High Frequency Indicators and Tracking Real Activity in North Macedonia¹
May 2021

High frequency economic indicators, such as industrial production or retail sales, are weakly correlated with the real gross domestic product (GDP) in North Macedonia. This makes tracking real activity in real time a challenging task. However, quarterly GDP growth may be decomposed into two important components: a fundamental economic signal and a random noise (a measurement error). A methodology discussed in this paper proposes to extract the economic signal that is highly correlated with a number of available high frequency indicators. The presented methodology may be used both as a nowcasting tool and also for construction of a monthly coincident economic indicator for North Macedonia.

¹ The methodology has been developed by Davor Kunovac (Consultant) and Sanja Madzarevic-Sujster (Senior Economist), and has been presented to the National Bank of the Republic of North Macedonia and the Ministry of Finance staff in early 2021. We are grateful for their comments, as well as those of Enrique Blanco, Sandra Hlivnjak, and James Sampi Bravo.
Contents

1 Introduction and Motivation
   1.1 Motivation and Objective
   1.2 What is Nowcasting?
   1.3 Methodology for Nowcasting and a Coincident Economic Indicator

2 Factor Model for Nowcasting and Constructing CEI
   2.1 Estimating the Factors: Basic Factor Model and PCA
   2.2 Nowcasting and Tracking Economic Activity with OLS Regression

3 Usefulness of High Frequency Indicators for Monitoring Real Activity in North Macedonia
   3.1 Evidence Using Real Industrial Production and Retail Sale
   3.2 Looking Beyond Industrial Production and Retail Sales

4 Results
   4.1 Separating Real Activity From Measurement Error
   4.2 PCA Analysis
   4.3 Nowcasting model and CEI for North Macedonia

5 Tracking Real Activity with Alternative Sources of Information

6 Conclusion

References

Statistical Annex
1 Introduction and Motivation

1.1 Motivation and Objective

The objective of this work was to evaluate the ability of a large number of high frequency indicators to track the real activity on a monthly basis in North Macedonia. For that purpose, time series of quarterly GDP in North Macedonia was tested against a large number of high frequency indicators. The conclusion is that the officially-released real GDP in North Macedonia poorly correlates with available high frequency economic indicators. This link is much weaker than one usually observed in European economies. Consequently, real activity—measured using official quarterly GDP—cannot be assessed with sufficient precision relying on dynamics of standard economic indicators such as industrial production, internal or external trade, other real economy indicators, or survey data. Official quarterly GDP, therefore, seems to contain excessive noisy component that dominates important economic content (signal) of the series. This noisy component is labeled herewith as the measurement error.

To be able to monitor the real activity developments in North Macedonia in real time, the methodology of filtering out the economic signal from the quarterly GDP growth is proposed. In that way, the large measurement error when assessing the level of real activity using official national accounts and high frequency indicators is eliminated. The main assumption used to decompose GDP growth into (i) an economic signal and (ii) a measurement error is that the economic signal must be consistent with a large number of available high frequency indicators. Noisy measurement error on the other hand is, by construction, unrelated to available economic indicators. To extract that particular component of GDP that is correlated to a large number of available high frequency indicators, we rely on standard factor models. These models are usually employed in nowcasting exercises. The main features of these models are elaborated in details below.

1.2 What is Nowcasting?

The official quarterly GDP data are released with a substantial lag. For instance, first estimate is usually available only two months after the end of a quarter, sometimes even later. In the
absence of official quarterly GDP data, real activity can still be monitored in real time using number of available high frequency indicators. This process is called nowcasting. Nowcasting is sometimes defined more formally as the prediction of GDP in the very recent past, the present, and the very near future (Banbura et al., 2013). Any nowcasting model relates (or bridge) official quarterly GDP to a large number of available economic indicators, usually available on a monthly basis (industrial production or retail sales, for example) with a shorter release time. Simple bridge models are usually used for this purpose (see for example Baffigi et al., 2004). The strategy of this approach is simple: monthly indicators are first aggregated to a quarterly level, and then the (quarterly) relationship between GDP and available indicators is estimated directly, often in the form of a simple Ordinary Least Squares (OLS) equation. That relation is then used for nowcasting purposes. Bridge models, although very intuitive and easy to implement, can simultaneously relate GDP to only a handful of monthly indicators, probably to not more than five or six because macroeconomic series are typically short. For that reason, more recent literature largely relies on factor models to relate potentially very large number of high frequency indicators to quarterly GDP. Being able to include a large number of predictors into regression framework may be of special interest in case when there are no many obvious highly correlated monthly predictors of GDP as in North Macedonia. In that case, useful information on high frequency dynamics in real activity is likely to be extracted from a large number of modestly correlated predictors and, therefore, factor models seem to be better suited for nowcasting purposes.

1.3 Methodology for Nowcasting and a Coincident Economic Indicator

A standard nowcasting factor model can be used to analyse real activity in North Macedonia for both past and current quarters. Such a model can be relevant in case when it is difficult to monitor real activity in real time given highly volatile GDP growth and its disconnect from available high frequency indicators. While standard nowcasting models rely on historical relation between GDP and monthly indicators to nowcast GDP growth in the current quarter, this relation may be also useful to construct a reliable Coincident Economic Indicator (CEI).

---

2 The official GDP is published on a quarterly frequency. In that context, here we label an indicator as a high frequency if it is being published more frequently compared to GDP - most often on a monthly basis.

3 See for example Giannone et al. (2008), Schumacher and Breitung (2008) or Kunovac and Špalat (2014) for some examples.
Indeed, given a large number of available high frequency economic indicators, using a standard nowcasting factor model, we were able to extract a portion of GDP dynamics that is consistent with those high frequency indicators. By definition, such a model derives an economic indicator that is correlated to high frequency indicators and may, therefore, be interpreted conveniently as a monthly \textit{coincident economic indicator} for North Macedonia. The specified factor model may directly be used for nowcasting the real activity. However, nowcasted real activity in such case does not represent the quarterly GDP growth, but rather real economic activity (economic signal).

2 \hspace{1em} \textbf{Factor Model for Nowcasting and Constructing CEI}

Nowcasting models usually rely on simple OLS bridge regressions or on more complicated \textit{factor models}. Implementation of bridge regressions is simple and only includes specifying a regression that correlates GDP growth with available high frequency indicators. For that reason, here we discuss some main features of factor models and their use for nowcasting real activity.

Given a large number of high frequency indicators used, the methodology first summarizes their overall variation in form of a handful of variables - factors. Having in mind that factors are available at higher frequency, usually monthly, one can exploit their dynamics and assess the real activity in real time using the information from a large number of available high frequency indicators. This strategy enables to simultaneously relate GDP and a large number, potentially several hundreds or thousands, of variables using a simple OLS regression. The key feature of this model is that factors, being \textit{unobservable} variables, are to be inferred from available \textit{observed} high frequency indicators and the structure of the model. To estimate the common factors consistently, here we rely on the Principal Components Analysis (PCA)\textsuperscript{4}.

The nowcasting model developed for North Macedonia is fully specified by the following two-step procedure:

1. Estimation of the common factors using a large number of available high frequency indicators; and

2. Nowcasting GDP using the factors within standard regression framework.

Some very basic properties of factor models are discussed below.

2.1 Estimating the Factors: Basic Factor Model and PCA

2.1.1 Factor Model

In this simple version of a factor model, we assume that each of the $N$ monthly indicators:

$$X_i = (x_{i1}, ..., x_{iT}), \text{ for } i = 1, ..., N,$$

can be represented as a linear combination of a small number ($r << N$, $r$ much smaller than $N$) of factors $F_1, ..., Fr$, given in the $T \times r$ matrix $F$:

$$X_i = F \Lambda_i + e_i$$  \hspace{1cm} (1)

where $\Lambda_i$ is the $r$-dimensional vector of factor loadings ("weights"), and $e_i$ are idiosyncratic errors. In other words, there exists a small group of $r$ common factors whose different combinations (determined with weights $\Lambda_i$) represent well the variation in all the available indicators $X_1, ..., X_N$. Factors $F$ (and the associated weights $\Lambda_i$) can be estimated consistently under certain (mild) conditions by employing the PCA (asymptotic properties are analysed in Bai, 2004 and in Stock and Watson, 2002). Alternatively, Kalman filtering may be employed to estimate the factors within a state space framework. Such a strategy is known to produce similar estimates of unobserved factors as the PCA, but has an important shortcoming in that it inevitably increases technical complexity of the model. There is, however, an important advantage of estimating factor models via state space techniques—they can handle missing observations’ problem easily. However, given this is not relevant for North Macedonia, a simpler strategy based on principal components is applied.

2.1.2 Estimating Factors via Principal Components Analysis

Here is some basic intuition behind the PCA. Let $\Sigma$ be the covariance matrix of high frequency indicators $X_1, ..., X_N$. Let $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$ be corresponding eigenvalues. The principle components are defined recursively as uncorrelated linear combinations of indicators $X_1, ..., X_N$, with important property: they must be of maximal variance. In that case, these linear combinations may appropriately capture underlying common drivers of a large group of variables and, therefore, may serve as a representative of overall real activity
in real time. All the variables are standardised to have zero mean and unit standard deviation before the PCA is applied.

There are two important remarks related to the construction and interpretation of principal components.

- **Remark 1 (Construction of principal components)** Let $\Sigma$ be the covariance matrix of vector $X = (X_1, \ldots, X_N)$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$ be eigenvalues and, a $x_1, \ldots, x_N$ the corresponding (unit norm) eigenvectors of $\Sigma$. Then the $i$-th principal component is given by:

$$Y_i = x_i^T X, i = 1, \ldots, N$$

Furthermore:

$$\text{Var} Y_i = \lambda_i \text{ and } \text{Cov}(Y_i, Y_k) = 0, \text{for } i \neq k.$$ 

- **Remark 2** Let $Y_1, \ldots, Y_N$ be the principal components from the remark 1. Then, the following holds:

$$\sum \text{Var} Y_i = \lambda_1 + \lambda_2 + \ldots + \lambda_N = \sum \text{Var} X_i.$$ 

The first remark says that the principal components are calculated as linear combinations of high frequency indicators from vector $X$, eigenvectors of the corresponding covariance matrix being the weights. The second remark ensures equality of the sum of the variances of the elements of the original vector $X$ and of constructed principal components. The principal components exhaust the overall variance of high frequency indicators.

A practical importance of these results is that in order to construct principal components it is sufficient to calculate eigenvectors and eigenvalues of covariance matrix of high frequency indicators. Then, to be able to assess the importance of these principal components in terms of their ability to capture the overall variance of available monthly indicators ($\sum \text{Var} X_i$), it is sufficient to assess the relative importance of pertaining eigenvalues in their total sum $\sum \lambda_i$. For example, $\frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots + \lambda_N}$ shows how much of overall variance of all the series ($\sum \text{Var} X_i$) is explained by a single principal component. PCA is successful in reducing the dimensionality of potentially huge data sets if a handful of principal components may capture a large share of total variation of a data set.
A related literature documents how these principal components may be used to estimate factors $F_1, \ldots, F_r$ from (1) (see Kunovac, 2007 for example).

### 2.2 Nowcasting and Tracking Economic Activity with OLS Regression

#### 2.2.1 Nowcasting

Let high frequency indicators $X_1, \ldots, X_N$ be those available at monthly frequency. Using the PCA, we first estimate monthly factors $f_t^M$ and aggregate them to quarterly factors $f_t^Q$. The GDP data is released with a lag relative to monthly indicators, that is, factors are available for periods $1, \ldots, T^q$, and GDP for one quarter period less $(1, \ldots, T^q - 1)$. For that reason, the following *bridge equation* is used for computing the forecast/nowcast of the GDP growth rate:

$$GD t^Q_t = \alpha + \beta f^Q_t + \epsilon_t.$$  \hfill (2)

The parameters of regression are estimated by using a simple OLS method on a sample: $1, \ldots, T^q - 1$. A nowcast is now evaluated at $T^q$ as the expected GDP, conditional on the available information of high frequency indicators summarised through common factors:

$$E(GDP_{T^q}^Q|f^Q_{T^q}) = E(\alpha + \beta f^Q_{T^q} + \epsilon_{T^q}|f^Q_{T^q}) = \hat{\alpha} + \hat{\beta} f^Q_{T^q}.$$ \hfill (3)

The nowcast $\hat{\alpha} + \hat{\beta} f^Q_{T^q}$ is constructed where parameters $\hat{\alpha}$ and $\hat{\beta}$ are known OLS coefficients and $f^Q_{T^q}$ are quarterly factors estimated using the PCA.

#### 2.2.2 Coincident Economic Indicator for North Macedonia

Finally, when all available high frequency indicators are firstly transformed into year-over-year (yoy) growth rates, and $f^M_t, f^Q_t$ and $GDP^Q_{T^q}$ are all also in yoy terms then:

$$CEI^M_t := \hat{\alpha} + \hat{\beta} f^M_t.$$ \hfill (4)

may be interpreted as a monthly coincident indicator ("monthly GDP growth"). Indeed, aggregating $CEI^M_t$ at a quarterly level using simple averages approximates conditional expectation implied by (2).
3 Usefulness of High Frequency Indicators for Monitoring Real Activity in North Macedonia

3.1 Evidence Using Real Industrial Production and Retail Sale

This section aims to evaluate the ability of a large number of main high frequency indicators to track the real activity measured by GDP in North Macedonia in real time. For comparison purposes, the analysis is conducted for a large group of European (both EU and non-EU) countries.\(^5\)

In the first example, the most basic nowcasting bridge regression is specified as:

\[
GDP_t = c + \alpha GDP_{t-1} + \beta IND_t + \delta RETAIL_t + \epsilon_t, \tag{5}
\]

where GDP growth is explained by its own lagged values and two key high frequency indicators—industrial production and retail sales, both in real terms. Such a simple specification is usually able to capture most of the variation of real GDP growth. In addition, given both industrial production and retail sales are available on a monthly basis, equation (5) can be used for nowcasting. In order to assess the ability of this simple bridge equation to track the GDP on a regular basis, we estimate equation (5) for 31 European countries and compare their \(R^2\) statistics.\(^6\) By doing this, we compare the overall share of variation in GDP growth that is explained by the two indicators and lagged GDP.

---

\(^5\) The total sample includes 34 European countries with data available on Eurostat: Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany (until 1990 former territory of the FRG), Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom.

\(^6\) The sample consists of all the countries for which equation (5) can be estimated over the period 2003q1 - 2020q1. This is the sample for which all the data are also available for North Macedonia. Only three countries dropped out according to this particular criterion: Turkey, Switzerland and Bosnia and Herzegovina. However, GDP in Turkey and Bosnia and Herzegovina may be explained by the estimated equation very well, while Switzerland is always an outlier given its structure of economy which cannot be explained with this simple equation.
Figure 1. $R^2$ Statistics from a Simple Nowcasting Specification

Source: Eurostat, National statistical offices, staff calculations.

Figure 1 compares $R^2$ statistics for 31 countries in the sample and clearly illustrates that most of the variation in GDP growth is well explained by dynamics of industrial production, retail sales and lagged GDP: on average 70 percent. In fact, almost all of that explanatory power is due to variation in two economic indicators; lagged GDP plays a minor role. In North Macedonia, however, only 10 percent of variation in GDP growth is explained in this way which is the lowest of all countries in the sample covering the period of 2003q1 - 2020q1. To gain more insights into such a striking difference between how well these indicators may explain fluctuations in GDP growth in North Macedonia and some other countries, Figure 2 compares its observed GDP growth and model in-sample forecast with that of a country with very high $R^2$ statistics, Italy. In contrast to Italy, GDP in North Macedonia is much more volatile throughout the sample than predicted by movements in some standard high frequency economic indicators.

Figure 2. GDP Growth and In-Sample Model Forecast for North Macedonia and Italy

Source: National statistical offices, staff calculations.

Finally, to additionally illustrate weak in-sample forecast accuracy of the model (5) for North
Macedonia, we calculated standard deviation of estimated residual from that equation on successive 12-quarter rolling windows (Figure 3). Standard error of estimated residuals for North Macedonia (in blue) has been consistently larger compared to other countries in our sample, with few exceptions. Towards the end of the sample, standard deviation for North Macedonia, however, has been decreasing, but is still larger compared to other countries in the sample.

**Figure 3. Standard Deviation of Estimated Residual of Equation (5) (a 12-quarter rolling window)**

Another important investigation is to check the corellations of each of the two high frequency indicators with related gross value added (GVA) NACE categories: B-E: Industry (except construction) and G-I: Wholesale and retail trade, transport, accommodation and food service activities, as published by the State Statistical Office. To do that, for each country $i$, we specify two simple OLS equations:

$$IND_{GVA\_BE\_it} = c + \beta IND_{it} + \varepsilon_t,$$

$$TRADE\_GVA\_GI\_it = c + \delta RETAIL_{it} + \varepsilon_t,$$

where $IND_{GVA\_BE\_it}$ and $IND_{it}$ denotes GVA category B-E (Industry except

---

7 Ireland has been omitted from this exercise due to a large structural break in 2015 when recorded GDP growth was about 23 percent.
construction) and a short term indicator of industrial production for country \( i \) respectively. \( TRADE_{GVA,GI,lt} \) and \( RETAIL_{lt} \) denote GVA category G-I (Wholesale and retail trade, transport, accommodation and food service activities) and a short-term indicator for retail sales, respectively. Figure 4 compares \( R^2 \) statistics resulting from these equations across a large number of European countries.

**Figure 4.** \( R^2 \) Statistics for GVA Categories vs Industrial Production (left chart) and Retail Sales (right chart) Indicators

![Graph showing \( R^2 \) statistics for different countries](image)

*Source: Eurostat, national statistical offices, staff calculations.*

For both categories, short term indicators in North Macedonia are, with few exceptions, significantly less correlated to corresponding GVA categories than it is the case with other countries in the sample. In fact, this correlation is practically non-existent and hardly useful for nowcasting or forecasting purposes. To illustrate that visually, Figure 5 compares observed GVA categories (B-E and G-I) and in-sample predictions from equations (6) and (7) for North Macedonia and the two countries where industrial production (Italy) and retail sales (Latvia) capture dynamics of corresponding GVA categories well.
Finally, taking into account that time series of retail sales is an important proxy for monitoring personal consumption, we also look at correlation between the two series. For that purpose, for country $i$ we specify a simple OLS regression:

$$CONSUMPTION_{it} = c + \beta RETAIL_{it} + \epsilon_t,$$  

(8)

where $CONSUMPTION_{it}$ denotes growth of real personal consumption from national accounts and $RETAIL_{it}$ denotes retail sales series. The results in terms of $R^2$ statistics are provided in Figure 6 which illustrates again that in case of North Macedonia retail sales series is weakly correlated with personal consumption from the national accounts and that for practical purposes when nowcasting real personal consumption this is not a useful indicator in North Macedonia.

Source: National statistical offices, staff calculations.
The main conclusions of this chapter may be summarized as follows:

- Industrial production and real retail sales are of very limited help in tracking dynamics of quarterly GDP growth in real time in North Macedonia;
- The two high frequency indicators are also poorly correlated with corresponding categories in GVA in North Macedonia;
- Looking at the GDP growth from expenditure side does not alter conclusions: personal consumption from national accounts and its corresponding short-term indicator—retail sales—are in North Macedonia, in fact, uncorrelated;
- Such a disconnect between main short term indicators and national accounts is largely an exception among European countries. There are, however, few European countries with similar features in that context, most often Malta or Luxembourg, but their structure of economy is much different from that of North Macedonia. The relative importance of industry in GVA in these two countries is, in contrast to North Macedonia, close to zero and the observed disconnect may be its natural consequence (Figure 7).
- Finally, observed disconnect is unlikely a consequence of potentially different structures of North Macedonia and other European economies. Figure 7 shows that the relative importance of industry and internal trade for the overall economy in North Macedonia is in fact, on average, very close to that of other European countries.

Source: Eurostat, national statistical offices, staff calculations.
3.2 Looking Beyond Industrial Production and Retail Sales

As discussed in the previous section, monthly indicators of industrial production and retail sales cannot explain variation in quarterly GDP in North Macedonia very well; in contrast, by using variation of these two indicators only, GDP growth can be well approximated in most European countries. Performance of such an oversimplified model may arguably be improved by taking larger number of available monthly indicators into account. This can be implemented using factor models as explained earlier.

Factor model allows for assessing correlation between GDP growth and a larger number of available high frequency indicators usually used to monitor real activity. The objective is to find other indicators, rather than the two already investigated, that may better explain dynamics of quarterly GDP growth. To investigate that in more detail, a group of almost 1,000 time series from various sources is investigated. All time series together with sources are reported in Annex Table A1. Overall conclusion of this analysis is that, on average, correlation between GDP growth and available group of a large number of high frequency indicators is weak and very unstable over the sample under analysis. Most importantly, it is very difficult to find a monthly indicator that is correlated to GDP growth consistently stronger than it is the case with previously analysed series of industrial production. In other words, this supports the view that aggregate real activity in North Macedonia, when measured by quarterly GDP, is really difficult to track accurately using available economic indicators. Without an in-depth analysis of the process of compiling quarterly national accounts in North Macedonia, it may be very hard to address the underlying disconnect between GDP and
monthly indicators in more detail. Another modelling strategy is thus proposed further.

Assuming that high frequency indicators are sufficiently reliable in tracking real activity,\(^8\) in what follows we propose the methodology that isolates important economic signal from quarterly GDP that is consistent with a large number of available economic indicators. By construction, in this way we construct an indicator that attempts to maximise its correlation to both GDP and a large number of available high frequency indicators.

4 Results

4.1 Separating Real Activity From Measurement Error

Quarterly GDP growth may be decomposed into two important components:

\[
GDP_t^Q = signal_t + error_t, \tag{9}
\]

where \(signal_t\) denotes the level of real activity in economy and \(error_t\) is unrelated to underlying economic conditions. It is also labeled herewith as a measurement error. The main assumption in decomposing GDP growth into the economic signal and the measurement error is that the economic signal is a portion of GDP dynamics that is consistent with a large number of available high frequency indicators. To separate that particular component from measurement error as in (9), we rely on factor model framework elaborated in detail in Chapter 2.

Using the PCA, we first estimate monthly factors \(f_t^M\) from a large group of high frequency indicators and aggregate them to quarterly factors \(f_t^Q\). We then estimate the following simple OLS specification to separate the signal and the measurement error:

\[
GDP_t^Q = signal_t + error_t = \alpha + \beta f_t^Q + \varepsilon_t. \tag{10}
\]

The argument used here is that it is the signal only, and not measurement error, that is determined by available economic indicators, so given OLS estimates \(\hat{\alpha} \) i \(\hat{\beta}\) we have:

\[
\begin{align*}
 signal_t &:= \hat{\alpha} + \hat{\beta} f_t^Q \tag{11} \\
 error_t &:= e_t (\text{OLS residual from(10)}). \tag{12}
\end{align*}
\]

\(^8\) High correlation between industrial production in North Macedonia and Germany, for example, supports that assumption.
The equation (10) may be used both for nowcasting $signal_t$ and also for constructing a coincident economic indicator using past historical data as earlier explained:

$$CEI_t^M := \hat{\alpha} + \hat{\beta} f_t^M.$$  \hfill (13)

### 4.2 PCA Analysis

In order to estimate (10) we first need to estimate factors $f_t^M$ from a large group of high frequency indicators $X_1, \ldots, X_N$. A recent literature on nowcasting, however, suggests that in terms of forecasting accuracy it is not always optimal to use the largest available dataset to extract factors from (see for example Boivin and Ng (2006) or Bai and Ng (2008)).\(^9\) It may, therefore, be a good practice to preselect initial dataset and to keep only those variables that are, according to a chosen criterion, sufficiently related to GDP.

There is, however, obvious trade-off involved in this preselection process: a sufficient number of high frequency indicators is needed to include all the available information, but, on the other hand, too many predictors may deteriorate the accuracy of the model. Here we use several informal criteria to preselect high frequency predictors: (i) a good correlation with GDP over different samples; (ii) a broad coverage of economic activities; and (iii) a frequent use in related literature. Obviously, such a selection of variables is not unique.

According to these criteria, 23 variables grouped into seven categories were preselected:

1. **Industry** (**industry\(_1\), \ldots, industry\(_6\)**): Index of industrial production-Total, Index of industrial production-Manufacturing, Index of industrial production-Intermediate goods industries, except energy, Volume index of industrial production-Germany, Volume index of industrial production-Serbia, Volume index of industrial production-Belgium;\(^{10}\)

2. **Internal trade** (**intrtrade\(_1\), intrtrade\(_2\)**): Retail trade, except of motor vehicles and

\(^9\) Indeed that was the case with the dataset used. When using the largest available dataset from Table A1, extracted factors would turn unrelated to GDP growth and thus useless for the purpose of our analysis.

\(^{10}\) A direct consequence of our preselection method is that some of the most important trading partners of North Macedonia are not included in the PCA, in contrast to some minor trading partners. For example, Belgium is a minor trading partner, but its industrial production is highly correlated to GDP in North Macedonia and is, therefore, included in our tested specification.
motorcycles (real), Wholesale trade, except of motor vehicles and motorcycles (real);

3. **External trade** (extrade$_1$, extrade$_2$): Export-thousand US dollars, Import-thousand US dollars;

4. **Economic sentiment** (sentiment$_1$, . . . , sentiment$_7$): Economic sentiment indicator (Germany), Economic sentiment indicator (European union), Economic sentiment indicator (Croatia), Economic sentiment indicator (North Macedonia), Industrial confidence indicator (Germany), Assessment of the production volume of the business entities over the past 3 months, Expectations for the volume of production during the next 3 months;

5. **Employment** (employment$_1$, employment$_2$): Employees in industry-Manufacture of fabricated metal products, except machinery and equipment, Employees in industry-Manufacture of other transport equipment;

6. **Money** (money$_1$, money$_2$): Monetary aggregate M2 in national currency, Broad money;

7. **Government budget** (government$_1$, government$_2$): Central Government Budget-VAT (net), Central Government Budget–Total Revenues;

The unobserved factor is estimated using the PCA method. Table 1 shows that 46 percent of overall variance of 23 high frequency indicators may be explained by a single factor, pointing to a strong common drivers underlying the chosen dataset.

<table>
<thead>
<tr>
<th>Table 1: PCA Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>
Overall, ten components is sufficient to explain almost (more than 90 percent of) total variance of the used dataset. To attach some interpretation to estimated factors, factor loadings of first four components are provided in Figure A1 in the Annex. The first principal component accounts for almost half of the overall variation and loads strongly and uniformly onto wide range of monthly indicators—domestic and foreign industry, sentiment indicators, and external trade. All loadings are of the same sign.11

A second component is by definition orthogonal to the first one, but seems to be more difficult to interpret. The third component, similar to the first one, loads on domestic industry strongly. It is, however, orthogonal to the first one that loads on both domestic and foreign activity and may, therefore, have an interesting interpretation—it captures that part of variation in industrial production in North Macedonia is driven by domestic, idiosyncratic shocks. The three components explain almost 70 percent of the overall variation in the dataset. First eight principal components are provided in Figure A2 in the Annex.

Finally, literature on factor models sometimes uses formal procedures to estimate the number of factors that optimally represent large datasets which may also be helpful in specifying nowcasting equation. However, these methods (see Bai and Ng (2002) for example) develop estimators of true number of factors in factor models relying on large cross sections (N) and large time dimension (T). In the case of North Macedonia, given a handful of variables with any meaningful correlation with GDP growth, N dataset is small, and, therefore, instead of relying on a formal procedure to determine number of factors, we opt for a simpler two-step procedure. First, a ‘targeting’ method is used, as described earlier, to select variables for factor analysis followed up with nowcasting specification which includes only those factors that explain variation in GDP in a statistically significant way.

4.3 Nowcasting model and CEI for North Macedonia

Once monthly factors $f_t^M$ are estimated, they need to be aggregated to quarterly frequency to estimate equation (10). There is, however, another modelling challenge involved here: there is evidence of instability in estimated OLS parameters of the equation. More precisely,

---

11 Principal components are identified up to the constant and therefore, up to the sign only. For that reason we focus on absolute value of estimated factor loadings. For example, the third principal component, as suggested by factor loadings in Figure A1, is negatively related to GDP.
standard Bai-Perron breakpoint test (see Bai and Perron 2008) points to a change in relation between GDP growth and estimated factors around the end of 2010. In order to account for that, the model is specified separately for 2007Q1-2010Q3 and 2010Q4-2020Q3. Accounting for parameter instability in this way is also a trade-off faced in implementing this methodology. By taking the structural break in specification into account, the model tracks observed GDP growth better, but at the same time diverges from the dynamics of high frequency indicators. And vice versa—ignoring the break, the model will be more consistent with a large number of monthly indicators at the cost of following GDP dynamics to a lesser extent. To illustrate that, Table A2 in the Annex reports correlations between real GDP, two model predictions (with and without a structural break), all 23 series used to extract principal components, and the first principal component itself.

As expected, the observed GDP is highly correlated to model predictions once it includes the structural break (83 percent) and less so when the break is ignored (65 percent). The weakest correlation with the GDP growth is with the first principal component because it is not directly related to GDP as oppose to other two indicators. On the other hand, by construction, the first principal component is correlated to all high frequency indicators stronger than the other two model predictions. There is an important practical recommendation resulting from this table: if one is primarily interested in tracking official GDP, irrespective of how well it correlates to high frequency indicators, a model with structural breaks included would be a preferred option.

The estimated specification may be directly used to estimate monthly series of real activity that is consistent with both GDP (through regression parameters $\alpha$ and $\beta$) and a large number of high frequency indicators (via estimated factors $f_t^M$), as shown by equation (13). This indicator, shown in Figure 8 (a), may, therefore, be interpreted as the coincident economic indicator for North Macedonia. Such an indicator is easily updated on a monthly basis after a release of important monthly economic indicators. Figure 8 (b) compares this monthly coincident indicator (in grey) and its 3-months moving average (in blue) with observed GDP growth in North Macedonia. Reported moving averages are by contraction

\[\text{Figure 7 also suggests that forecast accuracy of a standard nowcasting specification using industrial production, retail sales, and lagged GDP deteriorated just around that period.}\]
in-sample predictions of the preferred specification that accounts for the structural break. Overall, the selected specification tracks GDP growth well and manages to capture main characteristics of a business cycle in North Macedonia. This specification, in total, explains around 70 percent of variation in GDP growth in terms of $R^2$ statistics.

**Figure 8. Coincident Monthly Indicator of Economic Activity (a) and In-sample Model Prediction (b)**

Source: National statistical office, staff calculations.

5 Tracking Real Activity with Alternative Sources of Information

In previous chapters, it was illustrated how weakly GDP growth is related to standard monthly predictors of real activity in North Macedonia. For that reason, it may be useful to explore if some alternative data sources, not being published by statistical offices on a regular basis, may help in monitoring real activity on a just-in-time basis. This is even more relevant since the start of the pandemic, when time series observations are dominantly affected by huge shocks. In addition, standard relations between variables have, to a certain extent, been disrupted, and usual correlations between GDP and some standard predictors have weakened globally. In case of North Macedonia it points to further weakening of already poor relations between GDP and its standard predictors.

The strategy of how to exploit information available in these alternative sources of information heavily depends on frequency of their release. For example, in case when many high frequency predictors are available possibly on a weekly or even daily basis, standard framework for nowcasting may be used to track the real activity on a very high frequency
Popular sources of data that attracted attention lately in this context are Google trends and Google mobility data.

A Google mobility index seems to trace levels in industrial production or retail sales in North Macedonia fairly well. However, this series is very short, starting in March 2020, and it is unclear how to backcast the series to be able to use it in a meaningful econometric analysis. Also, given a weak link between the two hard economic indicators and GDP growth these data are not useful to track GDP in real time, but can be of help to track real activity more generally.

In order to explore the possibility that what people search for in North Macedonia may help in signaling the state of the business cycle, we specified a simple model based on Google trends data. For that purpose, a database containing a group of monthly searches for overall 142 words that may indicate the dynamics of business cycle was created. The sample starts in 2011 and covers the period up to the end of 2020. The focus is on some specific words, both in Cyrillic and Latin alphabets that are related to macroeconomy and finance in general, business cycle, labor market, pandemic, etc. Then, by focusing on an indicator that may be useful in tracking the evolution of the economy during the recent crisis, only those searches were investigated that, in contrast to the earlier period, seem to be affected particularly by the COVID 19 pandemic. In that way, we filtered overall 42 key words as predictors. Further, the PCA is applied on that group of time series and tested if these principal components may help in predicting the real activity. For that purpose, we specified a simple bridge equation and related the estimated monthly frequency indicator (Figure 8(a)) with the estimated principal components.
Figure 9. Coincident Monthly Indicator of Google Trends Data

Source: Google, staff calculations.

Figure 9 compares in-sample fit from that model with the estimated monthly indicator. Overall, the analysis suggests that Google trends data may be a very useful and timely indicator of real activity in North Macedonia, in particular during the recent shock. In contrast to the official statistics, these data series are available on a weekly basis. Figure 9 also suggests that information available from this data may be useful in detecting turning points of the business cycle—the two series enter the recent recession simultaneously. However, the magnitude of the recession cannot be captured fully by relying on these data only.

6 Conclusion

The main finding of the analysis is that, unlike in majority of European countries, GDP in North Macedonia is weakly correlated with standard high frequency indicators. A GDP series is very noisy and a detailed examination of its methodology might be warranted since it is very difficult to monitor GDP growth in real time based on higher frequency data.

To overcome this, this paper proposes the methodology to separate the economic signal from the measurement error in GDP. The economic signal is by construction correlated with a large number of high frequency indicators whose overall variation is summarized using the PCA.

In an illustrative specification in this paper, we extract principal components from a group of 23 monthly indicators. Additionally, to account for changing relationship between high frequency indicators and GDP growth over time, the preferred model specification also includes the structural break observed in GDP growth in North Macedonia in 2010. The final
model explains around 70 percent of variation in the overall GDP growth in terms of $R^2$ statistics, and may be used both for nowcasting and constructing the coincident economic indicator for North Macedonia. Such an indicator is easily updated on a monthly basis after the release of important monthly economic indicators.

Finally, in an environment where GDP is practically unrelated to standard high frequency predictors it may be useful to exploit some alternative indicators when tracking the real activity in real time. In that context Google trends data have been explored as a predictor of real activity. The analysis indeed shows that they can help signal the state of the business cycle much earlier in comparison to models that rely on official statistics.
References


Figure A1. Factor Loadings
Table A1. High Frequency Indicators - Categories and Sources

<table>
<thead>
<tr>
<th>Series / Group of Series</th>
<th>Time Span</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP, Chain linked volumes</td>
<td>2000Q1-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Indices of industrial production</td>
<td>2000M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Completed construction works (real)</td>
<td>2010M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Tourist arrivals and tourist nights spent</td>
<td>2000M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Deflated turnover sales in retail trade</td>
<td>2010M01-(Eurostat); 2011M01-(Makstat)</td>
<td>Eurostat / Makstat</td>
</tr>
<tr>
<td>Passenger, freight and mail traffic at the airports</td>
<td>2011M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Purchase and sale of agricultural products</td>
<td>2010M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Number of new employment contracts</td>
<td>2003M12-</td>
<td>Employment service agency (av.gov.mk)</td>
</tr>
<tr>
<td>Exchange rates (EUR/MKD, USD/MKD)</td>
<td>2000M01-</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Global price of Brent Crude</td>
<td>2000M01-</td>
<td>Federal Reserve Bank of St. Louis</td>
</tr>
<tr>
<td>Average paid gross wage per employee (denars)</td>
<td>2005M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Average paid net wage per employee (denars)</td>
<td>2005M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Exports, imports, trade balance and coverage of imports by exports (in EUR&amp;USD)</td>
<td>2004M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Consumer Price Index</td>
<td>2002M01-(COICOP); 2005M01- (subindices exc. energy&amp;food)</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Industrial producer price indices</td>
<td>2010M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Business tendencies in manufacturing industry</td>
<td>2001M04-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Electricity (MWh)</td>
<td>2012M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Natural gas (million nm3)</td>
<td>2012M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Economic sentiment indicator</td>
<td>2008M05- (N. Macedonia)</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Consumer confidence indicator</td>
<td>2012M05- (N. Macedonia)</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Industrial confidence indicator</td>
<td>2008M05- (N. Macedonia)</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Volume index of industrial production</td>
<td>2000M01- (N. Macedonia)</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Indices of employees in industry</td>
<td>2005M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Turnover indices in industry</td>
<td>2010M01-</td>
<td>State Statistical Office (Makstat)</td>
</tr>
<tr>
<td>Dataset</td>
<td>Date</td>
<td>Source</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Index of the real effective exchange rate of the denar (2015=100)</td>
<td>2002M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>NBRNM’s Balance sheet</td>
<td>2003M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Other depository corporations’ (banks and saving houses) balance sheet</td>
<td>2003M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Depository corporations’ (NBRNM, banks and saving houses) survey</td>
<td>2003M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Weighted interest rates on outstanding amounts on granted loans and received deposits</td>
<td>2005M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Loans granted by other depository corporations (banks and savings houses)</td>
<td>2003M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Banks’ and savings houses loans granted to households</td>
<td>2009M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Broad money and components</td>
<td>2003M01-</td>
<td>NBRNM</td>
</tr>
<tr>
<td>Central Government Budget (consolidated)</td>
<td>2006M01-</td>
<td>Ministry of Finance</td>
</tr>
<tr>
<td>Retail and wholesale trade, except of motor vehicles and motorcycles (real)</td>
<td>2002M01-</td>
<td>Constructed series from old nominal series</td>
</tr>
</tbody>
</table>
Figure A2. First Eight Principal Components
<table>
<thead>
<tr>
<th>var</th>
<th>GDP</th>
<th>Model (With Break)</th>
<th>Model (No Break)</th>
<th>PC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>1.00</td>
<td>0.83</td>
<td>0.65</td>
<td>0.57</td>
</tr>
<tr>
<td>Model (with break)</td>
<td>0.83</td>
<td>1.00</td>
<td>0.77</td>
<td>0.67</td>
</tr>
<tr>
<td>Model (No Break)</td>
<td>0.65</td>
<td>0.77</td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>PC1</td>
<td>0.57</td>
<td>0.67</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>extrade2</td>
<td>0.45</td>
<td>0.47</td>
<td>0.53</td>
<td>0.77</td>
</tr>
<tr>
<td>government1</td>
<td>0.44</td>
<td>0.61</td>
<td>0.66</td>
<td>0.57</td>
</tr>
<tr>
<td>sentiment5</td>
<td>0.22</td>
<td>0.34</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>intrade1</td>
<td>0.43</td>
<td>0.47</td>
<td>0.56</td>
<td>0.47</td>
</tr>
<tr>
<td>sentiment1</td>
<td>0.29</td>
<td>0.42</td>
<td>0.66</td>
<td>0.76</td>
</tr>
<tr>
<td>extrade1</td>
<td>0.43</td>
<td>0.47</td>
<td>0.67</td>
<td>0.88</td>
</tr>
<tr>
<td>ind4</td>
<td>0.45</td>
<td>0.56</td>
<td>0.81</td>
<td>0.92</td>
</tr>
<tr>
<td>money2</td>
<td>0.36</td>
<td>0.39</td>
<td>0.24</td>
<td>0.41</td>
</tr>
<tr>
<td>ind3</td>
<td>0.40</td>
<td>0.46</td>
<td>0.52</td>
<td>0.65</td>
</tr>
<tr>
<td>ind2</td>
<td>0.54</td>
<td>0.60</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>sentiment2</td>
<td>0.32</td>
<td>0.48</td>
<td>0.71</td>
<td>0.75</td>
</tr>
<tr>
<td>ind5</td>
<td>0.36</td>
<td>0.40</td>
<td>0.63</td>
<td>0.67</td>
</tr>
<tr>
<td>money1</td>
<td>0.32</td>
<td>0.34</td>
<td>0.44</td>
<td>0.59</td>
</tr>
<tr>
<td>intrade2</td>
<td>0.54</td>
<td>0.63</td>
<td>0.66</td>
<td>0.59</td>
</tr>
<tr>
<td>employment1</td>
<td>0.20</td>
<td>0.32</td>
<td>0.51</td>
<td>0.46</td>
</tr>
<tr>
<td>ind6</td>
<td>0.50</td>
<td>0.58</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td>sentiment7</td>
<td>0.37</td>
<td>0.49</td>
<td>0.77</td>
<td>0.88</td>
</tr>
<tr>
<td>employment2</td>
<td>0.29</td>
<td>0.19</td>
<td>0.26</td>
<td>0.14</td>
</tr>
<tr>
<td>ind1</td>
<td>0.63</td>
<td>0.67</td>
<td>0.73</td>
<td>0.76</td>
</tr>
<tr>
<td>sentiment6</td>
<td>0.37</td>
<td>0.39</td>
<td>0.69</td>
<td>0.81</td>
</tr>
<tr>
<td>government2</td>
<td>0.65</td>
<td>0.72</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td>sentiment4</td>
<td>0.65</td>
<td>0.67</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>sentiment3</td>
<td>0.45</td>
<td>0.53</td>
<td>0.89</td>
<td>0.85</td>
</tr>
</tbody>
</table>