Nowcasting Real GDP growth in South Africa

Alain Kabundi, Elmarie Nel and Franz Ruch

Abstract

This paper uses the nowcasting approach to forecast, in real-time, real GDP growth in South Africa from 2010Q1 to 2014Q3. We exploit the flow of information underlying the state of the economy which is available at a higher frequency to estimate GDP growth before the end of each quarter. One of the major challenges that face forecasters, policymakers, and economic agents is having a clear view of the state of the economy in real time. Unlike weather forecasts, many economic variables are only available at low frequency and with generous lags, making it difficult to have information on the state of the economy even after the end of the quarter. The pseudo out-of-sample forecasts show that the performance of the nowcasting model is comparable to those of professional forecasters even though the latter enhance the forecasting accuracy with judgement. The nowcast model also outperforms all other benchmark models by a significant margin.

Keywords: Nowcasting, Factor Model, Bayesian VAR, Forecasting

JEL classification: E52, C53, C33.

Preliminary and incomplete. Please do not quote.

1. Introduction

One of the major challenges that face forecasters, policymakers, and economic agents is having a clear view of the state of the economy in real-time. Unlike weather forecasts, real Gross Domestic Product (GDP) is only available at low frequency and with a generous lag, making it difficult to have information on the state of the economy even after the particular quarter has passed. For example, real GDP in the fourth quarter in South Africa (SA) becomes available eight weeks after the end of the quarter. This conundrum complicates the work of policymakers and forecasters who require realtime information on GDP in order to make accurate policy decisions and forecasts. Significantly, a substantial amount of higher frequency economic information released between the start of the quarter and the release of the official GDP number can help to predict the current state of the economy, the near future, and near past. This is done using Nowcasting.

Nowcasting, as introduced by Giannone, Reichlin, and Small (2008), provides a framework for the assimilation of a large number of timely economic series that matter for economic growth in a mixed frequency and asynchronise environment. The prediction of the current state of the economy is known as "nowcast" and the prediction of the near past is called "backcast". Statistical models such as a nowcasting model outperforms a naïve constant growth model in forecasting GDP growth at very short horizons, i.e. the current state and the near future, while are beaten by the latter model in predicting the medium- to long-term (see Giannone, Reichlin, and Small, 2008; Banbura, Giannone, and Reichlin, 2011; Banbura, Giannone, Modugno, and Reichlin, 2012).

So far there has been no attempt at estimating a nowcast model in Africa, and particularly for South Africa. This paper bridges the gap in the literature by estimating a nowcast model of real GDP growth in SA. It uses 21 series covering real economy, nominal variables, and the financial sector. Like most nowcasting models, the dataset contains variables published in different frequencies: daily, monthly, and quarterly. In addition, information flows are asynchronous resulting in missing data toward the end of the sample. The sample period is from 1990 to 2014, while the pseudo out-of-sample period is from 2010Q1 to 2014Q3. The forecasting exercise is performed on the 15th of each month.

The results show that the nowcasting model performs comparably to consensus forecasts of quarterly growth by Reuters and Bloomberg. This is despite the disadvantage of this approach being atheoretic and unable to incorporate judgement as is the case in the Reuters and Bloomberg survey forecasts. A pertinent example in South Africa occurred in the first quarter of 2014 when a prolonged strike in the platinum sector resulted in a contraction in real GDP growth, something that survey participants could adjust for but wasn't captured in the data. If we exclude this point the nowcast model outperforms survey results. The nowcast model also outperforms all other benchmark models by a significant margin.

The rest of the paper is organised as follows. Section 2 describes the literature and methodological issues around nowcasting. Section 3 describes the nowcast model. We discuss the data used in Section 4. In addition, we discuss the approach used to select the factors and their identification. Section 5 discusses the results of pseudo out-of-sample forecasting. Section 6 provides some concluding remarks.

2. Literature review

Giannone, Reichlin, and Small (2008) propose an approach that uses factor analysis in a data-rich environment together with the Kalman filter as a comprehensive and powerful technique for forecasting the near past, the current state, and the near future of GDP growth rate in the US. The model performs as good as professional forecasting and outperforms all other competitive models. There are important findings from their analysis. First, the nowcast model outperforms the naïve constant growth over the short-run, especially for the current quarter. Second, the proposed model performs equally well as the professional forecasts, despite the fact that the latter includes judgement. Third, the performance of the nowcast model improves as new information becomes available and towards the end of the quarter. This means that the forecaster can incorporate information progressively upon their release and increase our understanding of the drivers of GDP growth. Finally, it is not necessarily the strength of the relationship between each variable and the target which matters, but most importantly the timeliness of each variable. For example, when PMI is released at the beginning of each month, there is no hard data available. In addition, the Bureau of Economic Research (BER) confidence indices of the corresponding quarter are published 25 and 12 days, respectively, before the end of the quarter. It implies that soft data are extremely important (see Kabundi, 2004 and Martinsen et al., 2014).

Since the seminal work by Giannone, Reichlin, and Small (2008), there has been an increase in popularity of these models in predicting different macroeconomic variables and for different countries. ECB (2008) uses it to forecast GDP growth in the euro area, while Matheson (2011) for tracking growth in 32 advanced and emerging-market economies. Liebermann (2012) provides a detailed study on the nowcasting of a large variety of key monthly macroeconomic releases. Modugno (2013) uses the same framework for nowcasting inflation. Instead of using key determinants of GDP, Angelini, Banbura, and Runstler (2010) forecast different components of GDP, and then aggregate them to obtain a nowcast of GDP. There are also now several applications in different countries. Kuzin, Marcellino, and Schumacher (2013) use model combination in nowcasting the German GDP, Matheson (2010) nowcast New Zealand GDP and CPI, Yiu and Chow (2011) and Giannone, Agrippino, and Modugno (2013) use it for China, Aastveit and Trovik (2008) build a nowcast model for Norway, Siliverstovs and Kholodilin (2010) apply the same approach for Switzerland, D'Agostino, McQuinn, and O'Brien (2008) for Ireland, Barhoumi, Darn and Ferrara (2010) nowcast the French GDP growth, de Winter (2011) shows the superiority of this technique for the Netherlands, Arnostova, Havrlant, Ruzicka and Toth (2011) confirms the findings by other scholars for the Czech Republic, and for Brazil Bragoli, Metelli, and Modugno (2014) show that nowcasting performs as well as predictions of professional forecasters.

Nowcasting introduces two challenges that a forecaster must overcome. First, given that data is released in different frequencies, the forecaster needs to reconcile the data into a single frequency. For example, manufacturing production, which is closely related to GDP, is published monthly. The problem of mixed-frequency is easily translated to a task of missing data. Traditionally, forecasters use bridge equations as means of reconciling mixed-frequency information. However, such a solution is suboptimal, as bridge equations are single regression models that can only accommodate few variables. Thus it is crucial to use a framework which can take into account a large number of monthly series which are able to predict GDP. Forni, Hallin, Lippi, and Reichlin (2003) and Stock and Watson (2003) proved that factor models are good candidates for turning the curse of dimensionality (large number of parameters relative to the number of observations) into a blessing of dimensionality. It means that forecasters benefit from rich information contained in a large panel of variables.

Second, data are not released in a synchronous fashion. For example, the Purchasing Manager's Index (PMI) is released at the beginning of each month, while trade variables are published with a month lag. The PMI for February is published at the beginning of March, while the exports and imports for February become available at the end of March. Similar to the mixed-frequency problem, the unbalanced panel at the end of the period is also looked at as a missing-data question. One way of solving this problem is to move down all series in order to have a balanced panel at the end of the period. The drawback of such approach is one does not take into account the true contemporaneous correlation that exists among variables at each period introducing a lead-lag relationship between variables. Evans (2005) and Giannone, Reichlin, and Small (2008) propose the use of the Kalman filter as a solution to this problem, given its ability to adapt to changing data availability.

3. The Model

The added advantage from nowcasting is its ability to use high frequency data (daily and monthly in our case) to estimate quarterly macroeconomic variables. Thus we estimate GDP in quarter *q,* $\hat{z}^{q}_{k|\vartheta_j}$, by using the information available during that quarter, denominated as Ω_{ϑ_j}. We estimate

$$\hat{z}^{q}_{k|\vartheta_j} = Proj\left[\hat{z}^{q}_{k}|\Omega_{\vartheta_j}\right]$$

As more information is released, this is integrated into the projection. Thus the forecast is made with the largest possible information set at each point in time such that

$$\Omega_{\vartheta_j} > \Omega_{\vartheta_{j-1}} > \Omega_{\vartheta_{j-2}} > \cdots$$

We estimate GDP using the dynamic factor model suggested by Doz, Gainnone and Reichlin (2005). Since the frequency of the dataset differs, introducing missing data points, and the common factors are unobserved, we need to use the Kalman filter. The Kalman filter is a discrete, recursive linear filter used to estimate unobservable variables in a system of equations given an information set (Pasricha, 2006). The Kalman filter is optimal; minimises the mean squared error estimator, if the observed variable and error are jointly Gaussian and is “best” in the class linear filters if this assumption is violated.

The factor, F_t, summarises the co-movements from the latest available information set. Thus we estimate the corresponding monthly GDP ($y_{t|\vartheta_j}$) to the quarterly series $z^q_{k|\vartheta_j}$ by

$$y_{t|\vartheta_j} = \mu + \Lambda F_t + e_{t|\vartheta_j}$$

where μ is a constant, $e_{t|\vartheta_j}$ is a white noise error, and F_t is the unobserved factor.[1]

As there can be more than one significant co-movement in the information set, there can be more than one factor needed in the model. We test the number of common factors using the modified Bai and Ng (2002) information criterion.[2] To determine the number of common factors, this method determines the factor which minimises the variance of the idiosyncratic component.[3]

Near-term forecasting rarely provides any insight to the marginal change in GDP. This is not the case with nowcasting. This model makes it possible to analyse the marginal impact of a specific variable on the GDP forecast, making it possible to analyse the drivers of the forecast. This ability is called the *News* of the model.

We estimate the *News* by

$$NEWS[z^q_k, \vartheta_j] = \hat{z}^q_{k|\vartheta_j} - \hat{z}^q_{k|\vartheta_{j-1}}$$

which is the difference between the forecast of GDP including ($\hat{z}^q_{k|\vartheta_j}$) and excluding ($\hat{z}^q_{k|\vartheta_{j-1}}$) the most recent datapoint of various variables used. Thus, indicating the impact of certain datapoints on GDP. This is a valuable function as the impact of recent releases can pinpoint latest developments in the economy and increase our understanding of the current state of the economy.

4. Data

The dataset contains 21 series covering real variables, nominal variables, and financial variables.[4] Giannone, Reichlin, and Small (2008) show that if variables are selected systematically including timely information concerning the current state of the economy, the cross-section need not to be very large. The real sector includes trade, production, and demand variables. The nominal variables are producer and consumer price indices, and the Brent crude oil price (in US dollar).

[1] We specify the common factors as $\mathrm{F_t = MF_{t-1} + Nu_t}$ where M is a $\mathrm{r \times r}$ matrix, N is a $\mathrm{r \times q}$ matrix of full rank r and $\mathrm{u_t}$ the common factor shocks.
[2] As proposed Alessi, Barigozzi and Capasso (2010).
[3] For more detail on the test specifications see Alessi, Barigozzi and Capasso (2010).
[4] Table 1 provides a complete list of variables. In addition, it shows the treatment, the source, the frequency, and the relevance of all variables.

Financial variables are the nominal effective exchange rate and the policy rate (repurchase rate). Most series are expressed in log differenced at the monthly frequency.

Quarterly series used are the capacity utilisation, consumer confidence, and real GDP growth. These series are transformed into monthly frequency with quarterly values set as third month observations and missing data for the remaining two months of the quarter. Then, we use the Kalman filter to handle the missing data issue. The nominal effective exchange rate, the only variable obtained at daily frequency, is transformed to a monthly frequency by taking average of 15 days for each month. All variables are transformed to induce stationarity. For professional forecasters, we use the forecasts based on the surveys conducted by Reuters published monthly, while the Bloomberg forecasts are released two days before the GDP release.[5]

The first task in nowcasting is the choice of variables to include. We use the view of market participants obtained from Bloomberg to decide on variables to include. The sixth column of Table 1 depicts the relevance of variables. Note that the policy rate matters more with a weight of 97.30, followed by trade balance with 94.59. Even though PMI seems less important than trade balance, it is timely, with only one day lag, while the latter is published with 31 days lag. Expect for PMI and confidence indices, all variables are expressed as quarterly growth rate. In addition, we use annual growth for vehicles sales.

Table 2 shows correlation coefficients between different variables. All variables are standardised to have the same unit of measurement. Most of the correlation coefficients are above 0.5, which is evidence that they move together. Importantly, they are highly correlated with GDP growth. PMI is the variable that mimics GDP growth quite closely. Even though, it is perceived as soft data, it is closely related to hard data and it is timely. Figure 1 shows how close PMI tracks GDP. It is incredible how PMI predicts turning points. It is mostly contemporaneous with GDP growth. It depicts a correlation coefficient of 0.8. This relatively high correlation is confirmed in Figure 1. In addition, Figure 1 confirms the observation in Table 2 about relatively strong correlation among real variables. It depicts evidence of strong co-movement with GDP growth which implies that they can serve as its determinants. Instead of choosing one variable over another, the model has the advantage of exploiting the data-rich information represented in Table 1. Nevertheless, traditional econometric models cannot handle using them simultaneously in one model due to overfitting. Factor analysis can.

The rationale behind this is that the dynamics in all these variables can easily be captured by few common factors. In so doing, the model mimics closely information taken into account by policymakers, who do not base their decision from observing a single variable. It is clear in Figure 2 that Factor 1 tracks the GDP growth quite well. Moreover, the correlation between GDP growth and Factor 1 is 0.80, which far above most correlations in Table 2, except PMI.

Factor 1 represents the real sector, while Factor 2 is closely linked with M3 money supply and the credit to private sector. We can then aggregate the two factors into a single index, the common component, as depicted in Figure 4, as a proxy of monthly GDP growth rather than

[5] The mean of the Reuters survey is used.

using the growth in manufacturing production. The constructed series follows GDP growth much better than the Factor. The nowcasting model allows us to update our nowcast of GDP growth with each new release. In addition, it is possible to evaluate the contribution of each variable in predicting GDP growth.

5. Empirical Results

This paper follows the approach proposed by Giannone, Reichlin, and Small (2008) and Banbura, Giannone, and Reichlin (2011) to nowcast GDP growth in SA using monthly and daily information from January 1990 to November 2014. We perform a pseudo real-time out-of-sample forecasting from the first quarter of 2010 to third quarter of 2014. We conduct the forecasting on the 15th of each month. The first task in nowcasting is to determine the number of factors to include. We use the Alessi, Barigozzi, and Capasso (2010) approach, an improvement of the most popular Bai and Ng (2002) methodology. We use one lag in the state equation for the estimation of factors. We compare our nowcasting model to both survey results as well as some benchmark models.

5.1. Determining the number of common factors and shocks

In order to determine the number of common factors to use in the model we implement a modified Bai and Ng (2002) information criterion as implemented by Alessi, Barigozzi and Capasso (2010). This method chooses the number of factors by minimising the variance of the idiosyncratic component of the approximate factor model. This is subject to a penalisation in order to avoid over-parameterisation. The information criterion is

$$\hat{r}_{c,N}^{T} = \underset{0 \leq k \leq r_{max}}{argmin} IC_{a,N}^{T*}(k)$$

where

$$IC_{a,N}^{T*}(k) = \log\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}\left(x_{it} - \hat{\lambda}_i^{(k)`}\hat{F}_t^{(k)}\right)^2\right] + ckp_a(N,T) \text{ for a=1,2}^{6}$$

For k common factors, N is the number of variables, T the number of observations, $\left(x_{it} - \hat{\lambda}_i^{(k)`}\hat{F}_t^{(k)}\right)$ the idiosyncratic error, c an arbitrary positive real number and $p_a(N,T)$ the penalty function. Alessi, Barigozzi and Capasso (2010) propose multiplying the penalty function by *c* since Hallin and Liska (2007) show that a penalty function, $p(N,T)$ leads to consistent estimation of *r*, the number of factors, if and only if $cp(N,T)$ does as well.

The only information available regarding the behaviour of $\hat{r}_{c,N}^{T}$ can be gleaned from analysing subsamples of sizes (n_j, t_j). For any *j*, we can compute $\hat{r}_{c,n_j}^{t_j}$ which is a monotonic non-increasing function in *c*. Therefore, there exist moderate values of *c* such that $\hat{r}_{c,N}^{T}$ converges

[6] See Alessi, Barigozzi and Capasso (2010) for the functional form of the penalty function.

from above to *r*. This result, however, needs to be independent of *j* for the criterion to be stable. This is measured by the variance of $\hat{r}_{c,n_j}^{t_j}$ as a function of *j*:

$$S_c = \frac{1}{J}\sum_{j=1}^{J}\left[\hat{r}_{c,n_j}^{t_j} - \frac{1}{J}\sum_{j=1}^{J}\hat{r}_{c,n_j}^{t_j}\right]^2$$

Figure 5 shows the estimated number of factors for our model. The vertical axis represents the number of factors while the horizontal axis represents an arbitrary positive real number *c*. We run the results over a number of sizes for the subsamples in order to get a robust result. In order to determine the number of factors we have to find the first value of $\hat{r}_{c,N}^{T}$ where S_c is zero. The results suggest that the number of factors should be two.

The other important choice that's required in the model is the number of common shocks included. As we only use two common factors, we can easily assume that one common shock will be sufficient. This is based on arguments made by Forni et al. (2005) and Bai and Ng (2007) that q is lower than k. This argument follows as economic fluctuations are driven by a small number of common shocks. However we test the number of common shocks using the information criterion specified by Onatski (2009). We test the null hypothesis of $q = q_0$ shocks versus the alternative hypothesis of $q_0<q\leq q_1$ shocks. The test (table 3) supports the arguments made by Forni et al. (2005) and Bai and Ng (2007), determining a common shock of one.

5.2. Forecasting results

We compare our nowcasting model to both survey results as well as real-time benchmark models. These models include:

- a random walk model (labelled RW);
- two autoregressive models with one and four lags (labelled AR1 and AR4);
- and two four variable vector autoregressive models with one and four lags (labelled VAR1 and VAR4). These models include consumer price inflation, the repurchase rate and the nominal effective exchange rate.

Contrary to most nowcasting models in the literature which include manufacturing production, we exclude this variable as it affects the forecasting bias. Wholesale trade and trade activity index portray similar effects in predicting GDP growth in South Africa. Figure 6 illustrates the negative impacts of the wholesale sale trade and the trade activity index rendering the forecast biased upward. By including these variables the model performs well in predicting the peaks, whereas it fails short in predicting the troughs. Hence, we decide to remove them from the analysis.

Note from Figure 6 that the GDP growth recorded in the out-of-sample period is less stable. Figure 7 represents the results based on nowcasting and the results of the professional forecasters based on the surveys conducted by Reuters. The results show an upward biased in forecasting by professional forecasters. In general they tend to predict relatively well, but they miss the troughs. Nevertheless it is worth mentioning the difficulty faced by forecasters in predicting turning points.

Table 4 compares the out-of-sample Root Mean Squared Forecast Errors (RMSFEs) of the nowcast model, the two surveys, as well as the benchmark models. The results show that both survey outcomes do better than the nowcast model with the lowest RMSFE of 0.36 by the Bloomberg survey. The Reuters survey results are marginally better than the nowcast model with a RMSFE of 0.952 versus 0.97. Importantly, however, this result is driven by one outlier in 2014Q1. If we exclude this datapoint, the nowcast model outperforms the Reuters survey. This is despite the fact that these surveys incorporate judgement, a luxury the nowcast model does not have. The nowcast model also outperforms all benchmark models (Figure 8).

6. Conclusions

In the process of economic forecasting much time is spent on the initial conditions, i.e. the starting point of the forecast. Those who forecast well generally have good initial conditions while those who perform poorly have bad ones. The principal victim of these initial conditions is real gross domestic product (GDP) growth; the single most relevant variable describing the path of the economy (whether justified or not) and used, in concert with inflation, to justify the direction of the monetary policy. Of course, real GDP data is released with a generous lag and in South Africa this is 6-8 weeks after the end of the relevant quarter. Significantly, however a substantial amount of economic information released between the start of the quarter and the release of the official GDP number can help to predict the likely outcome. Nowcasting can bridge this divide.

Nowcasting provides a framework for the assimilation of a large number of timely economic series that matter for economic growth. The model uses a dynamic factor model to summarise this dataset providing a "nowcast" of current quarter growth. This framework provides a mechanism to determine the "marginal impact of new data releases" on real GDP growth. The benefit of this marginal impact or "news" means that a forecaster can quantify the evolution on economic activity in real-time and improve her initial conditions. The forecaster can also provide policymakers with the likely meaning of a large number of economic series on the path of real economic activity.

This paper estimates a nowcast model of real GDP growth in South Africa and shows that this model performs comparatively well to consensus forecasts of quarterly growth by Reuters. This is despite the disadvantage of this approach being a theoretic and unable to incorporate judgement as is the case in the Reuters survey. A pertinent example in South Africa occurred in the first quarter of 2014 when a prolonged strike in the platinum sector resulted in a contraction in real GDP growth, something that survey participants could adjust for but wasn't captured in the data. Encouragingly, if you remove this data point, the nowcast model outperforms the survey. The nowcast model also outperforms all other benchmark models by a significant margin. Given that a relatively short period of real-time data was used caution should be used when interpreting these results. Further monitoring is required.

References

Alessi, L., Barigozzi, M. and Capasso, M. (2010): "Improved penalization for determining the number of factors in approximate factor models," *Statistics & Probability Letters*, 80(23-24): 1806-1813.

Aastveit, K., and T. Trovik (2012): "Nowcasting norwegian GDP: the role of asset prices in a small open economy," *Empirical Economics*, 42(1), 95–119.

Angelini, E., M. Banbura, and G. Runstler (2010): "Estimating and forecasting the euro area monthly national accounts from a dynamic factor model," *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2010(1), 7.

Arnostova, K., D. Havrlant, L. Ruzicka, and P. Toth (2011): "Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators," *Czech Journal of Economics and Finance*, 61(6), 566–583.

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

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

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2012): "Now-Casting and the Real-Time Data-Flow", in G. Elliott and A. Timmermann, eds., Handbook of Economic Forecasting, Volume 2, Elsevier-North Holland, forthcoming.

Barhoumi, K., O. Darn, and L. Ferrara (2010): "Are disaggregate data useful for factor analysis in forecasting French GDP?," *Journal of Forecasting*, 29(1-2), 132–144.

Bragoli, D., Metelli, L., and Modugno, M. (2014): "The Importance of Updating: Evidence from a Brazilian Nowcasting Model," Discussion Series 2014-94, Board of Governors of the Federal Reserve System

D'Agostino, A., K. McQuinn, and D. O'Brien (2008): "Now-casting Irish GDP," Research Technical Papers 9/RT/08, Central Bank & Financial Services Authority of Ireland (CBFSAI).

de Winter, J. (2011): "Forecasting GDP growth in times of crisis: private sector forecasts versus statistical models," DNB Working Papers 320, Netherlands Central Bank, Research Department.

Diron,

ECB (2008): "Short-term forecasts of economic activity in the euro area," in *Monthly Bulletin, April*, pp. 69–74. European Central Bank.

Evans, M. D. D. (2005): "Where Are We Now? Real-Time Estimates of the Macroeconomy," *International Journal of Central Banking*, 1(2).

Forni, M., Hallin, M., Lippi, M. and Reichli, L. (2001): "Coincident and Leading Indicators for the Euro Area, " *Economic Journal, Royal Economic Society*, 111(471): 62–85.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003): "Do financial variables help forecasting inflation and real activity in the euro area?," *Journal of Monetary Economics*, 50(6), 1243–1255.

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

Giannone, D., Agrippino, S.M. and Modugno, M. (2013): "Nowcasting China Real GDP" Mimeo

Hallin, M. and Liska, R. (2007): Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association,* 102(478), pp.603–617.

Kabundi, A. (2004): "Estimation of Economic Growth in France Using Business Survey Data", IMF Working Papers 04/69, International Monetary Fund.

Kuzin, V., M. Marcellino, and C. Schumacher (2013): "Pooling versus Model Selection for Nowcasting with Many Predictors: An Application to German GDP," *Journal of Applied Econometrics*, 28(3): 392-411.
Martinsen, K., F. Ravazzolo , F. Wulfsberg (2014): "Forecasting Macroeconomic Variables using Disaggregate Survey Data," *International Journal of Forecasting*, 30(1): 65-77.
Matheson, T. (2011): "New Indicators for Tracking Growth in Real Time," IMF Working Papers 11/43, International Monetary Fund.
Matheson, T. D. (2010): "An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys," *Economic Modelling*, 27(1), 304–314.
Modugno, M. (2011): "Nowcasting inflation using high frequency data," *International Journal of Forecasting*, 29(4): 664-675.
Stock, J. H., and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.
Yiu, M. S., and K. K. Chow (2010): "Nowcasting Chinese GDP: Information Content of Economic and Financial Data," *China Economic Journal*, 3(3), 223–240.

Table 1: List of Variables

Variables	Sources	Frequency	Lag	Relevance	Treatment
Kagiso Purchasing Manager Index	BER	M	1	70.27	1
Total Retail Trade Sales	STATS SA	M	44	64.86	5
Real Wholesale Trade Sales	STATS SA	M	46	0.00	5
Mining Production	STATS SA	M	44	27.73	5
Private Credit	SARB	M	29	51.35	1
Electricity Consumption	STATS SA	M	39	10.81	5
Motor Vehicle Sales	NAAMSA	M	2	32.43	5
SACCI Business Confidence	SACCI	M	3	54.05	5
Exports	SARS	M	31	94.59	5
Imports	SARS	M	31	94.59	5
Money Supply M3	SARB	M	29	78.38	5
Oil price - US dollar (Brent crude)	OECD	M	0		5
Trade Activity Index	SACCI	M	12		1
Trade Expectations Index	SACCI	M	12		1
Gold Production	STATS SA	M	44	50.00	5
Business Cycle Indicator	SARB	M	51	24.32	5
Consumer Price Index	STATS SA	M	20	72.97	5
Producer Price Index	STATS SA	M	30	78.13	5
Repo Rate	SARB	M	0	97.30	1
Nominal Effective Exchange Rate	SARB	D	0		5
BER Consumer Confidence	BER	Q	-12	67.57	1
Capacity Utilisation	STATS SA	Q	67	0.00	1
Gross Domestic Product	STATS SA	Q	57	62.16	5

Note: the treatment 1 refers to variables at level and 5 refers to log difference

Table 2: Correlation Coefficients

	GDP	PMI	Retail	Electricity	Vehicle	BCI	TEI
GDP	1						
PMI	0.80	1					
Retail	0.67	0.70	1				
Electricity	0.66	0.67	0.50	1			
Vehicle	0.58	0.62	0.59	0.50	1		
BCI	0.64	0.66	0.64	0.63	0.70	1	
TEI	0.67	0.76	0.65	0.71	0.68	0.65	1

Table 3: Hypothesis testing for the number of common shocks

Null hypothesis	P-value	Decision	Alternative
H_0: q=1	0.327	Do not Reject null	H_1: $1 < q \leq 2$
H_0: q=2	0.571	Do not Reject null	H_1: $2 < q \leq 3$
H_0: q=3	0.820	Do not Reject null	H_1: $3 < q \leq 4$

Table 4: Out-of-sample Root Mean Squared Forecast Errors

		Nowcast	Bloomberg	Reuters	AR(1)	AR(4)	VAR(1)	VAR(4)	RW
2010	Q1	0.515	0.289	1.252	1.614	2.227	0.420	1.508	4.589
	Q2	0.532	0.541	0.959	0.717	0.473	1.941	1.540	3.209
	Q3	1.051	0.656	0.559	0.403	0.149	1.542	2.550	2.594
	Q4	0.682	0.245	1.145	1.866	1.982	0.746	0.513	4.445
2011	Q1	1.202	0.695	1.445	0.970	0.714	0.004	0.717	4.845
	Q2	1.768	0.286	2.176	2.898	2.917	3.645	4.103	1.264
	Q3	0.937	0.352	1.252	0.207	0.128	1.025	2.326	1.448
	Q4	0.726	0.065	0.448	1.376	1.186	0.769	0.441	3.165
2012	Q1	0.447	0.494	0.008	0.237	0.638	0.591	0.431	2.744
	Q2	1.063	0.106	0.375	0.502	0.635	0.460	0.168	3.194
	Q3	0.663	0.269	1.261	1.772	1.756	2.371	1.617	1.231
	Q4	0.273	0.431	0.392	0.486	0.609	0.385	1.464	2.131
2013	Q1	1.029	0.709	1.641	1.381	1.490	1.580	2.168	0.891
	Q2	0.423	0.212	0.158	1.639	1.536	1.017	0.683	3.038
	Q3	1.431	0.258	1.535	2.150	2.292	2.608	1.950	0.742
	Q4	1.649	0.434	1.121	2.529	2.535	2.170	1.753	3.834
2014	Q1	2.860	0.422	1.729	4.044	3.984	4.545	4.468	0.622
	Q2	0.780	0.292	0.562	0.222	0.076	0.007	0.403	0.608
	Q3	0.408	0.051	0.076	0.222	0.633	0.365	1.205	1.449
RMSFE (excl. 2014Q1)		**0.866**	**0.355**	**0.909**	**1.177**	**1.221**	**1.203**	**1.419**	**2.523**
RMSFE		**0.970**	**0.358**	**0.952**	**1.328**	**1.366**	**1.379**	**1.579**	**2.423**
Relative to RW		**0.400**	**0.148**	**0.393**	**0.548**	**0.564**	**0.569**	**0.652**	**1.000**

Figure 1: Factor 1 and Real Variables

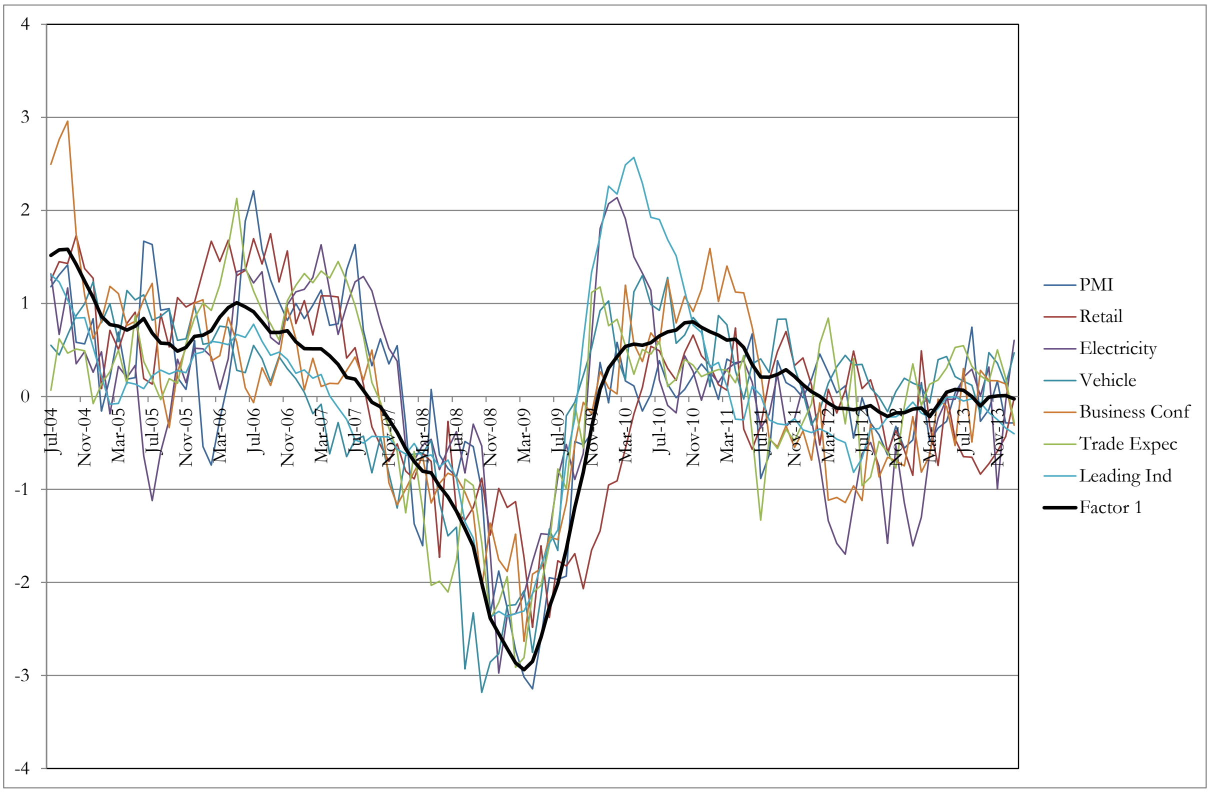

Figure 2: Factor 1 and GDP growth

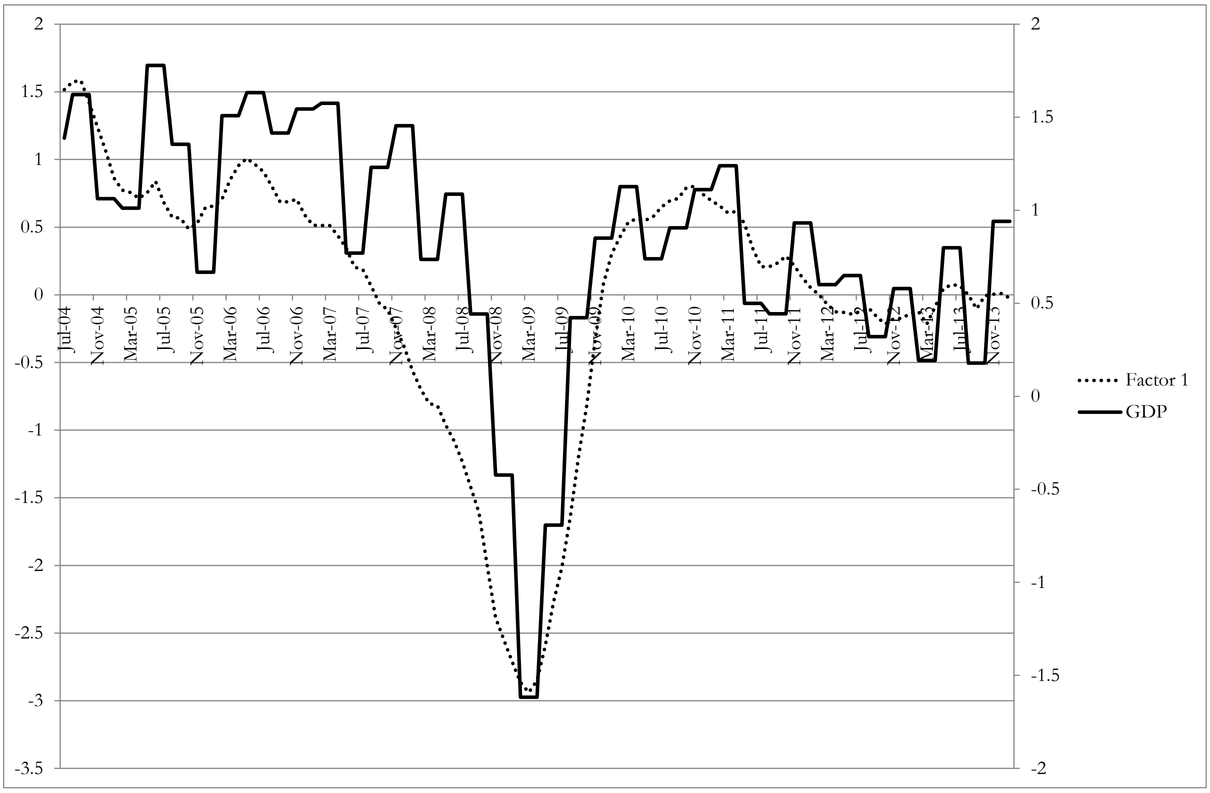

Figure 3: Factor 2, M3, and Credit to Private Sector

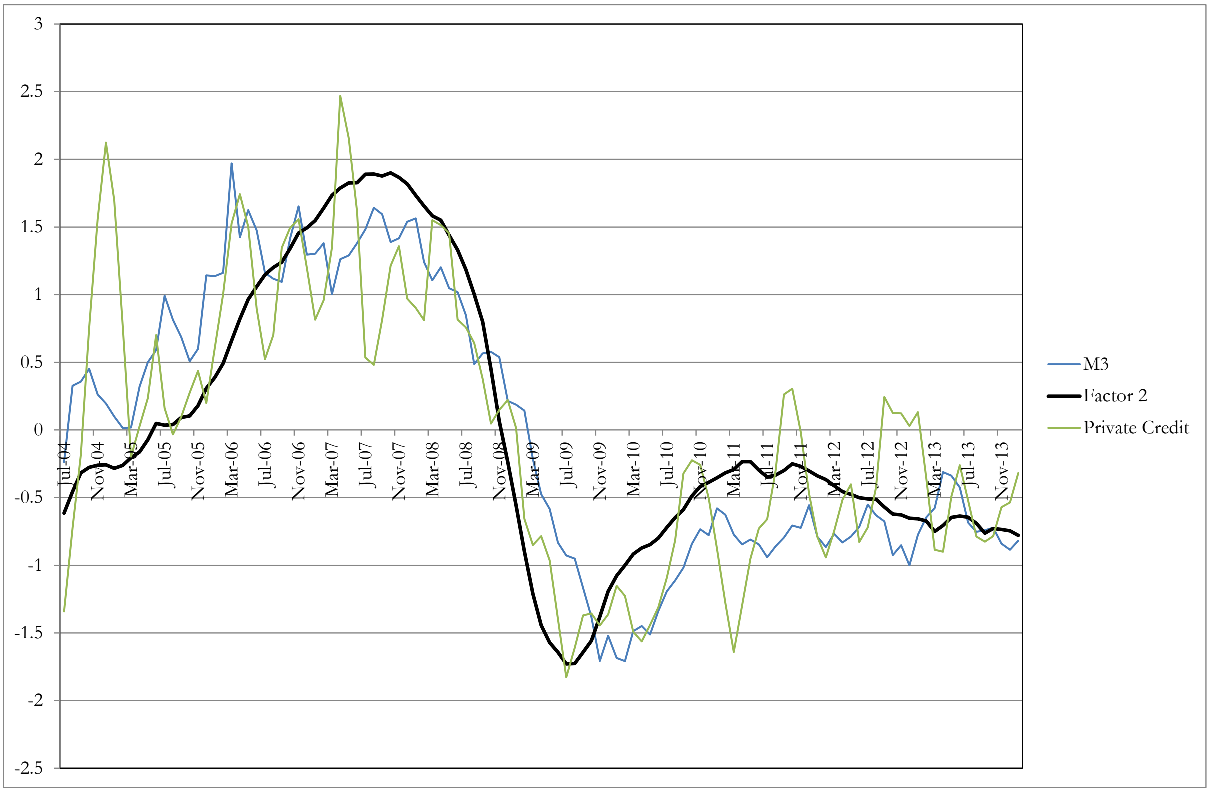

Figure 4: Common component and GDP growth

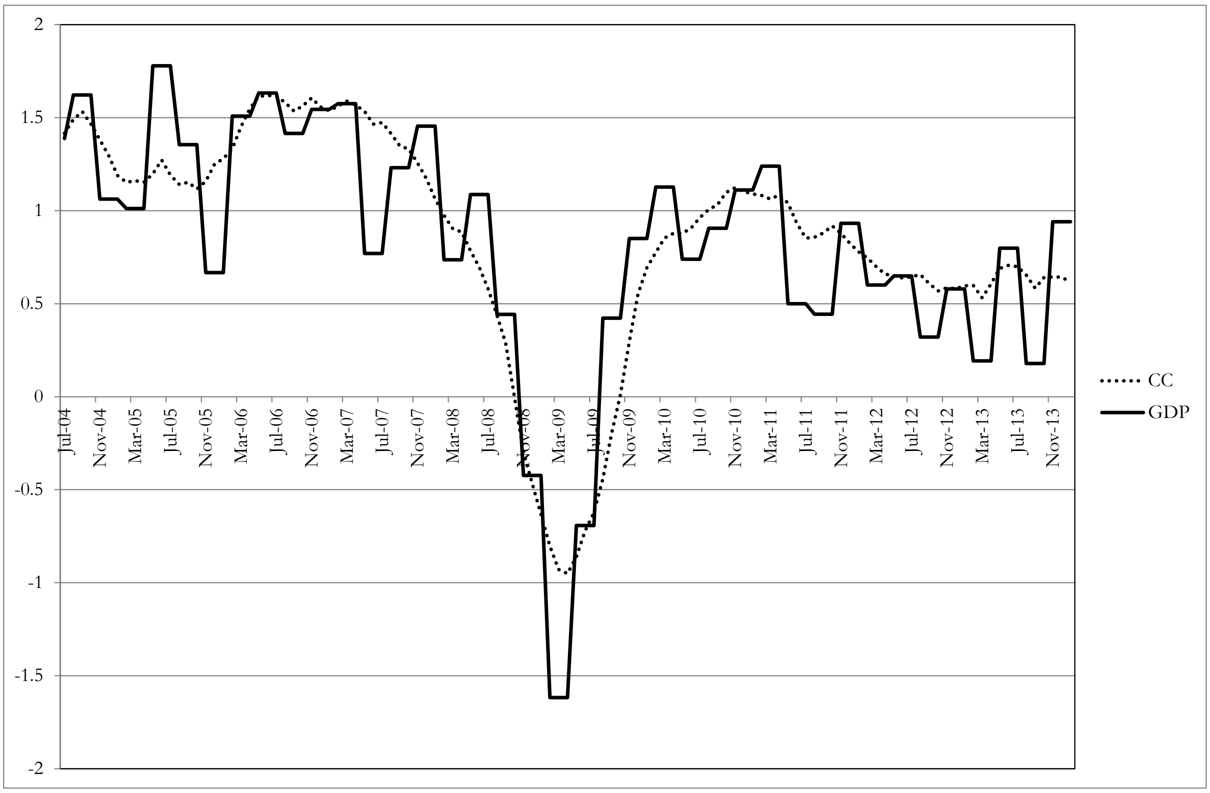

Figure 5: ABC criterion for common factors.

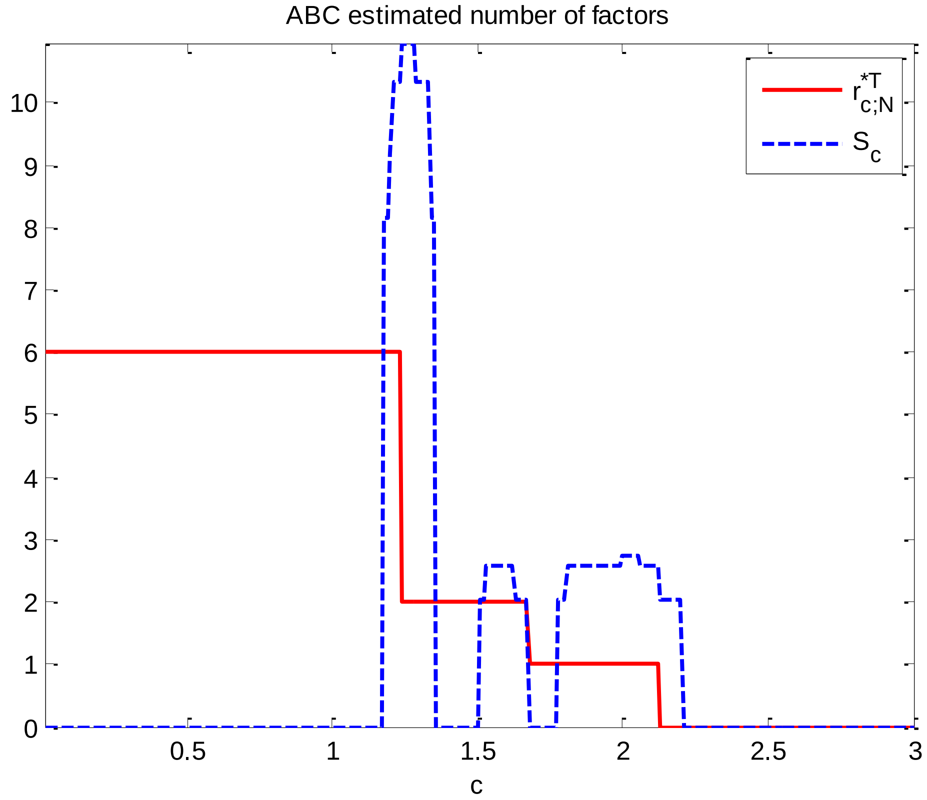

Figure 6: Model with and without Wholesale Trade and Trade Activity Index

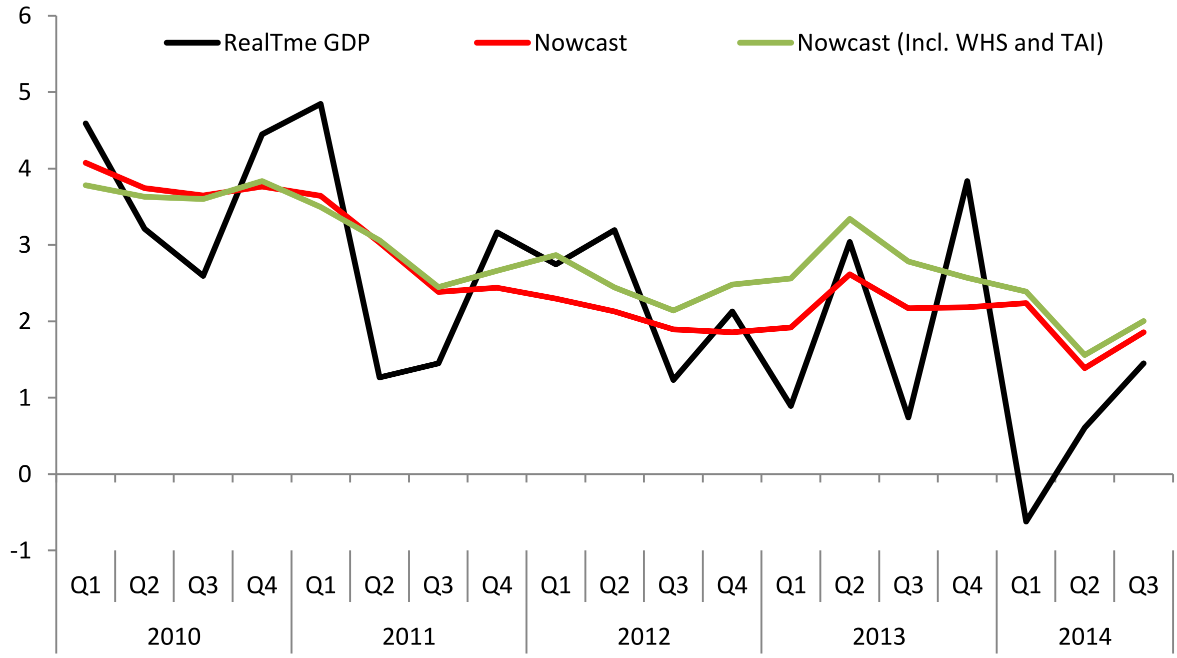

Figure 7: Nowcast vs Professional Forecasters

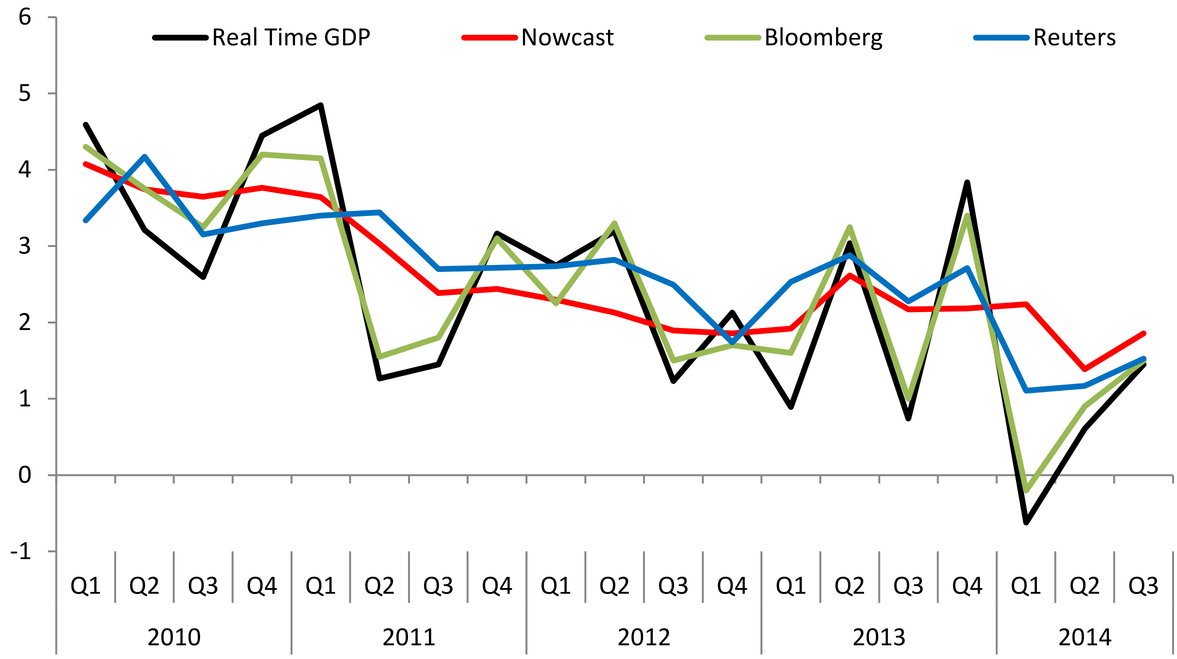

Figure 8: Nowcast vs AR(1) and VAR(1)