding a suitable monthly instrument that has a close relationship with the unobserved monthly growth and then using the partial least squares (PLS) scores instead of the principal components.

*We are grateful to Sercan Eraslan for his review and advice. We thank Ariel Mansura and Eyal Argov for valuable suggestions and discussions. This paper presents our preliminary findings and is being published exclusively to facilitate discussion on the topic. The authors are solely responsible for any error or omission present in the paper. The views expressed here are solely those of the authors.

From the theoretical perspective, in the CDFM framework quarterly forecasts are constructed from the filtered values of the unobserved monthly GDP growth, dynamics of which is derived mostly from the factors. Stock and Watson (2002) showed that it is possible to produce consistent estimates of the latent factors using principal components when both the number of observations and the number of monthly indicators tend to infinity. This feature was utilized by Brave et al. (2019) which constructed a new "big data" index of U.S. economic activity using a large panel of monthly indicators. However, these amounts of data are unattainable in small economies like Israel. Consequently, applying principal components to a small monthly dataset may result in low accuracy of the factors which would be passed to the implied monthly growth levels, and hence introduce a systematic error in the quarterly nowcasts. More recently, Groen and Kapetanios (2016) showed that PLS factors have similar properties to those found by Stock and Watson (2002) for PCA. Moreover, PLS retains the optimality characteristics even in the weak factor case for which it is known that PCA becomes inconsistent. Thus, if it is possible to find another monthly (instrument) variable which depends on the same factors as the unobserved monthly GDP growth, using PLS can provide a valuable alternative to PCA in small economies like Israel the number of monthly indicators is limited.

The paper is organized as follows. Section 2 presents the econometric methodology. Section 3 discusses our out-of-sample forecasting experiment and other practical issues related to the use of "soft" leading indicators and imputation methods in real-time applications, and Section 4 concludes.

2 Methodology

Let $x_t = (x_{1,t}, x_{2,t}, ..., x_{n,t})'$ with $t = 1, 2, ..., T$ be a vector of n monthly series which have been transformed to stationary and standardized. A dynamic factor model (DFM) assumes that it is possible to decompose x_t in terms of two unobserved orthogonal components representing common and idiosyncratic factors. The model is specified as follows:

$$x_t = \Lambda F_t + \varepsilon_t,\ \varepsilon_t \sim N(0, R) \tag{1}$$

where F_t is an $(r \times 1)$ vector of unobserved common factors which, Λ is an $(n \times r)$ matrix of their loadings, and ε_t is an $(n \times 1)$ vector of the idiosyncratic components. The factors are assumed to have the following stationary VAR(p) representation:

$$F_t = \sum_{s=1}^{p} \Phi_s F_{t-s} + u_t,\ u_t \sim N(0, Q) \tag{2}$$

where Φ_s are $(r \times r)$ matrices of autoregressive coefficients. The related inference and forecast procedures can be carried out using the standard Kalman filter techniques (see, for instance, Hamilton, 1994, Ch. 13).

Following Bräuning and Koopman (2014), we start from specifying the dynamics of the unobserved monthly GDP growth – y_t. Consider a benchmark model assuming that the logarithm of the GDP follows a drifting random walk giving the following dynamics of y_t:

$$y_t = \mu + \varepsilon_{y,t}, \tag{3}$$

where $\varepsilon_{y,t} \sim N(0, \sigma^2_{\varepsilon y})$, μ represents the trend component, and $\varepsilon_{y,t}$ is the error term. Its performance can be further improved by augmenting the model with a set of factors derived from x_t, giving the following representation:

$$y_t = \mu + \Lambda_{yx} F_t + \varepsilon_{y,t}.$$

In the state space form the measurement equation can be written as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \Lambda_{yx} \\ \Lambda \end{bmatrix} F_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_t \end{pmatrix} \tag{4}$$

In macroeconomic applications, the number of monthly indicators can be high relative to the number of observations, which can significantly complicate the estimation routine and introduce additional variance to the model. To overcome this difficulty, Bräuning and Koopman (2014) introduced a collapsed dynamic factor model that applies the dimension reduction transformation on the measurement equation. The idea is to use a transformed version of the measurement equation (4) pre-multiplied by the transformation matrix

$$P = \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix},$$

where A is an $(r \times n)$ matrix. The adjusted measurement equation therefore becomes

$$\begin{pmatrix} y_t \\ Ax_t \end{pmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \Lambda_{yx} \\ A\Lambda \end{bmatrix} F_t + \begin{pmatrix} \varepsilon_{y,t} \\ A\varepsilon_t \end{pmatrix},$$

while the state equation remains unchanged. To reduce the incurred information loss, Bräuning and Koopman (2014) construct A using the principal component weights. Denote by $\hat{F}_t = A_{pc}x_t$ the r principal components associated with the largest eigenvalues of the data matrix $(x_1, \ldots, x_T)'$. By writing $\hat{F}_t \approx F_t + error$, pre-multiplying (4) by $P = \begin{bmatrix} 1 & 0 \\ 0 & A_{pc} \end{bmatrix}$ gives the collapsed dynamic factor model

$$\begin{pmatrix} y_t \\ \hat{F}_t \end{pmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \Lambda_{yx} \\ I_r \end{bmatrix} F_t + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_t^{pc} \end{pmatrix}. \tag{5}$$

Since GDP growth is observed quarterly, to incorporate it with the monthly indicators within the same dynamic factor system, the state space representation needs to be further adjusted. To do so we adopt the framework proposed by Mariano and Murasawa (2003). Let GDP_t^Q denote the observed quarterly level of GDP and GDP_t^M be its unobservable monthly counterpart. We further define $Y_t^Q = \log(GDP_t^Q)$ and $Y_t^M = \log(GDP_t^M)$. Then, the unobserved monthly logarithmic growth rate y_t is equal to ΔY_t^M. To bridge between the observed quarterly data and the monthly series, we also define the GDP growth as a partially observed monthly variable

$$\begin{cases} y_t^Q = Y_t^Q - Y_{t-3}^Q & t = 3, 6, 9, ... \\ NA & o.w. \end{cases}$$

and apply the approximation of Mariano and Murasawa (2003):

$$y_t^Q = \frac{1}{3}y_t + \frac{2}{3}y_{t-1} + y_{t-2} + \frac{2}{3}y_{t-3} + \frac{1}{3}y_{t-4} \quad t = 3, 6, 9, ...$$

Finally, the series are modeled jointly using the suitable expansion of the state equation.

To summarize, the estimation of the CDFM consists of two steps. First, the information contained in the monthly variables is summarized in a small number of factors using the principal component analysis (PCA). If some of the observations are missing, imputation is performed using the iterative PCA technique proposed by Stock and Watson (2002). In the second step, the monthly factors are modeled jointly with the quarterly GDP growth in a properly adjusted dynamic factor model. In the current work, due to the aforementioned limitations of the PCA procedure in small markets like Israel where the number of monthly variables is limited, to construct the factors we opt to using the partial least squares (PLS) method with the Total revenue index as an instrument.

3 Result

Here we discuss a variety of aspects of our empirical results. First, we illustrate the advantages of the CDFM in nowcasting GDP growth. Then, we demonstrate its usefulness in monitoring the monthly changes in economic activity. Finally, we study various challenges of forecasting in real-time and imputation using survey data.

Our monthly data set consists of 40 economic and financial indicators. An overview of the series, their sources and transformations can be found in Table 1 in the online supplement. All monthly indicators are adjusted for price changes, seasonality, and are transformed to be stationary. All foreign series are presented in fixed USD prices.

Following the discussion in Sections 1 and 2, we estimate CDFM (hereafter - $CDFM^{PLS}$) using the PLS scores with the Total Revenue index being used as an instrument. Its predictive performance is then compared with the CDFM using PC scores (hereafter - $CDFM^{PC}$), dynamic factor model (DFM) and the current Bank of Israel staff forecast (hereafter - Bridge) built as a combination of an econometric model (for more technical details see see Krief, 2011) and a judgmental forecast. To make the nowcasts in the CDFM and DFM, for each quarter in the test sample we re-estimate the model using the monthly and quarterly data from January 2000 up to one of our three variants of the information set. For compatibility with the last Bridge validation report which separates the judgments effect, our forecast evaluation period is taken from the first quarter of 2010 to the second quarter of 2019. For the CDFM, the imputation of missing data at the beginning of the sample is done using the methodology of Stock and Watson (2002).

As was noted in Section 2, collapsing significantly reduces the number of parameters and hence the forecast variance. However, at the same time, it incurs the loss of information (Bräuning and Koopman, 2014). Consequently, it would be interesting to compare the predictive performance of the CDFM aginst the ordinary DFM to see whether the benefits from dimensionality reduction overweight the information loss. Our DFM specification is based on the same set of variables as CDFM with the block structure proposed by Bok et al. (2017), and is fitted using the Expectation Maximization algorithm as described in Bańbura and Modugno (2014). The models' out of sample prediction root mean squared error (RMSE) and the mean absolute error (MAE) are reported in Table 1. In all three setups, $CDFM^{PLS}$ gives the lowest prediction error. It should be noted that the initial nowcast is made one month before the Bridge. Thus, it improves not only on the accuracy but also on the timeliness of the forecast. It is worth nothing that judgements reduce the Bridge RMSE by 0.5. Thus, without a judgmental part, depending on the available data, $CDFM^{PLS}$ gives a reduction in the RMSE of 30% to 38%.

Table 1: Nowcasting error summary

	CDFM^{PLS}		CDFM^{PC}		DFM		Bridge	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
by 1 month	1.84	1.38	1.92	1.53	1.92	1.47		
by 2 months	1.75	1.34	1.82	1.43	1.90	1.45	1.89	1.45
by 3 months	1.73	1.34	1.78	1.40	1.98	1.43		

Following Brave et al. (2019) we construct a new activity index from the filtered values of the state corresponding to the current unobserved monthly GDP growth. As can be seen from Figures 1 and 2, the new indicator (hereafter - TETA) broadly coheres with the Israeli Composite State of the Economy Index (hereafter - ICI) during various stages of the cycle. However, it has a much higher variability which partly follows from the use of a broader information set and a binding restriction on the monthly growth rates to sum up exactly to the predicted quarterly growth (or to the actual, when available).

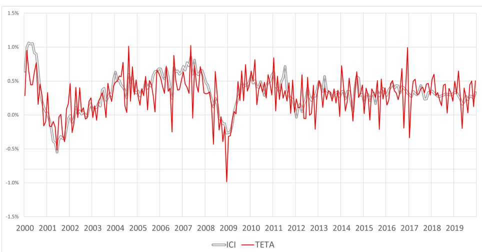

Figure 1: Monthly Economic Activity Indices: TETA alongside ICI (2010/1-2019/6).

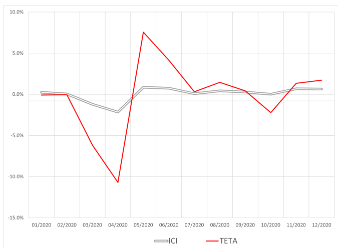

Figure 2: Monthly Economic Activity Indices: TETA alongside ICI (2020/1-2020/12).

The advantages of this higher sensitivity became apparent during 2020. As can be seen from Figure 2, during the whole period ICI continues to hover around zero and shows almost no indication of the crisis. In contrast, we observe a sharp drop in the new index coinciding with the emergency measures enacted in March which plunges even deeper in April. The marked improvement in May also corresponds to the easing of the lockdown restrictions followed by the decline in July during the second lockdown. Finally, the negative values in October coincide with the third wave of restrictions. This is in stark contrast with the ICI which remained at almost the same level since May. This demonstrates the ability of our new index to draw a more precise picture regarding the current state of the economy and to indicate earlier turning points.

Producing accurate nowcasts is most ambitious yet also most important at the periods of rapid changes and distress as was during the COVID-19 crisis in 2020. We conduct a real-time nowcasting experiment for the period from January 2020 to January 2021. For this purpose, we reconstructed the true data vintages at the weekly frequency and produced a sequence of forecasts for each quarter based on the expanding set of information.

Before, we dealt with the patterns of missing data arising solely from the differences in the historical length of the series. In practical applications it is highly common that due to non-synchronous data releases and publication delays there is an irregular pattern of missing observations at the end of the sample as well (it is frequently named as a "jagged edge" problem). As before, this issue could be treated routinely with the EM algorithm (Stock and Watson, 2002). In this method, the imputation of missing observations is done through the recursive procedure of computing the scores and filling the missing values using projection. However, many of the important leading indicators, such as production and revenue indices, are available with a significant delay (see Table 2 in the online supplement for more details on the release timeline) and the imputation accuracy based on the limited number of more timely released series may be insufficient.

Another interesting approach of dealing with missing data, which is currently applied in the Federeal Reserve Bank of Atlanta (see Higgins, 2014), was proposed by Giannone et al. (2008). In this method the initial factors are computed using the Kalman smoother. Here, the factor extraction process is designed in a way that it will simply put no weight on the missing observations rather than trying fill them using the available data. The method was less suitable for the pseudo real time forecasting experiment in the previous section because the constructed balanced monthly panel would be too short.

Many important series, such as industrial production, revenue, and labor indices have more timely released "soft" proxies. For instance, since 2011 Israeli Central Bureau of Statistics conducts a monthly compulsory survey (named a Business Tendency Survey, hereafter -BTS) among firm managers in the Manufacturing, Construction, Trade, Hotels, and Services industries, in which they provide an assessment (on a five-point scale, from "marked decrease" to "marked increase") of the current situation of their business and the outlook for the near future. The questionnaires are related to the companies' main parameters such as output, domestic and export sales, employment, and prices. Balances of opinion from such surveys is proven useful in forecasting GDP growth (see for instance Pichette and Rennison, 2011; Chernis and Sekkel, 2017). In the CDFM framework, these variables can be added directly to the monthly panel or to be used for external imputation of the series they are aimed to proxy outside the model (hereafter - two stage extraction). To test which extraction method is better, we evaluate the average precision of the nowcasts for the two alternatives with the model without the survey proxies serving as a benchmark.

Table 2 provides a breakdown of the average nowcasting error by survey inclusion method for each quarter. Several observations are in order. Prior research has emphasized the usefulness of "soft" indicators for nowcasting, and of the BTS in particular. Based on our analysis, for both factor extraction methods, BTS-based indices can significantly improve the timeliness and accuracy of the nowcasts when these are used for the end points imputation. Compared to the benchmark model without surveys, it reduces the RMSE by approximately 21.9% and 39.8% for the EM and GRS methods respectively. In contrast, while the CDFM framework is designed to deal with any number of indicators, it appears that we still should be careful with introducing unnecessary noise to the system. Expanding the monthly panel with the survey proxies of some of the already included variables leads to inferior forecasting performance. In our experiment, compared with the benchmark, the RMSE of the models with a survey block is higher by 48.5% and 16.7% for the EM and GRS methods respectively. For the two stage method both factor extraction methods shown similar MAE, and GRS had an RMSE lower by 6%.

Table 2: 2020 nowcast error summary by survey data inclusion method

	Q1	Q2	Q3	Q4	RMSE	MAE
EM without a survey	-1.94	-2.85	3.19	1.50	2.37	2.47
EM two-stage	-2.01	-0.95	2.75	1.53	1.81	1.93
EM with a survey block	-2.16	-3.89	5.70	1.12	3.22	3.66
GRS without a survey	-2.05	-4.85	1.97	2.13	2.75	3.01
GRS two-stage	-1.95	1.36	1.95	1.91	1.79	1.81
GRS with a survey block	-2.20	-5.78	2.55	2.14	3.17	3.51

4 Discussion and Conclusion

In this paper, we applied an extension of the CDFM to produce the nowcasts for Israeli quarterly GDP growth. Our pseudo-real-time forecasting experiment shows that the model is able to produce nowcasts as accurately as the judgmental ones. In addition, we have shown that using instruments can reduce the information loss from the dimension reduction and thus improve the accuracy of the forecasts. Our real-time nowcasting experiment during the COVID-19 crisis underlines the importance of using more timely released survey-based indicators for end points imputation of the "hard" data. Moreover, similarly to Brave et al. (2019) we applied the same framework to produce a new monthly index of economic activity. The index is shown to be able to give a more precise picture regarding the current state of the economy and to indicate earlier turning points.

References

Banbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bańbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti (2017). Macroeconomic nowcasting and forecasting with big data. Staff Report 830, New York, NY.

Brave, S. A., R. A. Butters, , and D. Kelley (2019). A new" big data" index of US economic activity. *Economic Perspectives, Federal Reserve Bank of Chicago 43*, 1–30.

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting 30*, 572–584.

Chernis, T. and R. Sekkel (2017, August). A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics 53*(1), 217–234.

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

Groen, J. J. and G. Kapetanios (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics Data Analysis 100*, 221–239.

Hamilton, J. (1994). *Time series analysis.* Princeton university press.

Higgins, P. (2014). Gdpnow: A model for gdp n̈owcasting. FRB Atlanta Working Paper 2014-7, Federal Reserve Bank of Atlanta.

Krief, T. (2011). Nowcasting model for GDP and its components. Technical Report 2011.01, Bank of Israel.

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.

Pichette, L. and L. Rennison (2011). Extracting information from the business outlook survey: A principal-component approach. *Bank of Canada Review 2011*(Autumn), 21–28.

Stock, J. and M. Watson (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 351–394.

Stock, J. and M. Watson (1991). A probability model of the coincident economic indicators. *Leading Economic indicators: new approaches and forecasting records 66.*

Stock, J. and M. Watson (1993). *Business cycles, indicators and forecasting*, Chapter A procedure for predicting recessions with leading indicators: econometric issues and recent experience. University of Chicago Press.

Stock, J. and M. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.