Measuring Sub-Regional Economic Activity: Missing Frequencies and Missing Data*

Gary Koop[†], Stuart McIntyre[‡], James Mitchell[§], Aubrey Poon[¶], Ping Wu[‖]

November 8, 2024

Abstract: Bayesian mixed-frequency vector autoregressions (MF-VARs) are commonly used to produce timely and high-frequency estimates of low-frequency variables. A typical application uses quarterly data on output, for a given country, and monthly indicator data to produce monthly estimates of national output. But, when working at sub-national levels, data limitations preclude the use of standard MF-VARs. The frequency mismatch is more complicated, key variables can have missing data, and release delays can be substantial. In this paper, we develop a novel MF-VAR which addresses all these issues and use it to produce historical estimates of sub-regional output growth in the UK. The model combines information in the annual sub-regional data (when available) with data from the UK regions and the UK as a whole. The model is estimated using variational Bayesian methods with shrinkage priors, reflecting the "big data" setup. We use our model to produce a new database of quarterly estimates of sub-regional GVA growth back to the 1960s, that importantly, because the MF-VAR imposes temporal and cross-sectional restrictions, is consistent with those official data that do exist. We illustrate the use of these new estimates by showing how they can used to characterize the considerable heterogeneity in sub-regional business cycle dynamics in the UK and contribute to our understanding of regional economic resilience.

---

*We thank the editors and an anonymous referee for helpful comments. This research has been funded by the Office of National Statistics as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

[†]University of Strathclyde; Economic Statistics Centre of Excellence (gary.koop@strath.ac.uk)

[‡]University of Strathclyde; Economic Statistics Centre of Excellence (s.mcintyre@strath.ac.uk)

[§]Federal Reserve Bank of Cleveland; Economic Statistics Centre of Excellence (james.mitchell@clev.frb.org)

[¶]University of Kent; Economic Statistics Centre of Excellence (a.poon@kent.ac.uk)

[‖]University of Strathclyde; Economic Statistics Centre of Excellence (ping.wu@strath.ac.uk)

1 Introduction

Mixed-frequency vector autoregressions (MF-VARs) are a popular tool for producing timely and high-frequency estimates of macroeconomic variables such as GDP growth, given that official estimates of GDP tend to be made available only at a lower frequency and are published with a lag. Most such applications of MF-VARs to temporal disaggregation are at the national level; e.g., see Schorfheide and Song (2015). But, in many countries, there is also a keen interest in issues of regional or sub-regional resilience and inequality, making it desirable to produce similar estimates at a finer level of geographical granularity. However, particularly at more spatially granular (sub-regional) levels, efforts to produce sub-regional historical estimates of GDP have been stymied by the limitations of the data. At the national level, the higher-frequency "indicator" data used to help inform on the (unobserved) monthly path of GDP growth are not "missing", have decent historical coverage, and are released fairly quickly. For instance, Schorfheide and Song (2015) produce estimates of monthly GDP in the US based on quarterly GDP data and several monthly national indicators. The release delay for quarterly GDP itself is short (i.e., the so-called "advance" estimate of US GDP is released by the Bureau of Economic Analysis roughly a month after the end of each quarter). National data has similar properties in other countries.

At the regional and sub-regional levels,[1] data availability is more restricted. In this paper, our methods are designed around the specific characteristics of the data available in the UK. But many countries face similar data challenges at the sub-national level and our methods could be used more widely, with minor alterations. In the UK, regional and sub-regional data are often lower frequency, and at times even completely unavailable. In particular, sub-regional gross value added (GVA) data are missing before 1998.[2] Aggregating up one level, the regional data are "better," in the sense that they are available in some form (in nominal terms pre-1998, and in real terms thereafter) back to 1970. But their frequency changes over time: they were annual through 2011, and then switched to quarterly from 2012. Quarterly output data for Scotland and Northern Ireland are available over a slightly longer time-period than the other ITL1 regions of the UK. Given our focus on sub–regional analysis, we treat the regional data as starting when the Regional Short Term Indicators (RSTI) data produced by the ONS in the UK, described below, become available from 2012. The goal of this chapter is then to incorporate sub-regional, regional, and national data on output growth, each with their own properties, into an MF-VAR so as to produce higher-frequency (quarterly) historical

[1]When we refer to regional (sub-regional) data, we mean at the International Territorial Level (ITL) 1 (ITL2) level. When we refer to output, we mean Gross Value Added (GVA).

[2]Real GVA and real GDP are closely related concepts. GVA is in basic prices, while GDP is in market prices (i.e., GVA plus taxes (less subsidies) on products equals GDP).

estimates of sub-regional output growth.

The MF-VAR we construct differs from the conventional national setup for temporal disaggregation, as seen in Schorfheide and Song (2015)), in several ways. First, the MF-VAR is much larger. Even if we only use GVA growth data, with one nation, 12 regions, and 41 sub-regions, our MF-VAR will involve 54 equations. Adding additional indicator variables, either at the national or regional level, will further increase the dimension of the model. Second, it involves two kinds of *cross-sectional restriction.* One is based on the fact that output for any region must be the sum of outputs for its sub-regions. The other is that national GVA is the sum of the regions. Third, it involves a very large number of unobserved missing values to fill in. This arises since quarterly output data are never in fact observed at the sub-regional level and only observed subsequently to 2012 for most regions. Fourth, the sub-regional data are completely missing, even at the annual frequency, prior to 1998. Finally, the frequency of the regional data changes over time. The MF-VAR that we develop accommodates all of these messy data features.

We use our MF-VAR to produce a time series of quarterly sub-regional output growth data back to the 1960s. We illustrate the use of these new data by characterizing sub-regional business cycle dynamics in the UK. Such an analysis is not possible with extant lower-frequency and incomplete sub-regional data. We find that sub-regions of the UK differ significantly in terms of the timing of their switches between recessionary and expansionary regimes. Our results are thus consistent with those in Owyang et al. (2005), who find that state-level business cycles in the US are also not all alike. We also show a growing tendency, since the UK recession of 2008, for sub-regions of the UK to enter "local" recessions, suggesting that business cycle dynamics in the UK have become more heterogeneous in recent years and increasingly out-of-sync with national phases. We end by using our new database to measure (sub-)regional "economic resilience" across several dimensions commonly considered in economic geography.

2 National, Regional, and Sub-Regional Data

We use "official" data from the Office for National Statistics (ONS) on GVA at the sub-regional, regional, and national (UK) levels. Three characteristics of these data are relevant: 1) the number of variables, 2) the sample span, and 3) their frequency.

- **UK GVA Data:** At the national level, the real GVA sample begins in 1966, and is at the quarterly frequency.

- **Regional GVA Data:** There are 12 ITL1 regions in the UK. The sample begins with annual regional GVA data from 1966 through 1998. These data are available only in nominal terms so, in the absence of regional inflation data, we deflate by the UK deflator.[3] The sample then continues with real-terms annual regional GVA data from 1998. Then, from 2012 to the present, the Regional Short Term Indicators (RSTIs) database is available.[4] This database provides quarterly real GVA for the ITL1 regions. They were first published in September 2019.

How can these regional GVA data be used to improve sub-regional (ITL2) output estimates? From 1966-1997, the regional data will have information useful for producing historical estimates of ITL2 GVA, since ITL2 data are not available for this period. Accordingly, we impose a cross-sectional restriction to ensure the model-interpolated ITL2 data, in a given region, sum to the observed ITL1 data. From 1998-2011, the ITL1 data do not provide any information we can exploit, since they are at the same frequency as the ITL2 data. From 2012 onwards, we can exploit the fact that the ITL1 data are published at a higher frequency than ITL2 data.

- **Sub-regional GVA Data:** There are 41 ITL2 sub-regions. Real GVA data for these sub–regions are available on an annual basis from 1998, and align with the regional data described above over this period.

  Regional and sub–regional GVA data for the UK are released with much longer and more variable release delays than equivalent national data. Moreover, there were a series of disruptions to the publication schedule for these data following the onset of the COVID-19 pandemic and, as a result, our sample period ends in 2019.

- **UK macroeconomic variables:** In addition, as in Koop et al. (2020b), we incorporate four additional UK-wide quarterly macroeconomic indicators within our model, to help capture intra-year economic dynamics. These are the consumer price index, the US Dollar exchange rate with the pound sterling, the Bank of England's Bank rate, and the Brent oil price.

3 Econometric Methods

In this section, we develop our MF-VAR which combines the sub-regional, regional, and UK data into one model so as to produce estimates of quarterly sub-regional GVA growth. We

[3]Note that as set out in the Appendix to Koop et al. (2020a), the definitions of the UK ITL1 geographies have changed over the period since 1966. For example, the South East region was divided into a South East England region and London part way through the sample. We adopt the same approach as that in Koop et al. (2020a) to address these historical geography changes at the ITL1 level. ITL2 geographies have remained constant over the period for which we observe their growth rates.

[4]https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/quarterlycountryandregionalgdp.

emphasize that because official sub-regional GVA data are not available at any frequency prior to 1998, and so over this period our model relies exclusively on the annual regional and quarterly UK output data, the historical estimates of sub-regional output that our model produces prior to 1998 are likely of lower quality than our post-1998 quarterly estimates. This is because, from 1998, the sub-regional quarterly estimates of output produced by our MF-VAR can condition on observed official estimates of annual sub-regional output, as well as quarterly regional and UK data.

3.1 Notation and Key Data Features

We begin by describing some variable definitions, relationships, and defining our notational conventions. All changes and growth rates referred to below are exact (not log-differenced).

- $t = 1, .., T$ runs at the *quarterly* frequency.
- $r1 = 1, .., R1$ denotes the $R1$ regions in the UK.
- $r2 = 1, .., R2$ denotes the $R2$ sub-regions in the UK.
- Superscripts $UK, 1, 2$ will denote UK, regional (ITL1), and sub-regional (ITL2) variables.
- y_t^{UK} is UK quarterly GVA growth. Data are available throughout the sample.
- $y_t^{1,r1}$ is quarterly GVA growth for region $r1$. Data are available from 2012 onwards.
- $y_t^{2,r2}$ is quarterly GVA growth for sub-region $r2$. Data are never available.
- $y_t^{UK,A}$ is UK annual GVA growth. Data are available throughout the sample.
- $y_t^{1,r1,A}$ is annual GVA growth for region $r1$. Data are available through the sample, but only for quarter 4.
- $y_t^{2,r2,A}$ is annual GVA growth for sub-region $r2$. Data are available from 1998 onwards, but only for quarter 4.

These variables are stacked into vectors, such that: $y_t^j = \left(y_t^{j,1}, .., y_t^{j,Rj}\right)'$ is the vector of ITLj regional quarterly growth rates, for $j = 1, 2$. $y_t^{j,A}$ are the analogous ITL2 annual growth rates. $y_t = (y_t^{UK}, y_t^1, y_t^2)'$ is the vector containing all the quarterly growth rates. $y_t^A = (y_t^{UK,A}, y_t^{1,A}, y_t^{2,A})'$ is the analogous vector containing all the annual growth rates.

Quarterly variables are constrained to add up to annual quantities via the following vector of inter-temporal restrictions, as popularized by Mariano and Murasawa (2003):

$$y_t^A = \frac{1}{4}y_t + \frac{1}{2}y_{t-1} + \frac{3}{4}y_{t-2} + y_{t-3} + \frac{3}{4}y_{t-4} + \frac{1}{2}y_{t-5} + \frac{1}{4}y_{t-6}. \tag{1}$$

Koop et al. (2020b) discuss why this is the appropriate temporal disaggregation restriction to use when modeling output (GVA) in exact growth rates.

In addition, we exploit the restrictions that UK quantities are the sum of regional quantities and, in turn, regional quantities are the sum of sub-regional quantities. These so-called cross-sectional restrictions take the form:

$$y_t^{UK} = \sum_{r1=1}^{R1} w_t^{1,r1} y_t^{1,r1} + \eta_{t,UK} \tag{2}$$

where $w_t^{1,r1}$ is the share of region $r1$ in UK GVA. We allow for an error in this restriction and assume $\eta_{t,UK} \sim N(0, \sigma_{UK}^2)$. The main reason to include an error is to acknowledge the fact that GVA from the regions does not exactly add up to UK GVA, due to the absence of the output from the UK Continental Shelf (UKCS) in any of the regional data. But there are other reasons for including an error. For instance, the weights in the restriction should vary over time at the quarterly frequency, but our data only allows us to calculate them at the annual frequency. Furthermore, data revisions can cause small inconsistencies between UK and regional data.

Next, we have the cross-sectional restrictions that arise from the fact that sub-regional output adds up to regional output. These restrictions take the form:

$$y_t^{1,r1} = \sum_{r2 \in r1} w_t^{2,r2} y_t^{2,r2} + \eta_{t,1,r1} \tag{3}$$

for $r1 = 1, .., R1$ where $r2 \in r1$ denotes all the sub-regions which lie in region $r1$ and $w_t^{2,r2}$ is the share of sub-region $r2$ in the GVA of the region it lies within. The errors are assumed to follow: $\eta_{t,r1} \sim N(0, \sigma_{1,r1}^2)$. However, we expect these errors to be smaller than the error in (2), since the issue relating to the UKCS is not relevant at the sub-regional level. But we still do not expect this sub-regional constraint to hold exactly, as the weights used in (3) are fixed at their observed 1998 values pre-1998 and, thereafter, are updated annually rather than quarterly. This approximation, forced on us by data limitations, again motivates inclusion of an error, $\eta_{t,1,r1}$, in (3). We choose priors for $\sigma_{1,r1}^2$ for $r1 = 1, .., R1$ to reflect our belief that these errors should be smaller than σ_{UK}^2.

Finally, we have the cross-sectional restriction that arises from the fact that sub-regional data add up to UK data:

$$y_t^{UK} = \sum_{r2=1}^{R2} w_t^{2,r2} y_t^{2,r2} + \eta_{t,UK2} \qquad (4)$$

for $r2 = 1, .., R2$, where the errors $\eta_{t,UK2} \sim N(0, \sigma_{2,r2}^2)$. Note that, at first sight, it may seem redundant to incorporate both cross-sectional restrictions, since (3) implies that (4) must hold. But y_t^{UK} is released more quickly than $y_t^{1,r1}$ and, for much of the sample, is available at a higher frequency. These facts could also be exploited to help produce timely and high-frequency nowcasts of sub-regional data. Intuitively, when regional data are available, our model exploits them as the best source of information to update our sub-regional estimates. But, when they are not available, this model would fall back to exploiting information in the UK data. In other words, (4) would be useful in periods where UK data are available but regional data have not yet been released.

3.2 Bayesian Estimation of the Mixed-Frequency VAR

The MF-VAR is a state space model where the measurement equations are given by the inter-temporal and cross-sectional restrictions of the preceding sub-section. The state equations of the MF-VAR are based on a VAR with y_t being the vector of dependent variables. The MF-VAR treats some of the elements of y_t as observed and some as unobserved. The unobserved values are treated as states in a state-space model. But it is worth emphasising that y_t contains all the quarterly growth rates for the UK, the regions, and the sub-regions and thus comprises $1 + R1 + R2 = 54$ variables, which is quite large. And, when we use data back to 1966, the sub-regional data at both annual and quarterly frequencies are unobserved until 1998. These facts motivate our use of Bayesian methods. These allow us to incorporate a shrinkage prior to deal with the fact that Φ_j for $j = 1, .., p$ requires many coefficients to be estimated. We use the adaptive Lasso prior. Bayesian inference is implemented using the variational Bayes algorithm of Gefang et al. (2020). The remainder of this sub-section provides details.

In the large VAR literature it is increasingly common to work with a transformation of the VAR which has a diagonal error covariance matrix and, thus, allows for each equation to be estimated separately. This is done for computational reasons since working with the reduced form VAR requires matrix manipulations involving the potentially enormous posterior covariance matrix of the VAR coefficients. To explain the transformation of the VAR, note that the error variance-covariance matrix of the reduced form VAR, Σ, can be decomposed as $\Sigma = LDL'$, where L is lower triangular with ones on the diagonal and D is a diagonal matrix. If we multiply both sides of the VAR by L^{-1} then the resulting transformed VAR will have a diagonal error covariance matrix, thus allowing for equation-by-equation estimation.

This transformation leads to a VAR of the following form:

$$y_t = X_t\beta + W_t\alpha + \epsilon_t, \tag{5}$$

where X_t contains lags of the dependent variables along with the quarterly UK macroeconomic indicators. W_t contains the appropriate contemporaneous elements of y_t implied by the transformation using the Cholesky decomposition of Σ. We can rewrite (5) as a series of independent equations, with the i^{th} equation being:

$$y_{i,t} = \tilde{\mathbf{x}}_{i,t}\theta_i + \epsilon_{i,t}, \epsilon_{i,t} \sim N(0, \sigma_i^2) \tag{6}$$

where $\tilde{\mathbf{x}}_{i,t}$ is a row vector with k_i elements and θ_i is a vector containing the elements of β and α pertaining to the i^{th} equation.

Following Gefang et al. (2023), we specify the prior distributions for the parameters in each equation:

$$\theta_i \sim N(0, \mathbf{V}_i), \tag{7}$$

$$\sigma_i^{-2} \sim G(\underline{\nu}, \underline{s}), \tag{8}$$

where G denotes the Gamma distribution.

The adaptive Lasso is a hierarchical prior which treats $\mathbf{V}_i$ as a matrix of unknown parameters with their own prior of the form:

$$\mathbf{V}_i = \text{diag}(\tau_{i,1}, \ldots, \tau_{i,k_i}). \tag{9}$$

Note that this allows for the different equations to have different prior shrinkage. The adaptive Lasso assumes:

$$\tau_{i,j} \sim Expon(\tfrac{\lambda_{i,j}}{2}), \quad \text{for } j = 1, \ldots, k_i, \tag{10}$$

where $Expon$ is the exponential distribution, with:

$$\lambda_{i,j} \sim G(\underline{a}_0, \underline{b}_0), \tag{11}$$

and k_i is the number of right-hand-side variables in equation i.

For estimation, we use the computationally efficient variational Bayes algorithm for VARs with an adaptive Lasso prior developed in Gefang et al. (2023) and extended to MF-VAR models in Gefang et al. (2020). Building on the recent work of Chan et al. (2023), to achieve further computational efficiencies, we then extend the Gefang et al. (2020) framework by utilizing a precision-based rather than Kalman filter-based method to estimate the missing quarterly values. Specifically, we implement the formulation of equation (9) of Chan et al. (2023), the joint conditional distribution of the missing data given the observed data, to approximate the missing latent states. We follow Gefang et al. (2023) and make relatively non-informative choices for the prior hyperparameters, $(\underline{a}_0, \underline{b}_0 \underline{\nu}, \underline{s})$. Specifically, we set $\underline{a}_0 = \underline{b}_0 = 0.001$, $\underline{\nu} = 5$, and $\underline{s} = 0.04$.

4 Historical Estimates of GVA Growth at the Sub-Regional Level

Having estimated our MF-VAR, we first look at its historical estimates of regional and sub-regional output growth to reassure that they satisfy the necessary temporal and cross-sectional aggregation restrictions. Then, we turn to illustrating the use of the new quarterly sub-regional historical database for business cycle analysis and to inform on regional economic resilience.

4.1 Estimating Unobserved Regional Quarterly GVA

We compare our model-based estimates of quarterly regional and sub-regional GVA growth to the official, in general lower frequency, data published by the ONS. We focus on the posterior median estimates from our MF-VAR. We proceed by undertaking two sets of comparison. First, we compare regional growth estimates to the aggregation of our sub-regional growth estimates. Given that a stochastic aggregation constraint is used to link these two estimates in the MF-VAR, we do not expect them to be exactly the same, but we expect them to be close. Second, we compare our sub-regional estimates to the published/observed annual data.

To undertake the first comparison, Figures 1 - 2 presents three estimates for each of the 12 (ITL1) regions. The first estimate shown, in red, is the time series of MF-VAR-based quarterly regional growth estimates (plotted at an annualized rate). The second, in black, is the published ONS data, which prior to 2012 are annual (hence the same growth rate is plotted for each quarter of a given year), thereafter they are quarterly. Note that once a year (in Q4), prior to 2012, the red and black lines intersect. This shows that the intertemporal restriction, (1), is working. Note also that, for Scotland and Northern Ireland, the post-2012 data are from the Scottish and Northern Ireland governments, respectively. The third series, in blue, is the weighted aggregation of the growth rates at the sub-regional level.

We draw out a few features of these three sets of estimates. The first is that, over much of our sample period, the three series align well. This suggests that the model, and in particular the various constraints within it, are holding. After the introduction of the quarterly regional data from 2012 onwards, the estimates align a little less well but still broadly track developments in the regional economies. There are a couple of reasons why the alignment is weaker in this period.

The first helps us understand the divergence between our estimates and the quarterly regional estimates in the 2012-13 period, specifically that our model calculates annual growth rates using the inter-temporal restriction, (1), which utilises 7 lags of the quarterly regional growth rates. Until 2013Q4, we do not observe all 7 lags of regional growth as "official" data. Rather, up until this point in time, the inter-temporal constraint involves a mixture of estimated states from our model and the new ("official") RSTI data. For example, when calculating annualized quarterly growth for 2012Q1 (the red line), our model uses just one-quarter of "hard" RSTI data and 6 quarters' worth of interpolated data for the unobserved quarterly regional growth rates.

The second reason is that, post-2012, the model reconciles three different datasets: the RSTI data from the ONS, the Scottish GDP data produced by the Scottish Government, and the quarterly Northern Ireland Composite Economic Index produced by NISRA. Each of these data sources is constructed in a different way. This explains why in comparing our model-based estimates, which constrain to published UK-wide data, to these published series we should not expect perfect alignment. Moreover, even though the ONS produce the annual regional data that we use in our model, as well as the quarterly regional GDP data that we use for the English regions and Wales, these data themselves do not temporally align exactly. Of course, in principle, they should (over matching sample periods). But release delays and data revisions complicate alignment.

Turning to the second comparison, Figure 3 plots the MF-VAR-based sub-regional growth estimates against the annual estimates from the ONS, for the period from 1998 for which the latter are available. For space reasons, we focus on showing estimates for a selected subset of 12 of the 41 sub-regions of the UK. Again Figure 3 shows that once a year, in Q4, the MF-VAR-based estimates intersect, as they should, with the official annual estimates available from 1998.

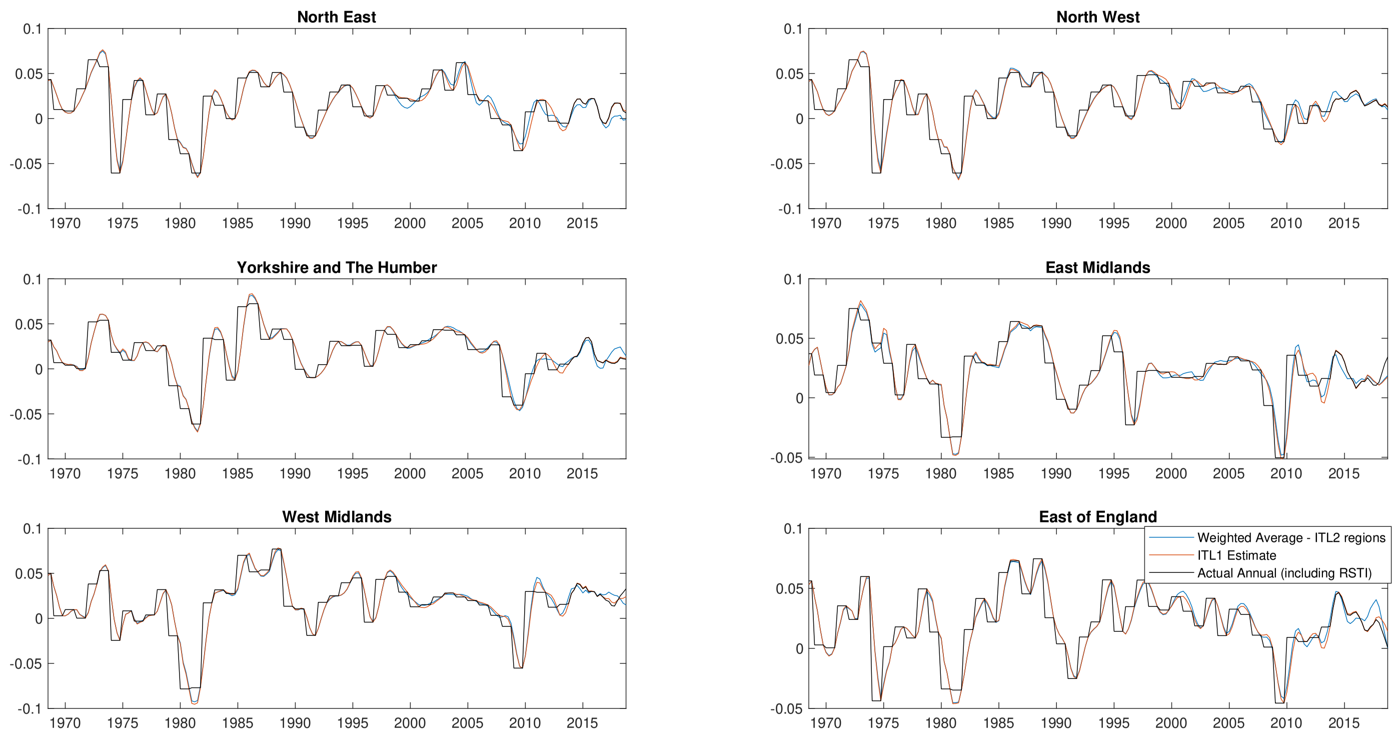

Figure 1: ITL1 quarterly GVA growth estimates (in %, annualized) and the cross-sectional (weighted) average of the ITL2 estimates from the MF-VAR, alongside the "official" annual estimates

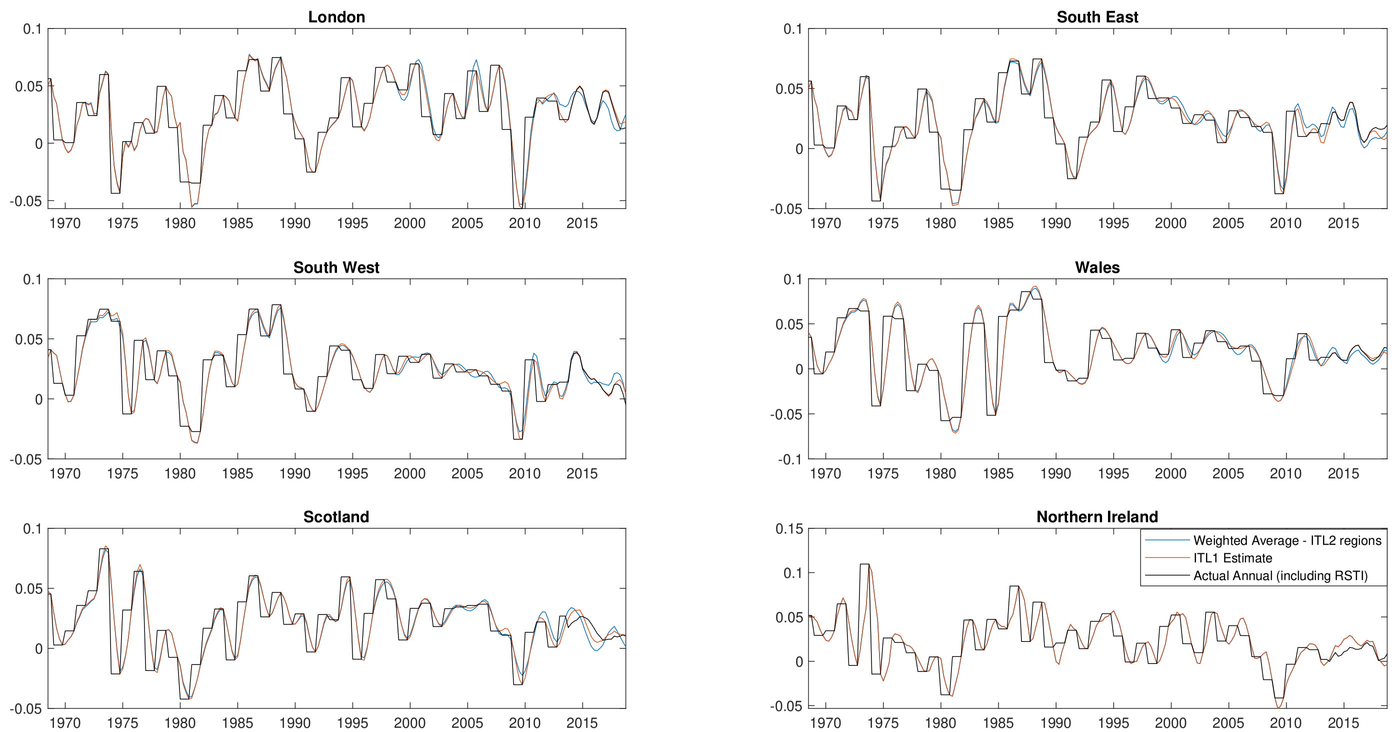

Figure 2: ITL1 quarterly GVA growth estimates (in %, annualized) and the cross-sectional (weighted) average of the ITL2 estimates from the MF-VAR, alongside the "official" annual estimates

Figure 3: ITL2 quarterly GVA growth estimates (in %, annualized) from the MF-VAR for 12 (of 41) sub-regions, alongside the "official" annual estimates

4.2 Heterogeneous Business Cycle Phases Across UK Sub-Regions

In this section, we illustrate how the new quarterly sub-regional data can be used to understand, at a higher temporal frequency than possible previously, the time profiles of recession and recovery in the sub-regions of the UK. We compare these with the five main recessions the UK as a whole has experienced since 1966. Our sample excludes, as discussed earlier, data covering the COVID-19 induced recession.

To identify the turning points that separate national and sub-regional business cycle expansions from contractions, we apply the popular nonparametric business cycle dating algorithm of Harding and Pagan (2002) to each of our sub-regional estimates (as produced by the MF-VAR) and to the UK data themselves, having transformed the growth estimates back into log-levels. We use as an input into the Harding-Pagan algorithm the median historical estimates of sub-regional GVA from our model. The Harding-Pagan dating algorithm seeks to formalize aspects of how the NBER date business cycles in the US, and has been found to match their turning points better than commonly used rules of thumb that characterize a recession as, for example, at least two consecutive quarters of negative growth.

The Harding-Pagan algorithm identifies five main recessions for the UK as a whole. These start in 1973Q3, 1974Q4, 1979Q3, 1990Q3, and 2008Q2. As seen in Figure 4, our new database informs that these UK downturns do tend to be accompanied, as we might expect, with downturns also occurring in many of the regions and sub-regions that comprise the UK. However, we also see that many regions and sub-regions were in recession when the UK was not. Additionally, the recessions of the 1970s were less pervasive than the later recessions, in the sense that even when the UK as a whole entered recession, fewer regions and sub-regions shared this same fate in the 1970s. The tendency for sub-regions of the UK to experience a recession when the UK did not was weaker in the Great Moderation period, between the UK recessions of 1990 and 2008. Interestingly, consistent with lacklustre growth for the UK as a whole, since the global financial crisis there appears to be an increased tendency for UK sub-regions to experience local recessions, with the share of ITL2 regions in a local recession often jumping sharply. This evidences a growing tendency for regional cycles to be more volatile and often decouple from the path of the UK as a whole.

To evidence further the apparent rising incidence of local recessions, that is sub-regional recessions that happen without a UK recession, in Figure 5 we zoom in to the period around the UK recession of 2008. Figure 5 plots the quarter in which each sub-region, region, as well as the UK as a whole, entered and exited the 2008 UK recession associated with the global financial crisis. This figure shows considerable variation as to the frequency and timing of recessions across sub-regions. While the UK as a whole has been in an expansionary phase since 2009Q3, all 41 sub-regions have experienced at least one recession since then. Many of these regional contractions were short-lived and intra-year, and would be missed without access to our new higher-frequency database.

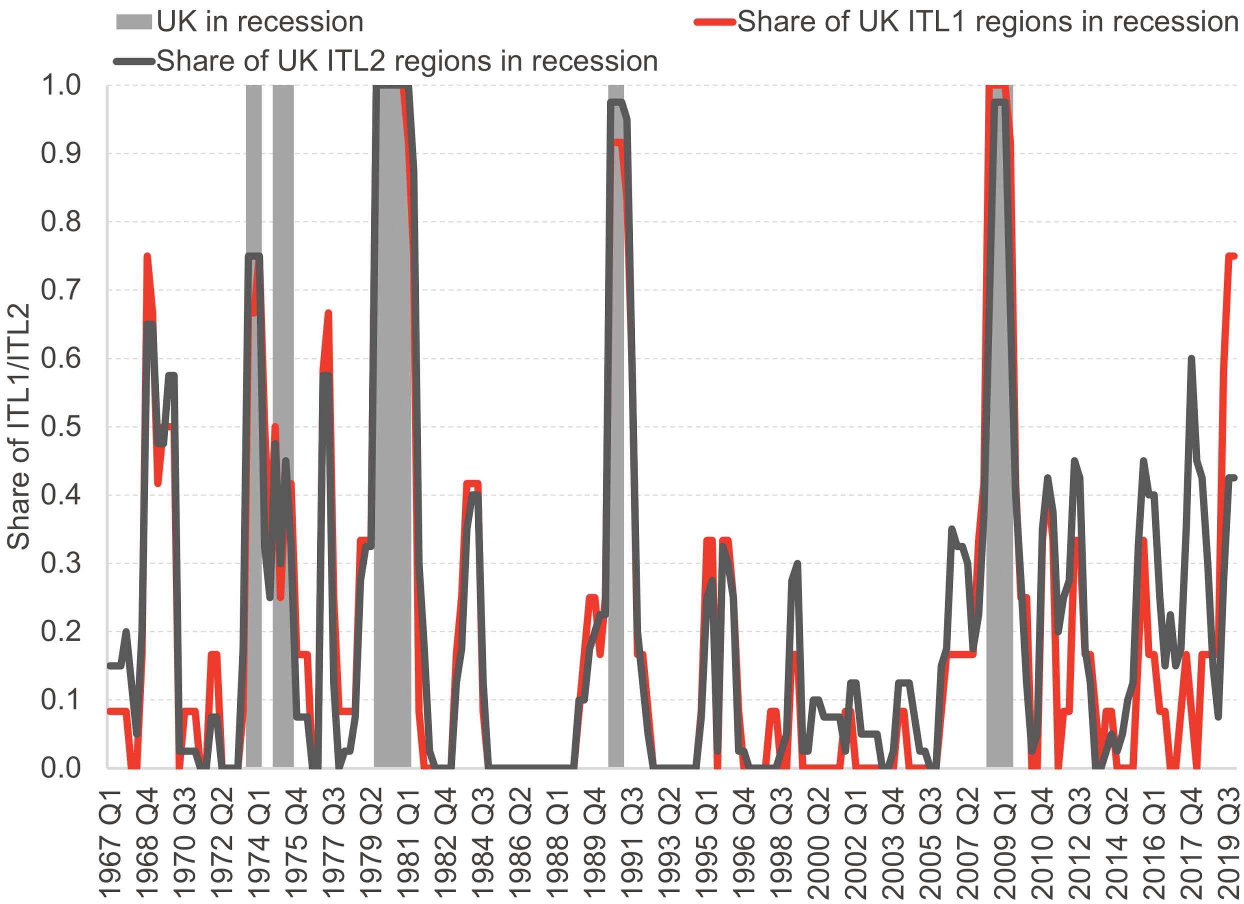

Figure 4: UK recession bands (vertical grey bars) alongside the share (proportion) of regions (ITL1) and sub-regions (ITL2) in (local) recession

Figure 5: Sub-region recessions (in white) and expansions (in black) by quarter from 2006

4.3 Further Evidence on Sub-Regional Business Cycle Heterogeneity

Our new database, plus the business cycle classification of the previous sub-section, let us take a deeper look at the link between national (UK) and regional/sub–regional business cycles, acknowledging that different regions/sub-regions may experience recessions at different scales, i.e., of differing magnitudes or depths. We compare the sensitivity of (sub-)regional economic growth to national economic growth with a measure of the relative (to the national economy) "gap" between a (sub-)region's average growth rates in expansionary and contractionary phases of the business cycle. This enables us to identify (sub-)regions which are more/less aligned to the national economy in terms of both overall growth and their business cycle properties.

We measure the sensitivity of regional output to movements in national output by the estimated β coefficient in the following CAPM-style regression model. This relates, at the quarterly frequency, regional or sub-regional growth (as estimated by the MF-VAR) to UK growth:

$$y_t^r = \alpha + \beta y_t^{UK} + \epsilon_t. \tag{12}$$

β thus measures the volatility or sensitivity of each region's or sub-region's growth to the UK (the "market") as a whole.[5] When $\beta > 1$, a region is relatively sensitive to movements in UK economic activity. When $\beta < 1$, the region is relatively insensitive.

In Figure 6, for each sub-region we plot OLS estimates of its β against, on the Y-axis, the difference in its average growth rate in expansionary and contractionary phases *relative* to the equivalent calculation for the UK as a whole. A value of unity on the Y-axis therefore indicates that a sub–region experienced the same gap between its average growth rates in expansionary and contractionary phases of the business cycle as the UK as a whole.[6]

Figure 6 illustrates that some sub–regions of the UK, most notably Inner London, have much larger differences between their expansionary and contractionary growth rates and, in turn, are much more sensitive to UK GDP movements than others. Given the importance of the aggregate London region (shown in red) to the UK economy, this is perhaps unsurprising, but these results make the point that London as a region is not homogeneous, with Inner and Outer London having rather distinct business cycle features. Sub–regions of the UK which are more rural tend to be found in the bottom left corner of Figure 6; they are both less sensitive to movements in UK GDP and experience much smaller differences between their growth rates in expansionary and contractionary phases of the business cycle.

Notable too is the North Eastern Scotland region, which over our sample period was

[5] We treat y_t^{UK} as weakly exogenous in (12), given that most regions/sub-regions (with the possible exception of London) are small relative to the size of the UK economy.

[6] ITL1 regions are shown in red, while ITL2 regions are shown in black.

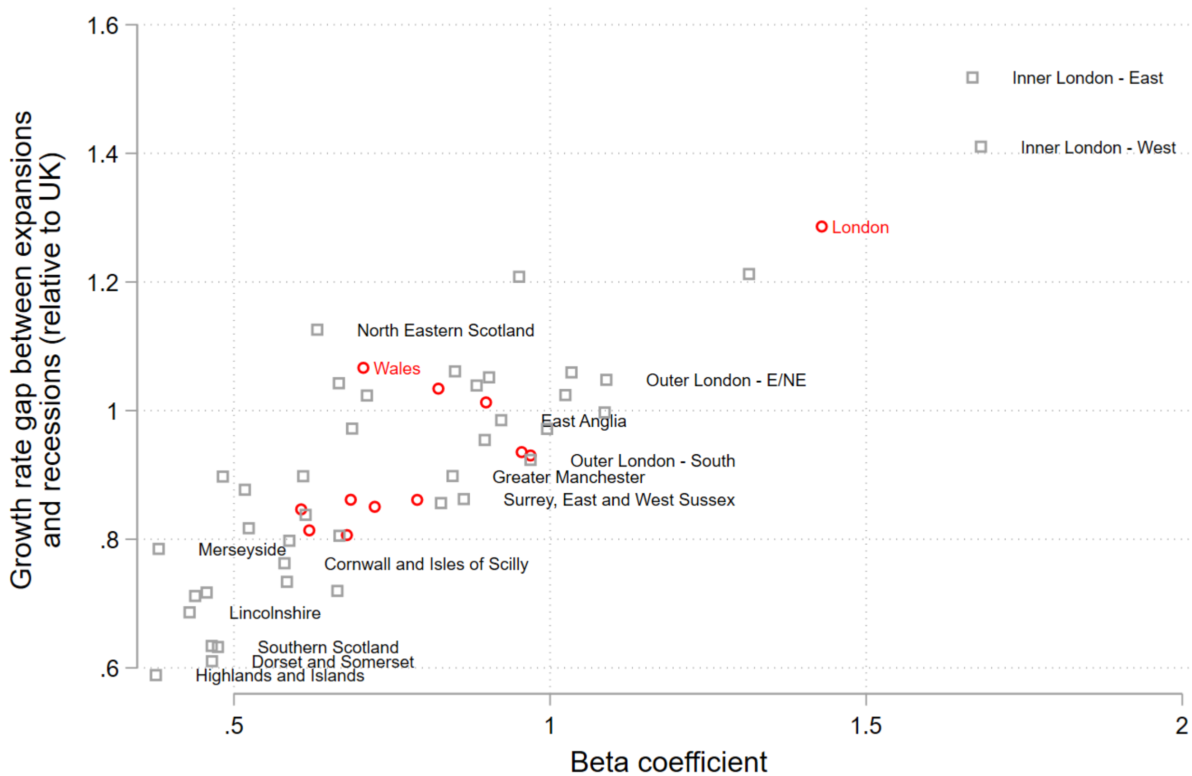

Figure 6: Relative depth of (sub-)regional business cycles (sub-regional average growth rate in expansions minus recessions) versus estimates of β (the sensitivity of each sub-region to the UK, as estimated via (12))

heavily engaged with oil and gas activity in the North Sea. It experiences larger differences in its expansionary versus contractionary growth rates than the overall UK economy and has less sensitivity to UK GDP movements. This is not that surprising, given that its economy is largely driven by global factors. Comparing to the aggregate Scotland region, we again see the advantages of considering business cycle features at a more disaggregated spatial scale, given the evident heterogeneity within regional economies. Continuing this theme, the next section demonstrates the insights that emerge from using our new data to consider different dimensions of regional and sub–regional economic resilience.

4.4 Economic Resilience

This section extends the analysis of how our new regional database sheds light on business cycle heterogeneities by assessing regional economic resilience. We follow the existing literature in economic geography by distinguishing between *resistance, renewal, recovery, and re-orientation*, which Martin (2012) defines as characterizing the dimensions of regional economic resilience.

Improvements in each of these dimensions reflect how a regional economy can become more resilient. This relates to the wider literature on regional economic resilience in the UK (Simmie and Martin, 2010; Martin, 2012; Fingleton et al., 2012; Martin and Sunley, 2015; Sensier et al., 2023). We focus on the first three of these measures of resilience, as the fourth – *re–orientation* – explores changes in a region's economic structure. This would be best explored by disaggregation of UK output not just into regions but sectors too.[7]

The first dimension of regional economic resilience is *resistance*, defined by (Martin, 2012, p.11) as: "...the vulnerability or sensitivity of a regional economy to disturbances and disruptions, such as recessions". Intuitively, it measures the extent to which a region is able to "resist" the effects of a negative economic shock. We measure this as the peak–to–trough change (in percentage points) in the level of output (but this could be equally applied to other metrics such as employment), reflecting that more resilient regions should see smaller drops in output in the face of an economic shock. The second dimension of resilience is *recovery*, which we measure as the number of quarters it takes a given sub-region to regain its pre-crisis (peak) level of economic activity. It reflects the expectation that more resilient regions should rebound more quickly from an economic shock. Thirdly, we measure *renewal* as the number of quarters it takes a given sub-region to regain its pre-crisis growth rate. This fits with Martin (2012) defining *renewal* as the "extent to which [the] regional economy renews its growth path: resumption of pre-recession path or hysteretic shift to a new growth path" (Martin, 2012, 12). The pre–crisis growth rate is calculated as the average annual growth rate for each region (or sub–region) from the beginning of the sample (1966 Q4) until the onset of each downturn. If it is the case that there has been a significant degree of damage done by the recessionary period to the core 'economic base' of a region, we might see this reflected in the region's growth rate experiencing a downward hysteretic shift.

In Figure 7 we present a heatmap of these three measures of economic resilience across the 1979 and 2008 UK recessions. Sub–regions are organized by country (England, Wales, and Scotland), and within England the sub–regions are further (broadly) organized from North to South. What emerges is a clear sense of the geographic (sub-regional) pattern of these two national downturns being rather different. In particular, across the three measures of resilience, the 1979 recession hit sub–regions in the South of England much less than those in the North. The 2008 recession, though, had a significant impact in the North but also in London according to the resistance measure of resilience, while the South of England was again much less affected in general, London aside. As expected, looking across the three measures of resilience, sub-regions tend to *renew* before they *recover*. It takes longer for economic growth

[7] Modeling such data would require estimation of an even larger version of our model, with even more missing data. We leave development of such a model to future research.

	1979			2008		
	Recovery Quarters	Renewal Quarters	Resistance %	Recovery Quarters	Renewal Quarters	Resistance %
Tees Valley and Durham	29	12	11	29	23	4
Northumberland and Tyne and Wear	26	11	10	16	8	5
Cumbria	33	12	7	6	6	2
Greater Manchester	28	11	11	16	15	4
Lancashire	27	11	13	24	7	8
Cheshire	24	11	13	15	8	6
Merseyside	28	12	9	10	21	2
East Yorkshire and Northern Lincolnshire	26	13	8	48	25	11
North Yorkshire	25	13	10	25	10	7
South Yorkshire	24	11	9	25	24	7
West Yorkshire	25	11	15	27	26	6
Derbyshire and Nottinghamshire	18	11	7	10	8	6
Leicestershire, Rutland and Northamptonshire	22	11	10	22	7	8
Lincolnshire	23	13	6	19	9	6
Herefordshire, Worcestershire and Warwickshire	25	11	17	11	8	6
Shropshire and Staffordshire	28	14	11	20	8	7
West Midlands	28	12	18	21	8	8
East Anglia	22	13	8	15	9	5
Bedfordshire and Hertfordshire	22	13	9	24	20	9
Essex	23	12	8	24	7	6
Inner London - West	23	12	10	16	11	8
Inner London - East	23	12	10	11	8	6
Outer London - East and North East	24	10	11	21	7	11
Outer London - South	22	10	9	44	8	10
Outer London - West and North West	19	10	8	18	10	9
Berkshire, Buckinghamshire and Oxfordshire	22	13	11	13	9	5
Surrey, East and West Sussex	21	13	8	9	8	4
Hampshire and Isle of Wight	22	13	6	9	7	4
Kent	24	14	7	21	7	7
Gloucestershire, Wiltshire and Bath/Bristol area	15	12	7	9	7	5
Dorset and Somerset	17	13	4	15	9	3
Cornwall and Isles of Scilly	23	13	7	20	9	6
Devon	22	11	6	9	7	4
West Wales and The Valleys	26	11	12	13	9	5
East Wales	25	11	11	12	9	4
North Eastern Scotland	26	24	9	6	9	2
Highlands and Islands	24	13	4	5	6	1
Eastern Scotland	24	13	8	18	18	3
West Central Scotland	24	13	6	19	11	5
Southern Scotland	24	13	5	21	8	5

Figure 7: Comparison of *Recovery*, *Renewal*, and *Resistance* of recessionary contractions in the UK in 1979 and 2008

rates in the recovery stage of a sub-region's business cycle to cumulate such that they push the sub-region back above its pre-recession level of output. Economic resilience is inherently a complex concept, and the heatmap in Figure 7 illustrates the diversity of experience between geographical areas and across different recessions. Some sub-regions are consistently blue/red across all three measures, but many are not. Instead, on some measures, these sub-regions seem more/less resilient than they do on other measures.

5 Conclusions

This chapter develops a "big data" mixed-frequency Bayesian VAR model to produce historical quarterly estimates of sub-regional output growth in the UK. The model takes on the challenge that extant official output growth data in the UK vary in their availability both across time, regions, and sub-regions. This makes for a complicated and evolving data landscape. Our MF-VAR model is estimated using variational Bayesian methods with shrinkage priors. It imposes not just temporal but hierarchical aggregation constraints, such that the sub-regional data "add up" to the region, in which the sub-regions resides, and, in turn, the regional data "add up" to the UK (national) data.

We use the new quarterly sub-regional historical database to show that sub-regions of the UK differ considerably in their business cycle dynamics, in terms of when they enter and exit recessions. Such cross-sub-regional variation in business cycle dynamics would be obscured if we focused on extant annual regional data. Our focus in this chapter is using the new data to establish these "stylized facts" about sub-regional business cycle dynamics. We defer formal economic explanation of them to future research. Understanding *why* these regional and sub–regional business cycles differ would no doubt benefit from extending our model further to incorporate sectoral detail. One could then model how the sectoral composition of the sub-regional economies, and changes in this over time, explains the observed business cycle heterogeneities.

Our model could also be used to produce nowcasts, that is, more timely (out-of-sample) real time estimates of sub-regional economic activity. Previous research (Koop et al. (2024)) has found that conditioning quarterly nowcasts for the UK regions (ITL1) on the more timely UK GVA estimate improves their accuracy. As more real-time data accumulate and thereby enable recursive out-of-sample simulations, future research should assess whether this result continues to hold at the more disaggregated sub-regional (ITL2) level.

References

Chan, Joshua CC, Aubrey Poon, and Dan Zhu (2023). "High-dimensional conditionally Gaussian state space models with missing data." *Journal of Econometrics*, 236(1), p. 105468.

Fingleton, Bernard, Harry Garretsen, and Ron Martin (2012). "Recessionary shocks and regional employment: Evidence on the resilience of UK regions." *Journal of Regional Science*, 52(1), pp. 109–133.

Gefang, Deborah, Gary Koop, and Aubrey Poon (2020). "Computationally efficient inference in large Bayesian mixed frequency VARs." *Economics Letters*, 191, p. 109120.

Gefang, Deborah, Gary Koop, and Aubrey Poon (2023). "Forecasting using Variational Bayesian inference in large Vector Autoregressions with hierarchical shrinkage." *International Journal of Forecasting*, 39, pp. 346–363.

Harding, Don and Adrian Pagan (2002). "Dissecting the cycle: a methodological investigation." *Journal of Monetary Economics*, 49(2), pp. 365–381.

Koop, Gary, Stuart McIntyre, and James Mitchell (2020a). "UK regional nowcasting using a mixed frequency vector auto-regressive model with entropic tilting." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(1), pp. 91–119.

Koop, Gary, Stuart McIntyre, James Mitchell, and Aubrey Poon (2020b). "Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970." *Journal of Applied Econometrics*, 35(2), pp. 176–197.

Koop, Gary, Stuart McIntyre, James Mitchell, and Aubrey Poon (2024). "Using stochastic hierarchical aggregation constraints to nowcast regional economic aggregates." *International Journal of Forecasting*, 40(2), pp. 626–640.

Mariano, Roberto S. and Yasutomo Murasawa (2003). "A new coincident index of business cycles based on monthly and quarterly series." *Journal of Applied Econometrics*, 18(4), pp. 427–443.

Martin, Ron (2012). "Regional economic resilience, hysteresis and recessionary shocks." *Journal of Economic Geography*, 12(1), pp. 1–32.

Martin, Ron and Peter Sunley (2015). "On the notion of regional economic resilience: conceptualization and explanation." *Journal of Economic Geography*, 15(1), pp. 1–42.

Owyang, Michael T., Jeremy Piger, and Howard J. Wall (2005). "Business cycle phases in u.s. states." *The Review of Economics and Statistics*, 87(4), pp. 604–616.

Schorfheide, Frank and Dongho Song (2015). "Real-time forecasting with a mixed-frequency VAR." *Journal of Business & Economic Statistics*, 33(3), pp. 366–380.

Sensier, Marianne, Anthony Rafferty, and Fiona Devine (2023). "The economic resilience scorecard: regional policy responses for crises recovery." *Regional Studies*, pp. 1–13.

Simmie, James and Ron Martin (2010). "The economic resilience of regions: towards an evolutionary approach." *Cambridge Journal of Regions, Economy and Society*, 3(1), pp. 27–43.