ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Regional Output Growth in the UK: Improving Estimates by Incorporating New Data Sources

GARY KOOP (STRATHCLYDE), STUART MCINTYRE (STRATHCLYDE), JAMES MITCHELL (FEDERAL RESERVE BANK OF CLEVELAND), AUBREY POON (OREBRO), AND PING WU (STRATHCLYDE).

Project overview

BACKGROUND AND CONTEXT: The Office for National Statistics (ONS) has long produced data on Gross Value Added (GVA) for the UK regions. However, until 2012 this data was produced at the annual frequency and with a release delay of approximately one year. In 2012, with the development of the Regional Short Term Indicator (RSTI) data set, regional GVA became available at the quarterly frequency and with a shorter, but still substantial, release delay of approximately six months. In addition, two of the ITL1 regions, Scotland and Northern Ireland, have produced their own quarterly GVA estimates (since 1998 and 2012 respectively) with a slightly different release delay.

In earlier work, we use the ONS's annual regional data and developed econometric methods for producing nowcasts and historical estimates of quarterly GVA growth in the UK regions in a more timely fashion. GVA growth data (for the UK and its regions) lay at the heart of the econometric models. We did add a small number of additional predictors, both at the national level (e.g. the exchange rate and inflation rate) and the regional level (e.g. the regional claimant count), to improve our estimates of regional GVA growth. But we did not incorporate the wider range of regional indicators that are increasingly being collected by the ONS and others.

WHAT WE HAVE DONE: We develop methods for extending our earlier model to include a wide range of regional indicators. We do this by incorporating regional factors into a regional MF-VAR. This enables us to exploit the information present in a wide range of regional variables. We then assess the effect of incorporating this information into our model in terms of improvements in the accuracy of our nowcasts of regional GDP growth.

A key issue which arises in trying to incorporate the wide range of candidate predictors that exist into our model is that many of these regional indicators are only available over a relatively short time period. We surmount this problem by creating regional factors which we then incorporate into our model. At the beginning of our sample, these regional factors reflect information in only a few regional variables, but as time passes and more variables become available these are included in the factors.

There are several ways these regional factors can be calculated and we compare the properties of these, and investigate to what extent they improve the regional nowcasts once they are included in the model.

WHAT WE FOUND: We have tried various ways of constructing these regional factors and considered different ways of including them in the MF-VAR. Although the differences across our various methods are small, they do exist. Our preferred model, which produces nowcasts which are better than the other models according to some forecasting metrics and as good as other models according to others is relatively parsimonious and computationally tractable. In particular, we find using a small number of factors (2) and a short lag length (1) works best.

Using the Tall Project (TP) method, which has computational benefits relative to EMPCA and theoretical benefits relative to Tall Wide (TW), to calculate the regional factors works slightly better than the other methods. Furthermore, the conventional strategy of selecting the factors with the highest eigenvalues is as good as the more complicated strategy of choosing factors based on their correlation with UK GVA growth.

Additional regional data

The predictors included in this model can be grouped into three main categories: measures of output, labour market data, and housing market indicators. Measures of output include data covering: construction sector output, retail sales, trade data, port traffic, tourism stays and spending, and business demographic information. The labour market data include headline employment and unemployment, and public sector employment, the claimant count rate, as well as the HMRC real-time information data from the PAYE system. The housing market data include monthly House Price published by the UK Government, as well as the quarterly UK house price index published by the Nationwide Building Society (more timely than the official data series), and rental prices for the private rental market.

Constructing the Regional Factors

We use three methods for constructing factors in the presence of missing data:

EMPCA: developed in Stock and Watson (2002) is an iterative algorithm, which makes it slower than the alternatives and it can also fail to converge.

TW: developed in Bai and Ng (2021) is not an iterative algorithm, but might be sensitive to the choice of variables and might not work well if one (or a few) of the variables is available for a short period of time.

TP: developed in Cahan, Bai, and Ng (2022) this approach is similar to TW, but uses auxilliary regressions to surmount some of the problems in the TW approach.

Project results

THE REGIONAL FACTORS – HOW DIFFERENT ARE THE FACTORS PRODUCED BY EACH METHOD?

Table 1: The three approaches (EMPCA, TW and TP) will produce regional factors. We examine the correlation across methods between the first three factors produced using each approach. The TP and TW approaches tend to produce factors that are more highly correlated with each other than with those produced using the EMPCA method

North East

	EMPCA F1	EMPCA F2	EMPCA F3	TW F1	TW F2	TW F3	TP F1	TP F2	TP F3
EMPCA F1	1.00	0.00	0.00	0.16	-0.65	-0.37	0.12	0.71	0.39
EMPCA F2	0.00	1.00	0.00	-0.92	-0.02	0.21	-0.90	-0.04	-0.32
EMPCA F3	0.00	0.00	1.00	-0.04	-0.72	-0.27	-0.09	0.75	0.14
TW F1	0.16	-0.92	-0.04	1.00	0.00	0.00	0.94	0.02	0.15
TW F2	-0.65	-0.02	-0.72	0.00	1.00	0.00	0.02	-0.97	0.11
TW F3	-0.37	0.21	-0.27	0.00	0.00	1.00	0.07	-0.23	-0.74
TP F1	0.12	-0.90	-0.09	0.94	0.02	0.07	1.00	0.00	0.00
TP F2	0.71	-0.04	0.75	0.02	-0.97	-0.23	0.00	1.00	0.00
TP F3	0.39	-0.32	0.14	0.15	0.11	-0.74	0.00	0.00	1.00

NOWCASTING REGIONAL GVA GROWTH – DOES ADDITIONAL REGIONAL DATA IMPROVE ACCURACY?

In running our model in pseudo-real-time we produce three estimates at each time point:

1) ***Nowcasts:*** estimates of growth in the quarter that the model is run and where we do not yet have official data for the UK as a whole,
2) ***Estimates:*** which reflect the allocation of the observed UK growth from the previous quarter (which has just been published at the time of the model being run) to the regions of the UK, and
3) ***Backcasts:*** estimates for the English regions and Wales of the yet to be released values of growth two quarters before (by this point data for Scotland and Northern Ireland have been published by the respective devolved administrations).

Our two evaluation metrics are the Root Mean Square Forecast Error (RMSFE) and the Continuous Ranked Probability Score (CRPS), which evaluate the point and density forecast performance of our models, respectively. In interpreting these, the lower the value the more accurate the estimates are.

Key conclusions:

1) there is a clear pattern of accuracy improving as we move from our nowcast to our estimates, to our backcasts;
2) Including the additional indicators does not seem to lead to any substantial improvement in the accuracy of the estimates or the backcasts, although there is an improvement in the nowcasts themselves.

Table 2: RMSFE

VB MFVAR - Adaptive Lasso (p=1 lag)	NE	NW	York	EM	WM	EE	LON	SE	SW	WA	SCOT	NI	Average
Nowcast	2.04	2.35	2.35	2.14	2.36	2.6	3.2	2.36	1.94	2.31	1.61	1.85	2.26
Estimate	0.53	0.55	0.55	0.56	0.51	0.58	1.12	0.47	0.60	0.63	0.17	0.37	0.55
Backcast	0.33	0.24	0.35	0.39	0.34	0.41	0.78	0.26	0.39	0.30	-	-	0.38
VB FAMFVAR (TP, p=1 lag, Nf=2 factors)													
Nowcast	1.85	1.77	1.78	1.88	1.75	1.78	2.77	1.62	1.64	1.87	1.36	1.5	1.8
Estimate	0.62	0.55	0.68	0.8	0.51	0.45	1.25	0.44	0.67	0.57	0.15	0.36	0.59
Backcast	0.38	0.32	0.47	0.53	0.36	0.36	0.81	0.26	0.44	0.27	-	-	0.42

Table 3: CRPS

VB MFVAR - Adaptive Lasso (p=1 lag)	NE	NW	York	EM	WM	EE	LON	SE	SW	WA	SCOT	NI	Average
Nowcast	0.97	1.00*	1.10	1.03	1.10*	1.06	1.29	1.06	0.97	1.04	0.67	0.78	1.01
Estimate	0.3	0.3	0.33*	0.34	0.32	0.29*	0.5	0.29	0.35*	0.36	0.11	0.21	0.31
Backcast	0.19	0.15	0.2	0.22	0.2	0.21	0.31	0.16	0.22	0.19	-	-	0.21
VB FAMFVAR (TP, p=1 lag, Nf=2 factors)													
Nowcast	0.91*	0.87*	0.89*	0.92*	0.86*	0.87*	1.23*	0.79*	0.82*	0.90*	0.56*	0.65	0.86
Estimate	0.36	0.31	0.35	0.4	0.31	0.28	0.58	0.27	0.35*	0.34*	0.11	0.21	0.32
Backcast	0.22	0.18	0.24	0.27	0.21	0.2	0.34	0.16	0.23	0.17	-	-	0.22

∗ denote 0.10 significance level for a two-sided Diebold and Mariano, 1995 test. The benchmark model is MF-VAR.

Key outputs

Koop, G., McIntyre, S.G., Mitchell, J., Poon, A., and Wu P., (2022), Regional Output Growth in the UK: Improving Estimates by Incorporating New Data Sources

University of Strathclyde Business School

ÖREBRO UNIVERSITET