AN ASSESSMENT OF MACROECONOMIC PERFORMANCE INDICATORS IN KAZAKHSTAN VIA FACTOR MODELS

by

Yelaman Rashitov

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

in

Economics

at

NAZARBAYEV UNIVERSITY
SCHOOL OF SCIENCES AND HUMANITIES

2020

1. Abstract

We use factor models to assess accuracy of macroeconomic performance indicators in Kazakhstan over 2011-2018. Specifically, the annual GDP growth rate and unemployment rate are analyzed based on the fundamental factors underlying the economy. The model is a version of dynamic factor models with exact factor structure. We exploit a two-step estimator that provides consistent estimates of the factors when $n \to \infty$ and $T \to \infty$. The findings indicate that there are two fundamental factors responsible for driving real and nominal variables in the economy. These two factors are just vectors whose entries were estimated via principal components and Kalman Filter. Although the growth rate of GDP from public sources indicate that there has not been a recession in the country, we find that the alternative GDP growth rate implied by the fundamental factors points to the recession in 2016. In addition, the finding is supported by the backcasts of unemployment rate which shows us that the unemployment rate started rising after the severe decline in oil prices that negatively affected the Kazakhstani economy. We also find that the official unemployment rate was below the backcasts of unemployment rate by 2 percentage points on average.

2. Introduction

Most of the modern literature on empirical macroeconomics have focused on either uncovering the fundamental sources of business cycle fluctuations or forecasting important endogenous variables in the economy. Especially, this has been the case for advanced economies with a large number of variables available for an adequatley long period of time. At the same time, the study of emerging economies has not progressed much with the rapid development of sophisticated empirical tools in macroeconomics as the issue lies not only in the amount of data available but also in their quality and accessibility. As a result, emerging economies have been treated unequally in the literature with main advancements in empirical macroeconomics being tested and used on the data for developed countries. In fact, the lack of quality and time series observations of the data on emerging economies makes the modern tools in macroeconometrics particularly useful for them in the first place. For example, factor models allow to project observable variables on few common factors to obtain an index of that variable whereas Bayesian methods allow to estimate models with a large number of parameters even if the number of time series observations is small. Hence, the issues handled by the modern macroeconometric techniques can largely resolve problems with the quality and quantity of data available in emerging economies.

In this paper we attempt to assess and construct alternative growth rate of GDP and unemployment rate for Kazakhstan using a large panel of time series on economic variables. Hence, we estimate a version of dynamic factor models (DFM) to extract common factors responsible for the comovement among the large number of variables in the economy. The application of factor models for the data in Kazakhstan is an interesting piece of research in its own right, but the comparison of official economic activity performance to the ones implied by the large number of variables is more relevant. Abilov et al. (2019) build a forecasting model for Kazakhstan and document inaccuracies and measurement errors present in the official macroeconomic data. As a result, there is not much trust economists can have in the official statistics which makes us use factor models to capture the dynamics of the business cycle that is due to common factors whereas measurement errors enter idiosyncratic components. But it does not mean that we do not trust on collected data at all. We assume that the data is credible enough to construct the factors mentioned above. Since the index is constructed using only the common factors, the measurement errors do not affect the factors, and hence the new economic performance indicators. Therefore, the factor model overcomes the problem of inaccurate statistics on economic activity and allows us to obtain more precise estimates of economic performance in Kazakhstan.

3. Literature review

Stock and Watson (1989) constructed a leading economic indicator for the US economy via a low-dimensional factor model estimated by maximum likelihood technique and Kalman filter. Aiolfi et al. (2006) adopt similar methodology and principal components to construct and analyze business cycles for Latin American countries. Galli (2017) builds a monthly business cycle index for the Swiss economy that can be updated in real time as the new information releases become available. Some well-known indexes constructed via factor models are Eurocoin and EuroSTING indexes for the Euro area, and the Aruoba-Diebold-Scotti Business Conditions Index (see Altissimo et al. (2010); Camacho and Perez-Quiros (2010); Aruoba, Diebold, and Scotti (2009)).

Although the abovementioned works share a similar objective, which is to construct a business cycle index, they use distinct estimation techniques to arrive at the final result. The methodology for estimating factors and parameters have gone through several stages of development as outlined in Stock and Watson (2011). The earliest versions of factor models in Geweke (1977) and Sargent and Sims (1977) were based on frequency domain methods. Stock and Watson (1989) employed the time domain of the data to estimate parameters and the factors via maximum likelihood and Kalman filter. The seminal works by Bai and Ng (2002), Forni et al. (2000), Forni et al. (2005), and Stock and Watson (2002) popularized the use of non-parametric methods by proving the consistency of principal components in estimating the space spanned by the factors. Principal component analysis appeared as a much easier way of obtaining the factors given the large cross-sectional dimension of the data. Boivin and Ng (2006) used a slight modification of principal components by weighting the data with the sample error covariance matrix whereas Forni et al. (2000) imposed a dynamic structure on the common component and non-orthogonality of idiosyncratic components. After obtaining the estimated factors one must project these factors on an observable variable to construct an index of that variable. Stock and Watson (2002) introduced a diffusion index forecast for predicting business cycles and proved consistency of the feasible forecast whereas Bai and Ng (2002) derived limiting distributions of parameters, forecasts and forecast errors in the factor-augmented regression.

The last fifteen years have seen a rapid progress in factor model estimation starting with the seminal work by Giannone, Reichlin, and Sala (2005) who introduced a two-step estimator of the factors by estimating parameters via ordinary least squares (OLS) based on principal components, and then using the Kalman smoother to arrive at better estimates of the factors. Giannone, Reichlin, and Small (2008) employ the same methodology to estimate the factors and nowcast the GDP growth of the US economy. A theoretical justification for the use of the two-step estimator was provided later in the work by Doz, Giannone, and Reichlin (2011) who proved the consistency of the estimated factors, but the main advantage of the two-step estimator is its ability to deal with missing observations in the data that allow us to address the issue of incomplete datasets. Doz, Giannone, and Reichlin (2006) show that the quasi-maximum likelihood (QML) estimator provides more accurate estimates of the factors compared with simple principal components and two-step estimators, when the model is subject to the misspecification of omitted cross-sectional correlation of idiosyncratic components. They also prove consistency of the quasi-maximum likelihood estimator when the time and cross-section dimensions of the data are large ($T \to \infty$ and $n \to \infty$). Overall, the techniques used for estimating factor models include the following: maximum likelihood estimation with Kalman filter, principal component analysis, two-step estimator, and QML estimator.

The structure of the paper is as follows. Section 4 presents two ways for the estimation of the factors which are deemed to be the major driving forces in the economy. Section 5 introduces the dataset that is the main input into the model. In Section 6 we discuss the results of the model and explain the estimated factors using factor loadings and graphical illustration. Finally, concluding remarks are made with the aim of raising further issues on the subject.

4. Model

4.1. Kalman Filter

State space model is a linear time series model consisted of observed and unobserved variables. This model has two main equations. The first one is a state equation:

$$S_t = FS_{t-1} + u_t \tag{1}$$

where u is normally distributed and F is the system matrix which is obtainted once we initialize the parametres via principal components. The second equation is a measurement equation:

$$Z_t = HS_t + v_t \tag{2}$$

where v is normally distributed. S_t is unknown but can be estimated using previous values available up to time period t-1. We denote this estimated as $S_{t|t-1}$. Then we assume that error term $S_t - S_{t|t-1}$ is normally distributed. So, our model has the following representation:

$$Z_t = HFS_{t|t-1} + v_t + H(S_t - S_{t|t-1}) \tag{3}$$

where $S_{t|t-1}$ is observable, v_t and $S_t - S_{t|t-1}$ are normally distributed. Therefore we can estimate the model above by maximum-likelihood method. If we substitute equation 1 into equation 2 we can get the formula below. Also, there is an assumption that the error term is normally distributed

$$v_t + H(S_t - S_{t|t-1}) \sim N(0, \Omega_t) \tag{4}$$

where $\Omega_t = \Sigma^v + H\Sigma^S_{t|t-1}H'$. Now we need to estimate the state variables at time given the information at time t-1.

$$S_t = FS_{t-1} + u_t \rightarrow S_{t|t-1} = FS_{t-1|t-1} \tag{5}$$

Then

$$Z_{t|t-1} = HS_{t|t-1} = HFS_{t-1|t-1} \tag{6}$$

By conditional expectations we can find minimum variance estimate of S_t provided observable Z_t

$$E(S_t|Z_t) = S_{t|t} = FS_{t-1|t-1} + K_t(Z_t - HFS_{t-1|t-1}) \tag{7}$$

where

$$K_t = (H\Sigma^S_{t|t-1})'(\Sigma^v + H\Sigma^S_{t|t-1}H')^{-1} \tag{8}$$

is called Kalman gain matrix. It can be updated using the following formula:

$$\Sigma^S_{t|t-1} = F\Sigma^S_{t-1|t-1}F' + \Sigma^u \tag{9}$$

$$\Sigma^S_{t|t} = (I - K_t H)\Sigma^S_{t|t-1} \tag{10}$$

The log likelihood function is

$$f(Z_1, Z_2, ..., Z_T|S_{1|0}, \theta) = f(Z_1|S_{1|0}, \theta)\Pi_{i=2}^{i=T} f(Z_i|Z_{i-1}, \theta) \tag{11}$$

First we give an initial value for theta, then estimate factors via Kalman filter. The next step is to plug them into log-likelihood function. Then we change the value of theta and repeat the procedure. We choose the combination of vectors and their associated factors that yield the highest value of log-likelihood function.

4.2. Principal component analysis

One of the most powerful tools in factor analysis is the principal components approach. Principal components estimator is used to estimate factors using eigendecomposition of the variance-covariance matrix. In other words, we use the correlation matrix of the dataset with n variables to calculate the factors.

That is,

$$\tilde{F}_t(N^{-1}W) = N^{-1}W'X_t \tag{12}$$

where $W = \tilde{\Lambda}$. $\tilde{\Lambda}$ is the matrix of eigenvectors of the sample variance-covariance matrix of X_t corresponding to the r biggest eigenvalues of of the correlation matrix. The matrix X is our dataset which contains 58 variables. Our next goal is to construct covariance matrix for those variables. The matrix of eigenvectors Λ can be obtained once we initialize the parametres via Principal Components using Matlab.

In principle, the principal component estimator can be considered as the solution of the following optimization problem

$$\min_{F_1,\ldots,F_T,\Lambda} V_r(\Lambda, F) \tag{13}$$

subject to

$$N^{-1}\Lambda'\Lambda = I_r \tag{14}$$

where $V_r(\Lambda, F) = \frac{1}{NT}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t)$. In fact, it can be shown that the optimization problem given above is equivalent to the following maximation problem

$$\max_{\Lambda} \Lambda'\tilde{\Sigma}_{XX}\Lambda \tag{15}$$

subject to

$$N^{-1}\Lambda'\Lambda = I_r \tag{16}$$

The principal components estimator of F_t is consistent for given T and $N \to \infty$. As a result, we obtain the estimated factors $\tilde{F}_t$ which are considered as the principal component estimates of the factors.

4.3. Two-step procedure

As for the two-step procedure we specify a state-space representation of the following form

$$
\begin{aligned}
X_t &= \Lambda F_t + e_t \\
F_t &= A F_{t-1} + u_t
\end{aligned}
\tag{17}
$$

where $u_t = [u_{1t},\, u_{2t}]'$ is the two-dimensional vector of common shocks that is multivariate normally distributed. First, we estimate the factors of the model via principal components estimator. As a result, we obtain $\tilde{F}_t$ and run the regressions in (17) by using ordinary least squares. That is, we run the OLS regression of observables variables X_t on the principal component estimates of the factors, $\tilde{F}_t$. In addition, the VAR(1) model is estimated for the factors. As a result, we obtain the coefficients of the state-space model denoted by $\hat{\boldsymbol{\theta}} = \{\hat{\Lambda}, \hat{B}, \hat{\Psi}, \hat{\Phi}\}$. Finally, we use these parameters of the state-space model by treating the factors in (17) as unobservable and calculate them via Kalman smoother. Hence, we obtain the two-step estimates of the factors denoted by $\hat{F}_t$ that are considered as the fundamental factors in the economy.

There are two classical assumptions that we make on the factor structure:

1. Common factors are pervasive:

$$\liminf_{n\to\infty} \left(\frac{1}{n} \Lambda' \Lambda \right) > 0 \tag{18}$$

2. Idiosyncratic factors are nonpervasive:

$$\lim_{n\to\infty} \frac{1}{n} \left(\max_{v'v=1} v' \Psi v \right) = 0 \tag{19}$$

More formally, we obtain the principal component estimates of the factors from the optimization problem given by (15) subject to (16).
Next we define $r \times r$ diagonal matrix whose diagonal elements are the largest r eigenvalues of covariance matrix, also let V be $n \times r$ matrix of respective eigenvectors under the normalization $V'V = I_r$. Then the principal component

estimates of the factors are given by:

$$\tilde{F}_t = V' x_t \tag{20}$$

where x_t is the covariance matrix. The matrices Λ and Ψ can be estimated by regressing the variables on factors:

$$\hat{\Lambda} = \sum_{t=1}^{T} X_t \tilde{F}_t \left(\sum_{t=1}^{T} \tilde{F}_t \tilde{F}_t \right)^{-1} = V \tag{21}$$

$$\hat{\Psi} = S - VDV' \tag{22}$$

We use VAR to estimate other parameters:

$$\hat{A} = \sum_{t=2}^{T} \tilde{F}_t \tilde{F'}_{t-1} \left(\sum_{t=2}^{T} \tilde{F}_{t-1} \tilde{F'}_{t-1} \right)^{-1} \tag{23}$$

$$\hat{\Phi} = \frac{1}{T-1} \sum_{t=2}^{T} \tilde{F}_t \tilde{F}'_t - \hat{A} \left(\frac{1}{T-1} \sum_{t=2}^{T} \tilde{F'}_{t-1} \right) \hat{A}' \tag{24}$$

Let P be $q \times q$ matrix with elements given by the biggest q eigenvalues of $\hat{\Phi}$ and M be $r \times q$ matrix with elements of respective eigenvectors:

$$\hat{B} = MP^{1/2} \tag{25}$$

The estimates are consistent if n and $T \to \infty$ Using the estimated parameters of factor models it is possible to reestimate the factors in the following way:

$$\hat{F}_t = Proj[F_t | x_1, ..., x_T], t = 0, 1, ..., T \tag{26}$$

Next, we replace estimated parameters in given equations and then apply Kalman smoother:

$$x_t = \hat{\Lambda} F_t + \xi_t \tag{27}$$

$$F_t = \hat{A}F_{t-1} + \hat{B}u_t \quad (28)$$

$$u_t \sim WN(0, I_q) \quad (29)$$

$$E(\xi_t \xi_t') = diag(\hat{\Psi}) \quad (30)$$

$$\hat{u}_t = P^{-1/2}M'(\hat{F}_t - A\hat{F}_{t-1}) \quad (31)$$

The Kalman smoother gives us the two-step estimates of the fundamental factors of the economy driving the variables in X_t.

5. Data description

The data is collected from the public sources of the Committee of Statistics of the Republic of Kazakhstan, National Bank, NAC Data is collected from public sources: 1) Statistics Committee of the Republic of Kazakhstan; 2) National Bank of the Republic of Kazakhstan; 3) NAC Analytica; 4) Kazakhstan Stock Exchange (KASE); 5) Bloomberg. We use monthly data on 57 economic variables that can be categorized by sectors: real sector, labor market, price indexes, financial sector, banking sector and external sector (see Appendix A). We also collect two time series on unemployment rate from independent sources. The first one is obtained from the Statistics Committee that publishes monthly data on labor market conditions. The other one is obtained from the Survey Centre of NAC Analytica that conducts survey of 3,000 households every month from June 2015. The unemployment rate index of NAC Analytica is used as an alternative measure of unemployment rate since there is more variation in this unemployment rate compared with the official figure that has changed a little over the last years (see Figure 1). We use monthly data only from 2010 to 2018 due to the difference in the methodology of the data collection process before and after 2010 in the Statistics Committee. There is also seasonality in a set of variables from the real sector and banking sector. Census X-13 has been used to eliminate seasonality present in the variables. We take yearly growth rates of variables that appear in levels. Other nonstationary variables are taken in first differences. When all variables are transformed into stationary form we standardize them to convert into the same scale of measurement.

5.1. Data issues

There is a difference in the methodology of calculating GDP in Kazakhstan and most of the developed countries. More specifically, if developed countries construct a quartely growth rate of GDP as the main indicator of economic activity, the main approach to calculating GDP in Kazakhstan is the production approach whereby output of each of the industries from the real sector are calculated every month and added up to produce the final GDP figure for the month. In developed economies, a quarterly GDP figure is used as the official economic activity indicator, where they adopt the expenditure approach to calculating GDP that is much harder to measure every month. The expenditure approach to calculating GDP comprises of collecting the data on consumption, investment, government spending and net exports to produce final GDP. Kazakhstan also uses the expenditure approach to calculating GDP, but it is not the main approach in the methodology of the Statistics Committee. The difference between the GDP from the production and expenditure approaches is almost 60 percent in some quarters as it was found in the paper by Abilov et al. (2019). This is one of the main reasons that generates doubts on the quality of official releases on the growth rate of GDP in Kazakhstan. In addition, the official unemployment rate in Kazakhstan tends to have a positive correlation of 0.41 with GDP growth rate, which is in stark contrast with mainstream view of negative correlation. The traditional view in economics states that unemployment rate is negatively correlated with GDP growth rate, because more workers are hired in times of strong aggregate demand that pushes up production. Overall, the unemployment rate does not seem to respond to the periods of expansions and contractions in economic activity as shown in Figure 1. Therefore, the alternative measure of unemployment rate is obtained from the Survey Centre of NAC Analytica which has a correlation of -0.61 with GDP growth rate in line with the traditional wisdom in economics. The survey contains only one question on the employment status of an interviewee. That is, the question asks which of the categories in the list of possible answers best describes his current employment situation. Unfortunately, the precise formulation of the question and the answers cannot be disclosed for privacy reasons, but this measure of unemployment rate is the only alternative measure available for Kazakhstan. The alternative measure of unemployment rate shall be used to form the backcasts for the period before June 2015 to compare it with the official unemployment rate.

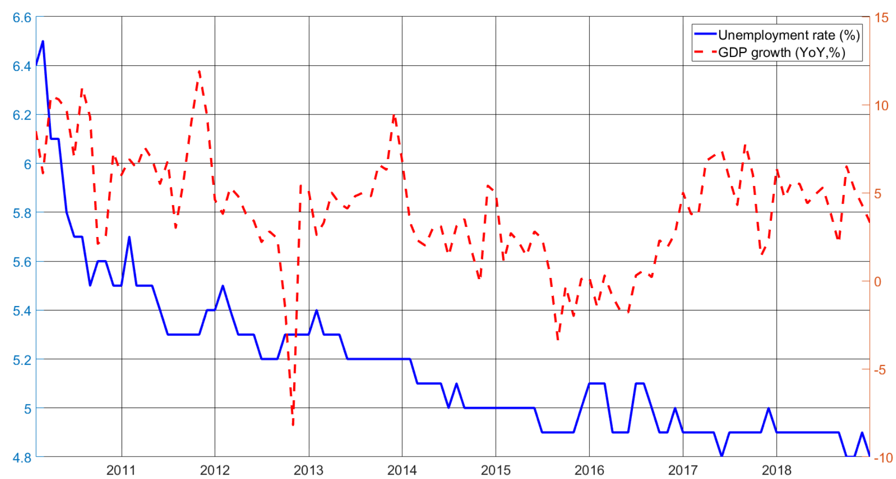

Figure 1: Official unemployement rate and GDP growth rate in Kazakhstan

6. Results

The model given by (17) is estimated via two-step procedure with two factors having a VAR(1) as the underlying process. The estimated factors are presented in Figure 2 and 3. The first factor reminds business cycles in the economy since it tends to fall abruptly during 2015 when there were both speculative attacks on the national currency of Kazakhstan and the decline in oil prices which resulted in economic downturn. The economy started to recover in 2016 when the value of the domestic currency and prices of export products of the country stablized in the world markets. In general, the first factor has a strong positive correlation with the variables such as per capita real income, wages, loans and volume indexes in the real sector (see Appendix). These variables tend to be procyclical with the business cycle, and hence the first factor can be considered as a factor driving real variables in the economy.

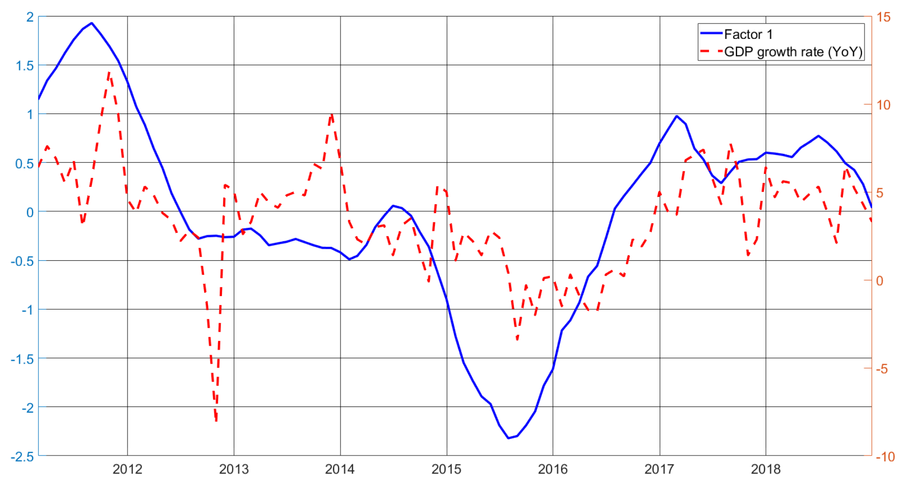

Figure 2: Dynamics of the factor underlying flutuations in real variables (factor 1)

The second factor declines significantly until June 2015 and skyrockets in January 2016. The factor closely reminds the behavior of inflation rate in the economy throughout the period, because the annual inflation rate slowed down in June 2016 to 3.3% and spiked after 80% devaluation of the domestic currency in August. As a result, the inflation rate rose above 10% in November 2015 to reach a peak of 17.7% in August 2016. Overall, the second factor positively correlates with price indexes and monetary aggregates, but it negatively correlates with all of the real sector variables. As a result, the second factor can summarize a fundamental source of fluctuations that underlie nominal variables since it tends to be procyclical with prices and countercyclical with real variables. Giannone et al. (2005) also find that there are two fundamental factors underlying the US economy which are responsible for fluctuations in real and nominal variables respectively.

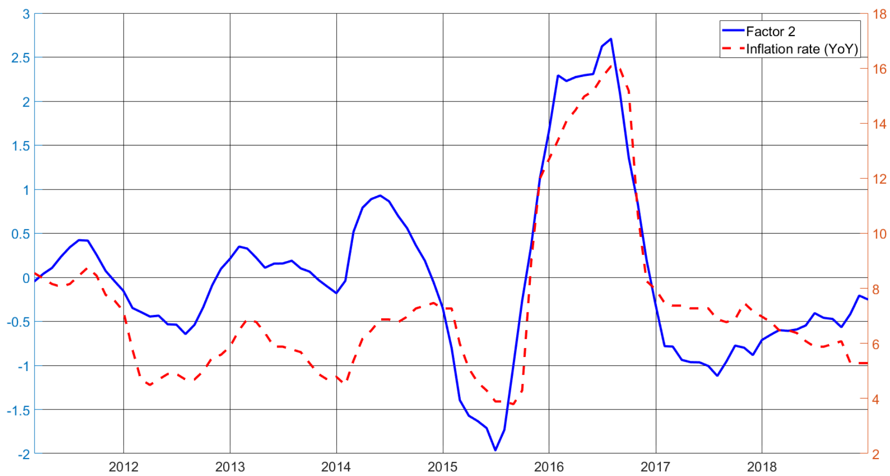

Figure 3: Dynamics of the factor underlying fluctuations in nominal variables (factor 2)

The fundamental factors underlying the economy are used for obtaining time series of annual growth rate of output that is theoretically free from measurement errors and alignments present in the official data. Hence, we estimate OLS regression of annual growth rate of output on the two factors under consideration. The estimated regression equation is given below.

$$\begin{aligned} y_t = \underset{(0.234)}{3.540} + \underset{(0.237)}{1.802}\hat{f}_{1,t} - \underset{(0.2405)}{1.014}\,\hat{f}_{2,t} \\ R^2 = 0.45,\ \bar{R}^2 = 0.44,\ T = 95 \end{aligned} \tag{32}$$

where y_t is annual output growth; $\hat{f}_{1,t}$ and $\hat{f}_{2,t}$ are the factors estimated via the two-step procedure and T is the sample size. The regression output in (32) shows us that the factors account for almost half of the variation in output growth, meaning that the other half is due to the measurement errors and data alignments not captured by the factors. Figure 4 shows the dynamics of actual output growth and fitted value of the regression given in (32). As the figure shows the fitted value repeats the pattern of annual output growth accurately, but it does not capture fluctations around the trend. However, output growth on monthly frequencies are less prone to alignment compared with quarterly GDP growth rates that are closely watched by international agencies. Hence, we calculate quarterly GDP growth rate using the fitted monthly output growth rates obtained from (32).

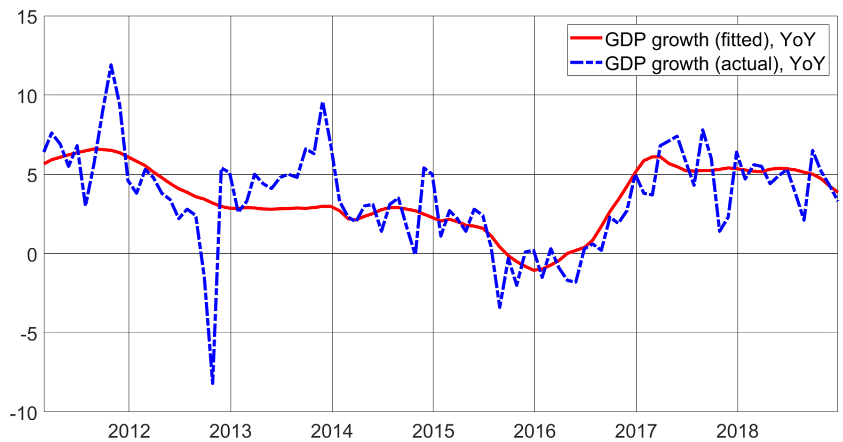

Figure 4: Comparison

Figure 5 shows us annual GDP growth rate obtained from the fitted monthly output growth by averaging the monthly growth rates over quarter. Hence, the figure shows us the annual growth rate of GDP on quarterly basis. It is clearly shown by the shaded area there were three quarters of negative GDP growth from 2015Q4 to 2016Q2 due to the sharp decline in oil prices and the downturn of the Russian economy which is a close trading partner of Kazakhstan. However, the figure cannot be compared with official quarterly publications on annual GDP growth since the Committee of Statistics publishes the data on GDP growth on cumulative basis. That is, the growth rate of GDP is calculated as the ratio of the first quarter relative to the first quarter of previous year whereas for the second quarter the official growth rate is the ratio of GDP of the first two quarters to the GDP of first two quarters of the previous year. They carry on in the same way for third and fourth quarters, meaning that the growth rate of GDP published for the fourth quarter coincides with the annual growth rate of GDP for the year.

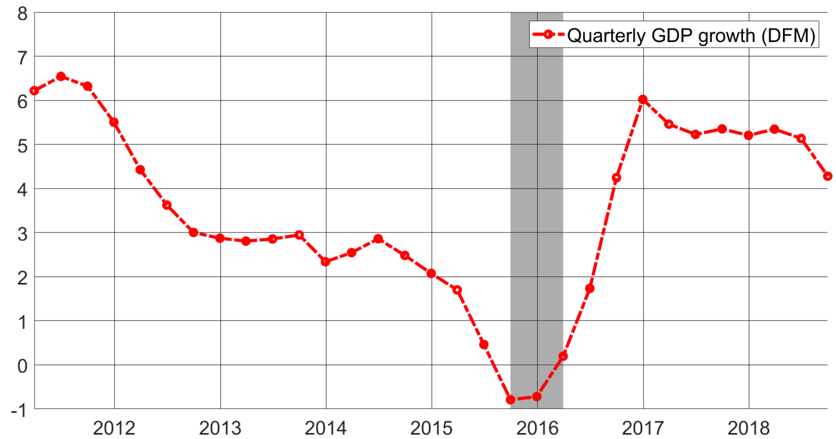

Figure 5: Annual GDP growth on quarterly basis

We convert the quarterly GDP growth rates obtained from the model in the cumulative quarterly growth rates in order to compare the growth rate of GDP implied by the factors and official GDP growth rates published by the Committee of Statistics. Figure 6 plots official GDP growth rate and the growth rate of GDP implied by the factors. As the figure shows both official and factors-implied growth rates of GDP had a downward tendency from 2013 to 2016. The official growth rate of the economy is above the growth rate implied by the factors from 2012 to the end of 2014 the largest difference occurring in the fourth of 2013 when the official growth rate peaked at 6% but the factors yield the growth rate of 2.9%. However, the gap closed over time until the trough of economic crisis in 2016. Although official GDP growth rate statistics deny the occurrence of recession in 2015 and 2016, the growth rate of GDP implied by the factors indicates that there was a recession in Kazakhstan at the beginning of 2016. However, the official GDP growth rates are lower from the fitted ones after 2017 by 1.5% on average when oil prices started to regain their position. In addition, Kazakhstan hosted an international event Expo-2017 that was accompanied by large spending programs on infrastructure and preparation works for the event during the year. As of fourth quarter of 2018 the official GDP growth rate stands at 4.1% whereas the growth rate implied by the factors yield 4.9%.

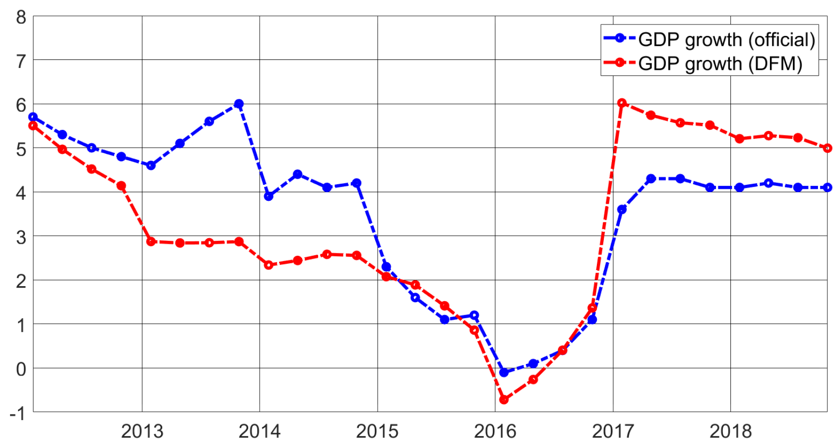

Figure 6: Annual GDP growth rate: official and fitted

Although the methodological framework for determining fundamental factors and theoretical rigor for assessing the GDP growth rate are compelling, there is not certainty the framework predicts accurately enough in out of sample. Hence, we also contruct unemployment rate figures for Kazakhstan using the unemployment rate figures available from the monthly surveys over 2015 to 2018. But we restrain from using official unemployment rates in the regression due to the lack of variation in the variables. The independent survey data contains monthly unemployment rates from June 2015 to December 2018. Hence, we use this sample to estimate the regression of unemployment rate on the two factors.

$$
\begin{gathered}
u_t = \underset{(0.169)}{6.822} - \underset{(0.171)}{0.908}\hat{f}_{1,t} - \underset{(0.134)}{0.163}\hat{f}_{2,t} \\
R^2 = 0.41,\ \bar{R}^2 = 0.38,\ T = 43.
\end{gathered}
\tag{33}
$$

where u_t is the unemployment rate. The fundamental factors explain 41% of variation in unemployment rate as calculated by the survey. It is also noteworthy to point out that the first factor which is deemed to be responsible for fluctuations in real variables affects unemployment rate negatively. That is, more favourable fundamental developments in real variables lead to a decrease in the unemployment rate in line with the common view in economic literature. It can also be seen from (33) that the second factor also affects unemployment rate negatively, but the coefficient is insignificant. Although the short run Phillips curve points to a negative relationship between inflation rate and unemployment rate which precisely corresponds to the sign of relationship given above, there is not much agreement in the literature on the influence of nominal variables on unemployment rate.

We use (33) to backcast unemployment rate until February 2011. Figure 8 shows unemployment rate from survey data, backcasts and official sources. The blue line represents the survey data used for estimation whereas the dashed red line represents backcasts obtained from the fundamental factors in the economy. The time series for official unemployment rate lies entirely below the survey data and backcasts, which points in the direction of either large measurement errors or alignments present in the data. The official unemployment rate neither fell nor rose during the recession in 2015 and 2016 whereas the backcasts show large increases in the unemployment rate from June 2014 when oil prices entered into free fall. The backcast of unemployment rate shows that it reached 8.9% in May 2015. Overall, the backcasts estimated using the fundamental factors indicate that the unemployment rate was on average 2 percentage points higher than the officially reported unemployment rate. The gap widens when the expansion of the economy slows down whereas the actual unemployment rate does not respond at all.

7. Conclusion

The research addresses the concern on public reports of macroeconomic indicators since inaccurate data prevents undertaking sound research and makes

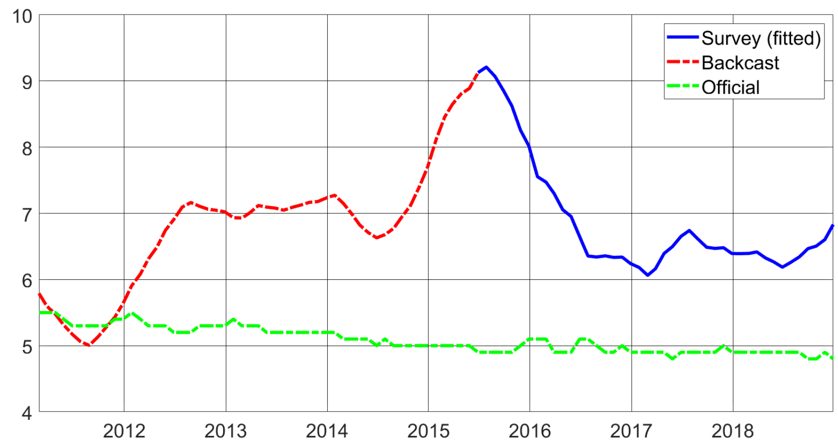

Figure 7: Unemployment rate

policy-making practice difficult based on the official data. Hence, we use a large monthly dataset on Kazakhstani economy to estimate the factors that act as the main fundamental forces driving the economy. The findings suggest that two factors are enough for capturing fundamental comovements among the large number of economic variables. We also estiablish that the first factor mainly drives real variables whereas the second factor is responsible for movements in the nominal variables. Since the factors are fundamental driving forces in the economy, we assume that the part of annual GDP growth rate not attributable to these factors correspond to measurement errors. As a result, the regression of annual GDP growth on the factors concludes that the factor driving real variables positively affects annual GDP growth rate whereas the second factor has a negative impact on it. The fitted values from the regression gives us annual GDP growth rate implied by the fundamental factors in the economy. The paper shows that there was a recession in Kazakhstan in 2016, even though official data does not record the recession and overestimates the growth rate of GDP.

In addition, we observe a constant unemployment rate in official data that does not match the states of the economy over the sample horizon. The official unemployment rate has been at 5% when GDP growth rate was 4% and 1.1%.

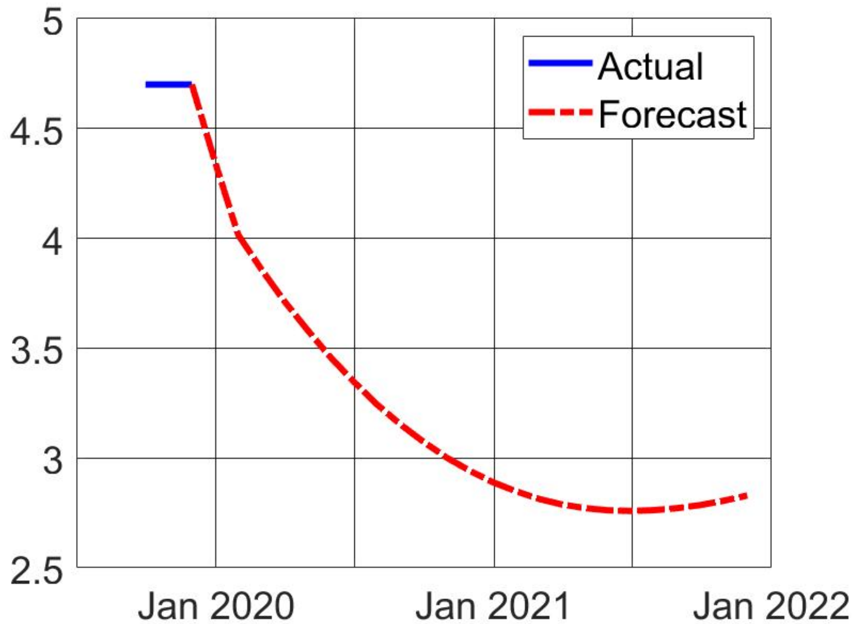

Figure 8: Monthly GDP growth rate

Hence, we use an alternative unemployment rate taken from monthly survey data and estimate backcasts of unemployment rate. It has been found the backcasts show that unemployment rate was rising from 2014 when the oil prices declined severely confirming the view that the economy entered a short period of recession in 2016. Overall, the unemployment rate implied by the fundamental factors lies 2 percentage points above the unemployment rate reported in official sources.

The main application of dynamic factor models in this paper is forecasting the macroeconomic performance indicators in Kazakhstan. Figure 8 represents the growth rate of GDP in the end of 2019 and forecasted GDP for upcoming 2021 and 2022. As we can see from the graph, monthly GDP growth rate seems to decrease significantly until the middle of 2021 and slightly increase in the beginning of 2022.

Measurement errors and alignments present in the data raise a concern for both researchers and policymakers. First, it is hard to conduct sound research based on the data that is subject to either measurement errors or alignments. Second, valid policy strategy cannot be designed in response to economic events that require an immediate solution. This paper gives alternative estimates on the macroeconomic performance indicators in Kazakhstan in order to shed light on the issues of official data quality.

8. Acknowledgements

Coauthor: Nurdaulet Abilov (NAC analytica), Nur-Sultan

I want to thank my advisors Professor Iraj Rahmani (Nazarbayev University) and Professor In Kim (Nazarbayev University) for their help. Also, I appreciate the support of my external advisor Charles Becker (Duke University)

References

N. Abilov, A. Tolepbergen, and K. Weyerstrass. A macroeconometric model for Kazakhstan. *Manuscript submitted for publication*, 2019.

M. Aiolfi, L. Catao, and A. Timmermann. Common factors in Latin America's business cycles. *IMF Working Ppaper 06/49*, 2006.

F. Altissimo, R. Cristadoro, M. Forni, M. Lippi, and G. Veronese. New Euro-COIN: tracking economic growth in real time. *The Review of Economics and Statistics*, 92, 2010.

S. Aruoba, F. Diebold, and S. Scotti. Real-time measurement of business conditions. *PDF Journal of Business and Economic Statistics*, 27, 2009.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

J. Bai and S. Ng. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74, 2006.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194, 2006.

M. Camacho and G. Perez-Quiros. Introducing the EuroSTING: short-term indicator of euro area growth. *Journal of Econometrics*, 25, 2010.

C. Doz, D. Giannone, and L. Reichlin. A quasi maximum likelihood approach for large approximate dynamic factor models. *ECB Working Paper No. 674*, 2006.

C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205, 2011.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized factor model: Identification and estimation. *Review of Economics and Statistics*, 82:540–554, 2000.

M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–839, 2005.

A. Galli. Which indicators matter? Analyzing the Swiss business cycle using a large-scale mixed-frequency dynamic factor model. *SNB Working Papers 8/2017*, 2017.

J. Geweke. *The Dynamic factor analysis of economic time series.* North-Holland, Amsterdam, 1977.

D. Giannone, L. Reichlin, and L. Sala. Monetary policy in real time. *NBER Macroeconomics Annual 2004*, 19:161–224, 2005.

D. Giannone, L. Reichlin, and L. Small. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55 (4):665–676, 2008.

T. Sargent and C. Sims. Business cycle modeling without pretending to have too much apriori economic theory. *Working Paper No.55, Federal Reserve Bank of Minneapolis*, 1977.

J. Stock and M. Watson. New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, page 351–393, 1989.

J. Stock and M. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, (97): 1167–1179, 2002.

J. Stock and M. Watson. *Dynamic factor models.* in the Oxford Handbook of Economic Forecasting, ed. by M.P. Clements, and D.F. Hendry, chap. 7, pp.35-39, Oxford University Press, Oxford, 2011.

Appendix A.

Variables	**Unit of measurement**	**Source**[1]
Short-term economic activity index	Index	Stat
Nominal income per capita	KZT	Stat
Real income	Index	Stat
The number of unemployed	Quantity	Stat
Unemployment rate	Percentage	Stat
Average monthly nominal salary per employee	KZT	Stat
Real wage index	Index	Stat
Consumer Prices Index	Index	Stat
Enterprise Price Index industrial manufacturers products	Index	Stat
Producer price index agricultural products	Index	Stat
Price index in construction	Index	Stat
Wholesale Price Index sales	Index	Stat
Tariff index for cargo transportation by all means of transport	Index	Stat
Service Tariff Index postal and courier for legal entities	Index	Stat
Service Tariff Index communications for legal entities	Index	Stat
Export Supply Price Index products	Index	Stat
Import Price Index product receipts	Index	Stat
Market prices housing	Index	Stat

[1] Stat - Comittee of Statistics of the Republic of Kazakhstan; NBRK - National Bank of the Republic of Kazakhstan; KASE - Kazakhstan Stock Exchange.

Volume of investment in fixed assets	Index	Stat
Volume of industrial products	Index	Stat
Gross output agricultural products (services), forestry and fisheries	Index	Stat
Volume of construction work	Index	Stat
Transportation of goods by all types of transport	Index	Stat
Cargo turnover of all types transport	Index	Stat
Scope of postal and courier services activities	Index	Stat
Scope of Services connection	Index	Stat
Loans in the economy STB (short-term)	KZT	NBRK
Loans in the economy STB (long-term)	KZT	NBRK
Loans in the economy (total)	KZT	NBRK
Loans to industry	KZT	NBRK
Rural loans in the economy	KZT	NBRK
Construction loans	KZT	NBRK
Credits to transport	KZT	NBRK
Loans to communications	KZT	NBRK
Loans to trade	KZT	NBRK
Loans to other industries	KZT	NBRK
Weighted average interest rate on loans	Percentage	NBRK
Ratio of overdue loans to total loans in the banking sector	Percentage	NBRK
Ratio of overdue loans over 90 days to total loans in the banking sector	Percentage	NBRK
Monetary aggregate M1	KZT	NBRK
Monetary aggregate M2	KZT	NBRK
Business lending rate	Percentage	NBRK
Household lending rate	Percentage	NBRK

Business lending rate	Percentage	NBRK
Household lending rate	Percentage	NBRK
MEOKAM (yields on government bonds with the maturity of 1 to 10 years)	Percentage	NBRK
MEUKAM (yields on government bonds with the maturity of from 5 years)	Percentage	NBRK
Yields on the short-term notes of the National Bank (central bank) notes (up to 1 year)	Percentage	NBRK
TONIA	Percentage	NBRK
Stock market index of KASE	Index	KASE
Price of oil (Brent)	US dollars	Bloomberg
Bloomberg commodity price index	Index	Bloomberg
Inflation rate in the US	Percentage	Bloomberg
Inflation rate in Russia	Percentage	Bloomberg
World Trade Index	Percentage	Bloomberg
Exchange rate of ruble	KZT	NBRK
Real index effective exchange rate	Percentage	NBRK
Exchange rate of US dollar	KZT	NBRK

Appendix B.

Variables	Factor 1	Factor 2
Short-term economic activity index	0.675	0.357
Nominal income per capita	0.479	-0.411
Real income	-0.065	0.036
The number of unemployed	-0.097	0.001
Unemployment rate	0.589	0.198
Average monthly nominal salary per employee	0.283	-0.001
Real wage index	0.035	0.743
Consumer Prices Index	0.855	0.315
Enterprise Price Index industrial manufacturers products	0.182	0.152
Producer price index agricultural products	0.742	0.077
Price index in construction	0.341	0.631
Wholesale Price Index sales	-0.017	0.649
Tariff index for cargo transportation by all means of transport	0.036	0.476
Service Tariff Index postal and courier for legal entities	0.112	-0.187
Service Tariff Index communications for legal entities	0.931	-0.109
Export Supply Price Index products	0.583	0.215
Import Price Index product receipts	-0.254	0.762
Market prices housing	0.468	-0.257
Volume of investment in fixed assets	0.245	-0.448
Volume of industrial products	0.512	-0.607

Gross output agricultural products (services), forestry and fisheries	-0.050	-0.104
Volume of construction work	0.142	-0.198
Transportation of goods by all types of transport	0.632	-0.158
Cargo turnover of all types transport	0.554	-0.376
Scope of postal and courier services activities	0.402	-0.487
Scope of Services connection	0.697	-0.156
Loans in the economy STB (short-term)	0.206	0.270
Loans in the economy STB (long-term)	0.190	0.602
Loans in the economy (total)	0.253	0.517
Loans to industry	0.248	0.419
Rural loans in the economy	-0.317	0.486
Construction loans	0.587	0.507
Credits to transport	0.264	0.528
Loans to communications	-0.113	0.006
Loans to trade	-0.043	0.234
Loans to other industries	0.387	0.058
Weighted average interest rate on loans	-0.297	-0.085
Ratio of overdue loans to total loans in the banking sector	0.184	0.121
Ratio of overdue loans over 90 days to total loans in the banking sector	0.287	0.183
Monetary aggregate M1	0.696	0.102
Monetary aggregate M2	0.694	0.331
Business lending rate	-0.177	-0.023
Household lending rate	-0.004	0.118
Business lending rate	-0.005	0.003
Household lending rate	-0.002	-0.007

MEOKAM (yields on government bonds with the maturity of 1 to 10 years)	0.034	-0.232
MEUKAM (yields on government bonds with the maturity of from 5 years)	-0.302	-0.113
Yields on the short-term notes of the National Bank (central bank) notes (up to 1 year)	-0.231	-0.029
TONIA	-0.085	-0.078
Stock market index of KASE	0.242	0.026
Price of oil (Brent)	0.866	-0.113
Bloomberg commodity price index	0.789	-0.022
Inflation rate in the US	0.913	-0.016
Inflation rate in Russia	-0.677	0.028
World Trade Index	0.420	-0.407
Exchange rate of ruble	0.292	0.732
Real index effective exchange rate	0.106	-0.866
Exchange rate of US dollar	-0.466	0.742