Report 45: Quantifying the economic response to COVID-19 mitigations and death rates via forecasting Purchasing Managers' Indices using Generalised Network Autoregressive models with exogenous variables.

Guy P Nason and James L Wei[1]

Department of Mathematics, Imperial College London, UK

[1]Corresponding author: james.wei19@ic.ac.uk

Summary

Knowledge of the current state of economies, how they respond to COVID-19 mitigations and indicators, and what the future might hold for them is important. We use recently developed generalised network autoregressive (GNAR) models, using trade-determined networks, to model and forecast the Purchasing Managers' Indices for a number of countries. We use networks that link countries where the links themselves, or their weights, are determined by the degree of export trade between the countries. We extend these models to include node-specific time series exogenous variables (GNARX models), using this to incorporate COVID-19 mitigation stringency indices and COVID-19 death rates into our analysis. The highly parsimonious GNAR models considerably outperform vector autoregressive models in terms of mean-squared forecasting error and our GNARX models themselves outperform GNAR ones. Further mixed frequency modelling predicts the extent to which that the UK economy will be affected by harsher, weaker or no interventions.

Quantifying the economic response to COVID-19 mitigations and death rates via forecasting Purchasing Managers' Indices using Generalised Network Autoregressive models with exogenous variables

Guy P. Nason and James L. Wei✉†

Imperial College, London, UK.

Summary. Knowledge of the current state of economies, how they respond to COVID-19 mitigations and indicators, and what the future might hold for them is important. We use recently developed generalised network autoregressive (GNAR) models, using trade-determined networks, to model and forecast the Purchasing Managers' Indices for a number of countries. We use networks that link countries where the links themselves, or their weights, are determined by the degree of export trade between the countries. We extend these models to include node-specific time series exogenous variables (GNARX models), using this to incorporate COVID-19 mitigation stringency indices and COVID-19 death rates into our analysis. The highly parsimonious GNAR models considerably outperform vector autoregressive models in terms of mean-squared forecasting error and our GNARX models themselves outperform GNAR ones. Further mixed frequency modelling predicts the extent to which that the UK economy will be affected by harsher, weaker or no interventions.

Keywords: multivariate time series, forecasting, joint medical and economic modelling, network time series.

1. Introduction

The severe impact of COVID-19 on global economic conditions cannot be understated, with UK real gross domestic product (GDP) contracting by a record 20% in the second quarter (Q2) of 2020 alone. Timely estimates of economic activity through the current turbulent period and forecasts of the impact of government virus mitigation strategies will be crucial for evaluating the costs and benefits of different policy interventions.

This article introduces new methods to forecast Purchasing Managers' Indices (PMIs) during the COVID-19 pandemic and quantifies the potential impact of government virus mitigation strategies. PMIs are survey-based leading indicators of the direction of economic activity. We choose to model PMIs as they are used extensively for the nowcasting of slower-moving headline measures of economic activity, such as GDP. See, for example, D'Agostino and Schnatz (2012). Our analysis builds on the class of generalised network autoregressive (GNAR) models (Knight et al., 2016; Knight et al., 2020), which

†*Address:* Dept of Maths, 180 Queen's Gate, London, SW7 2AZ, UK. Correspondence: james.wei19@ic.ac.uk

model multivariate time series in a highly parsimonious fashion by associating the dependencies between constituent series with an underlying network structure. Hence, GNAR models can characterise a rich diversity of dependence structures, whilst incurring a lower risk of overfitting when compared to classical vector autoregression (VAR) models, translating into superior forecasting performance. Indeed, our work below demonstrates that GNAR models outperform VAR models by 33% based on mean-squared forecasting error for forecasting PMIs from historical PMI data.

A goal here is to model PMIs in terms of historical PMIs, PMIs from neighbouring countries via our network model and (i) stringency measures of government intervention policies enacted to mitigate the spread of COVID-19 and (ii) the monthly number of confirmed COVID-19 deaths per million. Incorporating information about past PMIs from neighbouring countries via a network structure improves forecasting as it is extra relevant information. Our network model uses countries' PMIs on the network nodes (vertices) and the strength of link between two countries (nodes) increases with increasing quantities of exports between the two. Figure 1 (below) shows an example of one of the networks we use. In addition, the incorporation of multiple exogenous time-dependent node-specific regressors is both new to GNAR modelling and practically essential for properly understanding the potential influence of (i) and (ii). Our new extension is named GNARX, where X indicates incorporation of fully time-dependent exogenous variables, which is in line with similar naming conventions, such as VAR to VARX.

We structure our paper as follows. Section 2 introduces the new GNARX specification and characterises it as a specially-constrained linear regression problem. Asymptotic properties of the corresponding estimators follow from this. Section 3 provides details of our multivariate time series and network data and discusses two experiments: the first compares out-of-sample forecasting performance of GNAR models (i.e. without using external regressors) with a VAR model, and the second evaluates the GNARX model estimates when the stringency indices and COVID-19 death rates are incorporated. Section 4 uses our new GNARX modelling ability to analyse UK economic conditions under three different scenarios pertaining to either tightening, easing or unchanged intervention policies over the next six months. Section 5 concludes with a brief discussion of the experimental results and avenues for future research.

2.   The GNARX mathematical model and notation

Definition. Let our multivariate real-valued time series $\{Y_{i,t}\}$ be observations collected on variable $i \in \mathcal{K} = \{1, \ldots, N\}$ at time $t = 1, \ldots, T \in \mathbb{N}$, where $N \in \mathbb{N}$ is the number of variables. We sometimes write the multivariate series as the vector time series $\{\mathbf{Y}_t\} = (Y_{1,t}, \ldots, Y_{N,t})^\intercal$, where $\intercal$ denotes transpose, for $t = 1, \ldots, T$. Now let $\mathcal{G}$ be a graph consisting of a set of vertices/nodes, $\mathcal{K}$, and an edge set denoted by $\mathcal{E}$. Let $i \leftrightsquigarrow j \in \mathcal{E}$ denote the presence of an undirected edge between nodes i and j and $i \rightsquigarrow j \in \mathcal{E}$ denote the existence of a directed edge from i to j, for $i, j \in \mathcal{K}$. For network time series a graph vertex $i \in \mathcal{K}$ corresponds to time series variable $\{Y_{i,t}\}$. Let $\{X_{h,i,t}\}$ be the h-th real-valued stationary exogenous node-specific regressor series for node i at time t, for $h = 1, \ldots, H \in \mathbb{N}$, the number of such regressors.

We introduce the GNARX$(p, \mathbf{s}, \mathbf{p}')$ model, where $(p, \mathbf{s}, \mathbf{p}') \in \mathbb{N} \times \mathbb{N}_0^p \times \mathbb{N}_0^H$, by

$$Y_{i,t} = \sum_{j=1}^{p} \left(\alpha_{i,j} Y_{i,t-j} + \sum_{r=1}^{s_j} \beta_{j,r} \sum_{q \in \mathcal{N}^{(r)}(i)} \omega_{i,q} Y_{q,t-j} \right) + \sum_{h=1}^{H} \sum_{j'=0}^{p'_h} \lambda_{h,j'} X_{h,i,t-j'} + u_{i,t}, \quad (1)$$

where the parameters $\alpha_{i,j}, \beta_{j,r}, \lambda_{h,j'} \in \mathbb{R}$ for all i, j, r, h, j', p is the autoregressive order of the model and also maximum order of neighbour time lags, p'_h (the hth element of $\mathbf{p}'$) denotes the maximum lag of the hth exogenous regressor involved in the model, and s_j (the jth element of $\mathbf{s}$, which we sometimes write as $[s]$) is the maximum stage of neighbour dependence for time lag j. For example, $s_1 = 2$ implies that nodes are dependent on the first lags of its first and second stage neighbours in the network $\mathcal{G}$. The set $\mathcal{N}^{(r)}(i)$ is the r-th stage neighbourhood set of node $i \in \mathcal{G}$ defined by

$$\mathcal{N}^{(r)}(i) = \mathcal{N}\{\mathcal{N}^{(r-1)}(i)\}/[\{\cup_{q=1}^{r-1}\mathcal{N}^{(q)}(i)\} \cup \{i\}], \quad (2)$$

for $r = 2, 3, \ldots$, $\mathcal{N}^{(1)} = \mathcal{N}$, where $\mathcal{N}(A) = \{j \in \mathcal{K}/A : i \rightsquigarrow j; i \in A\}$. The $\omega_{i,q} \in [0,1]$ in (1) are network connection weights, typically set as the inverse of some prior notion of distance between nodes, as in Knight et al. (2020) using a similar approach to that used in the lifting scheme described by Jansen et al. (2009). Sometimes we consider the global-α model where $\alpha_{i,j} = \alpha_j$, i.e. use the same $\{\alpha_j\}_{j=1}^p$ sequence for each node. Consequently, the local-α model is where the $\alpha_{i,j}$ sequence is different for each node i.

The GNARX model cannot be captured by the standard GNAR model of Knight et al. (2020) by incorporating the exogenous regressors into $\mathbf{Y}_t$, as p and p'_h may differ. The network vector autoregression model proposed by Zhu et al. (2017) does incorporate dependence on node-specific covariates, but unlike model (1), they do not vary with time and have limitations on other parameters/quantities in the network. Full specifications for both of these models can be found in the Supplementary Material Section 1.

GNARX models share the same advantages with respect to missing data as GNAR models, especially when compared to vector autoregression-based approaches. This is because, e.g., VAR models try to estimate parameters that link one specific variable to another. If, for example, one of those variables disappears for a significant block of time, then it can be difficult to estimate the parameter associated with any of its pairings. By contrast, in a GNARX model, a variable pair is just a nearest neighbour and every such pair in the data contributes to the 'nearest neighbour' parameter, one of the $\beta_{j,r}$. So, even if one node suffers from significant missingness, there will be usually many other pairs so that the respective parameter is still well-estimated. More details on this, particularly on how this works for rth-stage neighbours, can be found in Section 2.6 of Knight et al. (2020).

GNARX Parsimony. The GNARX$(p, \mathbf{s}, \mathbf{p}')$ model with a single exogenous regressor ($H = 1$) contains a total of $M = Np + \sum_{j=1}^{p} s_j + p' + 1$ parameters. By comparison, the corresponding vector autoregression model with exogenous variables (VARX) containing p lags of the modelled series $\{\mathbf{Y}_t\}$, and p' lags of the exogenous series $\{\mathbf{X}_t\} = \{(X_{1,t}, ..., X_{N,t})^{\intercal}\}$, can be given in the reduced form by

$$\mathbf{Y}_t = \phi_1 \mathbf{Y}_{t-1} + \cdots + \phi_p \mathbf{Y}_{t-p} + \Lambda_0 \mathbf{X}_t + \Lambda_1 \mathbf{X}_{t-1} + \cdots + \Lambda_{p'} \mathbf{X}_{t-p'} + \mathbf{u}_t, \quad (3)$$

which contains $P = N^2(p + p' + 1)$ parameters. The parsimonious nature of GNARX compared to VARX is clear when parameter growth is $\mathcal{O}(N)$ for the former and $\mathcal{O}(N^2)$ for the latter. Indeed, it has been suggested by Bernanke et al. (2005) that vector autoregressions are not often used for macroeconomic datasets that contain more than six to eight time series. The parsimony of GNARX, combined with the flexibility to model dynamic networks representing multiple covariates, ensures our methodology is readily applicable to many applications involving the forecasting of multivariate time series.

GNARX Stationarity. Under the assumption of stationarity of $\{X_{h,i,t}\}$ for all h, i and parameter constraints

$$\sum_{j=1}^{p} \left(|\alpha_{i,j}| + \sum_{c=1}^{C} \sum_{r=1}^{s_j} |\beta_{j,r,c}| \right) < 1 \quad \forall i \in 1, \dots, N, \tag{4}$$

the GNARX process $\{\mathbf{Y}_t\}$ is stationary. We prove this in Section 2 of the Supplementary Material. For simplicity, we will assume $H = 1$ for the remainder of this section only, replacing $X_{h,i,t}$ with $X_{i,t}$ and $\mathbf{p}'$ with p'. Extending the following results to multiple external node-specific regressors is straightforward.

Precise model definition and theoretical statistical properties. With the $H = 1$ assumption, the GNARX$(p, \mathbf{s}, p')$ model can be expressed as a vector autoregression of the form (3), but where coefficient matrices are restricted so that $\phi_k = \text{diag}\{\alpha_{i,k}\} + \sum_{r=1}^{s_k} \beta_{k,r} W^{(r)}$, $[W^{(r)}]_{l,m} = w_{l,m} \mathbb{I}\{m \in \mathcal{N}^{(r)}(l)\}$, and $\Lambda_k = \lambda_k I_N$. In matrix form, the model can be written as $Y = BZ + U$, where $Y = [\mathbf{Y}_{p^*+1}, \dots, \mathbf{Y}_T]$, $p^* = \max(p, p')$, $B = [\phi_1, \dots, \phi_p, \Lambda_0, \Lambda_1, \dots, \Lambda_{p'}]$, $Z = [\mathbf{Z}_{p^*}, \dots, \mathbf{Z}_{T-1}]$, $\mathbf{Z}_t^{\mathsf{T}} = [\mathbf{Y}_t^{\mathsf{T}}, \dots, \mathbf{Y}_{t-p+1}^{\mathsf{T}}, \mathbf{X}_{t+1}^{\mathsf{T}}, \dots, \mathbf{X}_{t-p'+1}^{\mathsf{T}}]$, and $U = [\mathbf{u}_{p^*+1}, \dots, \mathbf{u}_T]$. Additional regressors can be incorporated by concatenating the corresponding coefficient and data matrices to B and $\mathbf{Z}_t$, respectively.

Following Lütkepohl (2005), the coefficient restrictions on B that result from the GNARX assumptions are $\text{vec}(B) = R\gamma$, where R is the $P \times M$ model matrix, defined in the Supplementary Material Section 4, $\underset{M \times 1}{\gamma} = (\gamma_1^{\mathsf{T}}, \dots, \gamma_p^{\mathsf{T}}, \lambda_0, \dots, \lambda_{p'})^{\mathsf{T}}$, with GNAR parameters that we are interested in given by $\gamma_k = (\alpha_{1,k}, \dots, \alpha_{N,k}, \beta_{k,1}, \dots, \beta_{k,s_k})^{\mathsf{T}}$ and all elementary vectors are of length N. With the global alpha assumption, $\alpha_{i,k} = \alpha_k$ for all $i = 1, \dots, N$, we have $\gamma_k = (\alpha_k, \beta_{k,1}, \dots, \beta_{k,s_k})^{\mathsf{T}}$ and $A = L$.

Given the VAR matrix representation and linear coefficient restrictions from above, Lütkepohl (2005) proposes the feasible generalised least squares (FGLS) GNAR parameter estimator to be

$$\hat{\gamma}^{FGLS} = \left\{ R^{\mathsf{T}} \left(ZZ^{\mathsf{T}} \otimes \hat{\Sigma}_u^{-1} \right) R \right\}^{-1} R^{\mathsf{T}} \left(Z \otimes \hat{\Sigma}_u^{-1} \right) \text{vec}(Y), \tag{5}$$

where the maximum likelihood estimator of Σ_u is

$$\hat{\Sigma}_u = T^{-1} \left(Y - \hat{B}Z \right) \left(Y - \hat{B}Z \right)^{\mathsf{T}}, \tag{6}$$

and where $\hat{B}$ satisfies $\text{vec}(\hat{B}) = R\hat{\gamma}$ and $\hat{\gamma}$ is the ordinary least squares estimator of γ. In the case of missing observations, the corresponding residuals are set to zero. Under the conditions of stationarity of $\{\mathbf{Y}_t\}_{t=1,\dots,T}$ and $\{\mathbf{X}_t\}_{t=1,\dots,T}$, consistency and asymptotic normality are proved in Section 2 of the Supplementary Material.

3.   Purchasing Managers' Indices: the data and forecasting experiments

3.1.   *Input data*

3.1.1.   *Purchasing Managers' Indices*

IHS Markit Composite Purchasing Managers' Indices (PMIs) are computed by aggregating the views of senior purchasing executives from around 400 companies on a wide range of economic variables pertinent to the manufacturing and services sectors. Composite PMIs can take values between zero and 100, with a reading above 50 indicating economic expansion. All PMI series have been seasonally adjusted using the X-12 ARIMA method; full details of the survey methodology are available at `https://ihsmarkit.com/products/pmi.html`. Currently, monthly composite PMI data are available for thirteen major economies, which are all included in our analyses. The longest series are from the UK, Italy and Germany, which start from Jan-98, and the shortest is from Australia, which starts from May-16.

    PMI series are diffusion indices, which treat a fixed set of component economic indicators effectively as dummy variables that represent whether the trend in the component is positive or negative. For PMIs, these component series are the survey responses of individual purchasing managers. It is therefore reasonable to assume that diffusion indices are stationary, given the well-documented cyclical nature of macroeconomic time series, see Zarnowitz (1991). Moreover, Augmented Dickey-Fuller (ADF) unit root tests on the in-sample period between Jan-98 and Dec-17 result in p-values above 0.1 for all series, indicating stationarity for the time being, against the alternative hypothesis of first-order non-stationarity. Broadly, the series also do not reject the null of second-order stationarity according to the wavelet-based tests of Nason (2013, 2016).

3.1.2.   *Network construction*

GNAR and GNARX models require us to specify a network structure to associate with the PMI dataset. The time series for a particular country's PMI corresponds to a single vertex in the network — one vertex for each country. Rather than attempting to learn the network structure indirectly from the time series as in Leeming (2019) or Knight et al. (2020), we suggest that a construction based on the flow of exports between countries that provides a reasonable reflection of the association between component series. This rests on the assumption that a country's business confidence is partly dependent on the level of economic activity of its major export partners. Since (i) model orders are chosen by BIC optimisation, (ii) network parameters are highly significant and (iii) forecast performance is improved when compared to models that ignore network information, these all suggest that our choice of networks is reasonable. Evaluating performance by forecast accuracy is honest and exacting.

    Our experiments evaluate two different networks:

(a) **Fully-connected trade network**. Each country is connected to every other country. Edge weight $w_{i,j}$ is set to be the export quantity between countries i and j. The weights are then normalized so that $\sum_j w_{i,j} = 1$ for all i. In this case, $s_j \in \{0, 1\}$ for all j.
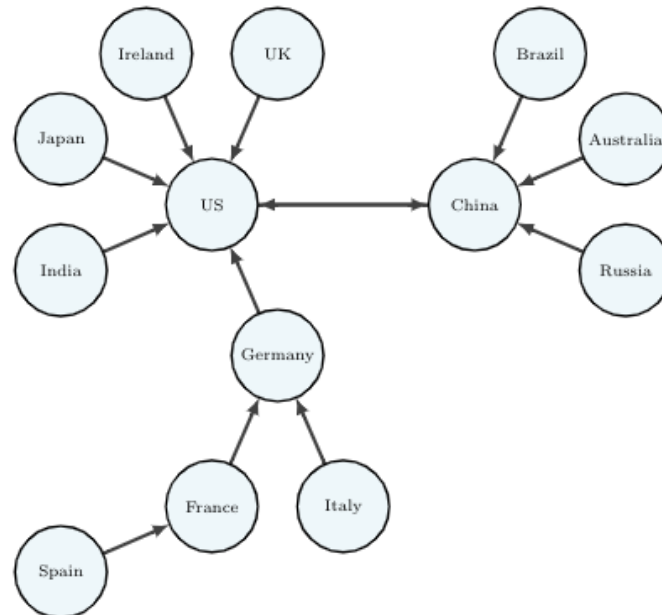
Fig. 1. Nearest neighbour network based on 2019 export data

(b) **Nearest neighbour trade network**. (Figure 1) Each country has a single outgoing edge to the country that it exports the most too.

Export data was obtained from the UN Comtrade International Trade Statistics Database (`https://comtrade.un.org/data/`), using the latest available observations.

3.1.3. Exogenous variables
The GNARX model's advantage is that it permits the multivariate time series $\mathbf{Y}_t$ to depend on vertex-specific external time series regressors $\mathbf{X}_t$; two series in our examples. The first, $X_{1,i,t}$, are differenced country stringency indices (Hale et al., 2020) for country i, which aim to quantify the severity of COVID-19 mitigation policies. These are composite indices that take values between zero and 100, with higher values reflecting

stricter measures. The stringency indices are mainly related to containment policies, such as the closing of schools, workplaces and public events. The full list of constituent indicators is available at `https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation`. Values of the stringency indices before the COVID-19 outbreak are automatically set to zero, rather than treated as missing.

The second regressor, $X_{2,i,t}$, is the first difference of country i new COVID-19 death rates, which may impact economic conditions via channels beyond the previously discussed interventions. For example, consumers may extrapolate a rise in the confirmed number of COVID-19 deaths into the future, leading to consumer precautionary saving in anticipation of greater future lockdown measures. As higher death rates would likely also be associated with stricter mitigation policies, omitting death rates would likely lead to model errors being correlated with the stringency index regressor, resulting in inconsistent parameter estimates. Our experiments use the daily number of confirmed COVID-19 deaths per million in a given country, averaged for each month (Roser et al., 2020), available at `https://ourworldindata.org/covid-deaths`. The sample correlation between the differenced stringency indices and the differenced COVID-19 death rate of the same country is only 0.46, suggesting that multicollinearity is not a significant problem. The correlation only rises modestly to 0.57 for the non-differenced series.

Figure 2 shows the UK PMI and the two exogenous variables starting from Jan 2019, where all series have been standardised to have zero mean and unit variance so that they can be compared on the same axes. The most prominent feature of the UK PMI series is the large six and eight standard deviation declines during Mar and Apr 2020. This fall in purchasing manager confidence coincides with the first UK lockdown (rise in stringency index), but also reflects broader worsening of global macroeconomic conditions caused by the COVID-19 pandemic, see, e.g., International Monetary Fund (2020).
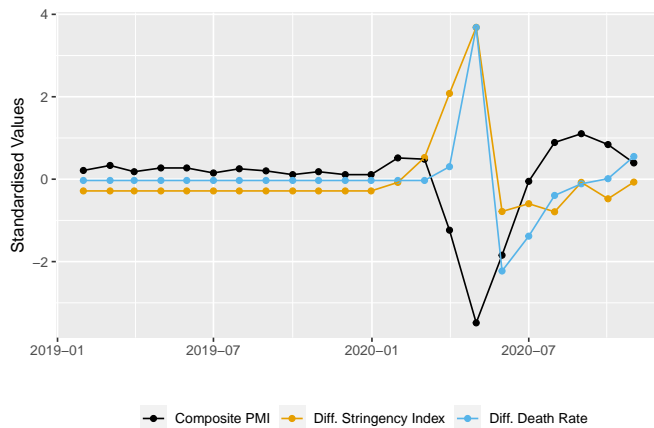

Fig. 2. Plot of UK Composite PMI together with UK stringency index and death rate covariates, where all variables have been standardised to have zero mean and unit variance.

Table 1. Out-of-sample composite PMI forecast performance between Jan-18 and Oct-20 using GNAR, VAR and naive forecasting. MSFE=Out-of-sample mean-squared forecasting error; SE = standard error

Model	Model order	Parameters	MSFE (SE)
Global-α GNAR, fully-connected net	2, [1, 1]	4	41.9 (8.2)
Global-α GNAR, nearest-neighbour net	2, [1, 1]	4	40.7 (7.4)
Local-α GNAR, fully-connected net	1, [1]	13	38.4 (6.8)
Local-α GNAR, nearest-neighbour net	1, [1]	13	39.3 (7.1)
VAR	2	288	57.3 (9.3)
AR	1	12	39.1 (7.0)
Naive forecast	-	-	40.6 (7.2)

3.2. GNAR forecasting performance

Here, model order is selected to minimise the Bayesian Information Criterion (BIC) subject to an order limit of $p = 12$. For computational feasibility, we adopt a stagewise approach, where p is first selected to minimise BIC, followed by s_j for $j = 1, ..., p$. This allows us to capture network effects lasting up to a year in duration. For the exogenous regressors in later experiments, we instead impose a lower order limit of three months, due to the lack of nonzero observations. In that case, p'_h for $h = 1, ..., H$ are optimised at the same time as s_j. All PMI observations up to Dec-17 are treated as the in-sample period for model order selection and estimation of both GNAR and VAR models. We refer readers to the Supplementary Material Section 7 for the results if model orders were instead selected to minimise forecasting errors in a subset of the in-sample period. Finally, we provide simulation evidence for the consistency of the model selection procedure using stagewise BIC optimisation for GNARX in the Supplementary Material Section 8, which shows similar performance to 'global search' BIC.

Unlike GNAR, VAR models do not easily handle missing values. Specifically, GNAR models deal with missingness by modifying connection weights (Knight et al., 2020, Section 2.6) and because of this advantage our GNAR model can make use of the *entire* unbalanced panel starting from Jan-98. If Australia were included, then the in-sample period for the VAR model would contain only 20 months and, for this reason, we exclude Australia from this analysis.

Table 1 shows the composite PMI forecasting performance of four different stagewise-BIC selected GNAR models, a VAR model, a naive forecast (predicting next from current) and univariate AR models. The latter involve a separate AR model estimated for each country, with order determined by BIC optimisation. For all but three countries, AR(1) was selected. However, for fair comparison with the GNAR and VAR models where autoregressive order is the same for every series, AR(1) was used for all countries. It is worth noting that AR(1) is identical to GNAR(1, [0]) with local-α. We have used mean-squared forecast error (MSFE) in our article, which computes the empirical average of the squared differences between our point forecasts and the eventual out-turn. The lowest forecast error is attained by the local-α GNAR model with the fully-connected network.

Furthermore, the mean-squared forecast error of all four GNAR models are sub-

Table 2. Local-α GNAR$(1, [1])$ parameter estimates (to two decimal places) using the fully-connected network. Parameter p-values were all $< 0.001\%$.

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
$\alpha_{\text{USA},1}$	0.93	$\alpha_{\text{Japan},1}$	0.87	$\alpha_{\text{Germany},1}$	0.90
$\alpha_{\text{UK},1}$	0.90	$\alpha_{\text{Italy},1}$	0.89	$\alpha_{\text{China},1}$	0.88
$\alpha_{\text{Spain},1}$	0.92	$\alpha_{\text{Ireland},1}$	0.92	$\alpha_{\text{Brazil},1}$	0.88
$\alpha_{\text{Russia},1}$	0.92	$\alpha_{\text{India},1}$	0.91	$\alpha_{\text{France},1}$	0.92
$\beta_{1,1}$	0.07				

stantially lower than that of the VAR model, indicating overfitting in the latter. The local-α GNAR$(1, [1])$ model contains only thirteen parameters, whilst the VAR(2) model contains $2N^2 = 288$ parameters. Our results indicate that when a simple model is appropriate, BIC optimisation for VAR may still lead to overfitting, but the same does not hold for GNAR models. Indeed, only three of the four GNAR models outperformed the naive forecast and one of the two local-α GNAR models outperformed the univariate AR models, reflecting the difficulty in forecasting economic indicators using only past values. On the other hand, the best GNAR model *does* improve the forecast performance and a few percent improvement can translate into substantial economic value when one is dealing with GDP-like magnitudes, see Section 4.

Table 2 shows estimates for the local-α GNAR$(1, [1])$ model, with p-values computed using HC2 robust standard errors, which are heteroskedasticity-consistent estimators first proposed by MacKinnon and White (1985) that adjust residuals using the projection matrix leverage values. These were calculated with the **sandwich** package of Zeileis (2004). In this experiment, the use of robust standard errors produces more conservative p-values.

Figure 3 plots the UK out-of-sample one-step-ahead rolling forecasts for this best-performing model, along with those from the VAR(2) model (see Supplementary Material, Section 6, for other countries). The VAR forecasts appear to underestimate the initial PMI decline at the start of the crisis period, but substantially overestimate the recovery. This leads to substantially higher mean-squared forecast error during the COVID-19 crisis period starting from Feb-20, relative to GNAR.

Overall, BIC choice selects for higher autoregressive order in the global-α GNAR models compared to the local-α variant, which also results in higher out-of-sample mean-squared forecast error. This suggests that the PMIs' autoregressive dynamics differs between countries, although the autoregressive parameters in Table 2 are similar. Moreover, BIC choice of the local-α models also favours exploiting the network structure implied by export flows, whether by using the fully connected or the nearest neighbour network: $\mathbf{s} \neq \mathbf{0}_p$ for both of the best-performing GNAR models.

3.3. GNARX forecasting performance with external regressors

Given the shortness of the exogenous regressors, it is not feasible now to evaluate out-of-sample forecasting performance for the GNARX models, so we switch to in-sample forecast evaluation. For example, predicting 'October' based on exogenous variables that only existed since early Spring, i.e. few observations, is not sensible. Here, we estimate model parameters on the whole series and then, given a time point, we use the estimates,
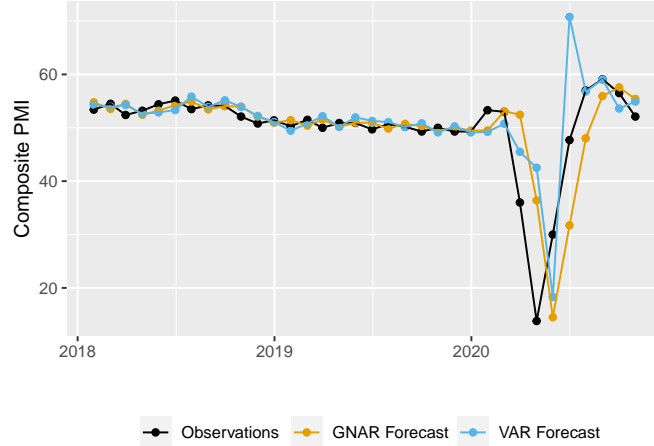
Fig. 3. Out-of-sample one-step-ahead rolling composite UK PMI forecasts using the local-α GNAR$(1, [1])$ and the VAR(2) models.

but use observations only from previous time points. In future, with more data, we will be able to switch to out-of-sample forecasting performance evaluation.

We can now reintroduce the shorter Australian PMI series, using the same method for handling missing observations as in Section 3.2. On our relatively short time series with missing data, vector autoregressions with exogenous variables (VARX) as described in, for example, Tsay (2013), which contain $N^2(p + p' + 1)$ parameters, are infeasible. Recall that some VAR models may be estimated as separate linear regressions for each country. A simple VARX(1,0) model would involve the estimation of 13 coefficients corresponding to the stringency indices and 13 corresponding to the death rates in each country. However, our dataset contains fewer than 13 nonzero observations for each exogenous variable of each country, clearly leading to a parameter identification problem.

Table 3 shows in-sample mean-squared forecasting errors for our stagewise-BIC selected GNARX models using the fully-connected network (results for the nearest neighbour network and no network were always worse and reported in the Supplementary Material, Section 11). The inclusion of exogenous variables results in substantially lower BIC and an almost halving of the mean-squared forecast error across all settings compared to GNAR without external regressors. The lowest BIC model is the global-α GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$, with estimates shown in Table 4 and p-values calculated as in Table 1.

Table 4 shows that the 1st and 5th positive autoregressive coefficients are significant. As before, current-period PMIs are also significantly positively dependent on previous-period PMIs of neighbours weighted by export volumes. On the other hand, the negative $\beta_{3,1}$ and $\beta_{5,1}$ parameter estimates are smaller in magnitude.

The zero-lag coefficient for COVID-19 intervention stringency indices is highly significant, matching our expectation of a strong coincident negative impact of containment policies on business confidence. A country's COVID-19 death rate also has a significant negative contemporaneous impact on the PMI, indicating that the pandemic might be

Table 3. In-sample composite PMI forecast performance between Jan-98 and Oct-20 using GNARX models with the fully-connected network. MSFE=In-sample mean-squared forecasting error.

Model	Model order	BIC	MSFE (SE)
Without external regressors			
Global-α GNAR	4, [1, 1, 1, 0]	12653	7.1 (0.9)
Local-α GNAR	3, [1, 1, 1]	12806	6.7 (0.9)
Stringency indices only			
Global-α GNARX	5, [1, 0, 1, 0, 1], 3	11323	4.2 (0.3)
Local-α GNARX	2, [1, 1], 3	11331	4.0 (0.3)
COVID-19 death rates only			
Global-α GNARX	4, [1, 1, 1, 0], 3	12020	5.5 (0.8)
Local-α GNARX	2, [1, 1], 3	12212	5.5 (0.8)
With both external regressors			
Global-α GNARX	5, [1, 0, 1, 0, 1], 3, 2	11175	4.0 (0.3)
Local-α GNARX	2, [1, 1], 3, 1	11199	3.8 (0.3)

Table 4. Global-α GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$ model parameter estimates using the fully-connected network and both exogenous regressors. † means p-value is < 0.001. Values significant at the 5% level in bold.

Param.	Est.	p-value %	Param.	Est.	p-value %	Param.	Est.	p-value %
$\boldsymbol{\alpha_1}$	0.83	†	$\boldsymbol{\beta_{1,1}}$	0.28	†	$\boldsymbol{\lambda_{1,2}}$	0.15	0.8
α_2	0.00	97.5	$\boldsymbol{\beta_{3,1}}$	−0.15	4.2	$\boldsymbol{\lambda_{1,3}}$	0.10	3.3
α_3	0.03	59.2	$\boldsymbol{\beta_{5,1}}$	−0.08	3.5	$\boldsymbol{\lambda_{2,0}}$	−0.91	0.2
α_4	−0.05	34.0	$\boldsymbol{\lambda_{1,0}}$	−0.31	†	$\lambda_{2,1}$	0.42	21.3
$\boldsymbol{\alpha_5}$	0.09	0.6	$\lambda_{1,1}$	−0.09	24.8	$\lambda_{2,2}$	0.41	11.3

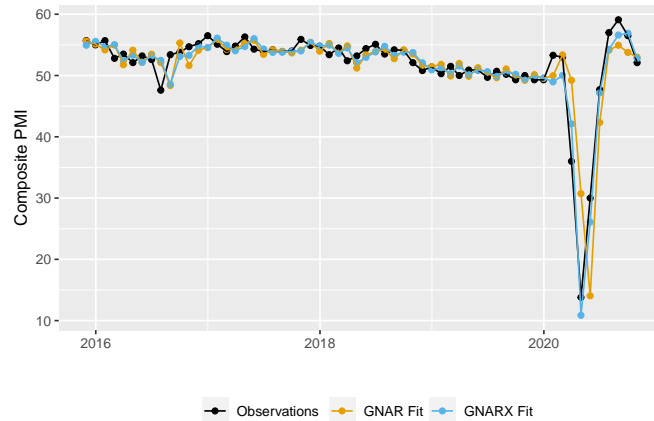**Fig. 4.** UK: Comparison of GNARX$(5, [1, 0, 1, 0, 1])$ and GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$ fits.

influencing business confidence through channels beyond policy interventions.

Figure 4 shows the global-α GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$ UK fits over the most recent five-year period, to illustrate the quality of fit during the COVID-19 period (other countries' charts are given in the Supplementary Material, Section 10). The inclusion of stringency indices and COVID-19 death rates allows the GNARX model to anticipate both the initial composite PMI decline at the start of the crisis and the subsequent recovery, leading to a much lower mean-squared forecasting error than the corresponding GNAR model.

4. UK scenario analysis

We now use the fitted the global-α GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$ model from the previous section to produce six-month forecasts of the UK composite PMI series under different assumptions on the evolution of the UK stringency index. For other examples of conditional forecasting on multivariate economic time series, we refer readers to the seminal paper by Waggoner and Zha (1999) and the more recent summary provided by Bańbura et al. (2015), which describe conditional forecasting techniques and applications for VAR models. We would note that these describe more difficult problem than ours, as we condition on exogenous variables. Figure 5 provides forecasts of the UK composite PMI under three different policy intervention scenarios for the six month period from Nov-20 to Apr-21, assuming no change in the stringency indices of other countries from Aug-20 or in COVID-19 death rates:

(a) **Easing COVID-19 intervention measures**. The stringency index decreases linearly from the Oct-20 value of 67.9 to 0.0 between Nov-20 and Apr-21.
(b) **No change in COVID-19 intervention measures**. The stringency index is constant at the Oct-20 value of 67.9 between Nov-20 and Apr-21.
(c) **Tightening in COVID-19 intervention measures**. The stringency index increases linearly from the Oct-20 value of 67.9 to 100.0 between Nov-20 and Apr-21.
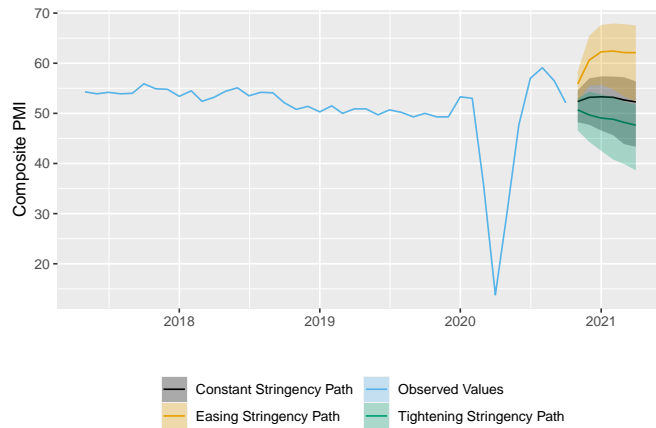
Fig. 5. Forecasts of UK composite PMI under three different scenarios for policy interventions using the GNARX$(5, [1, 0, 1, 0, 1], 3, 2)$ model. 95%-coverage prediction intervals estimated using 1000 bootstrap samples.

As might be expected, the GNARX model predicts significantly higher UK composite PMI under assumption of the UK stringency index falling to zero, holding death rates constant. Using bootstrap prediction intervals, computed using the algorithm introduced in the Supplementary Material Section 12, the expected PMI under the easing scenario exceeds the 97.5th percentile of the PMI under the tightening and constant stringency scenarios after only one month. The Apr-21 prediction of PMI is 62.1 in the easing scenario (which would be a record high) and 47.6 in the tightening scenario. In all cases, the estimated sampling distribution of the PMI is negatively skewed.

We now estimate how these composite PMI paths would translate into headline quarter-on-quarter real GDP growth, using data sourced from `https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyq/qna`. At the time of writing, 2020 Q3 is the latest available observation of UK GDP. Given that the GDP series is sampled quarterly, while the PMI series is sampled monthly, we assume a mixed-frequency linear relationship between the GDP growth and PMI series, using the MIDAS regression model (Ghysels et al., 2004). Analysis of both the unrestricted MIDAS model and the restricted MIDAS model, using exponential Almon lag polynomial weights, suggest that quarterly GDP is primarily dependent on the "2/3" lag of composite PMI. E.g., 2020 Q3 GDP depends only on the Jul-20 PMI observation in our simplified model.

Table 5 shows conditional forecasts for 2020 Q4, 2021 Q1 and 2021 Q2 UK GDP growth using the PMI predictions produced by GNARX for each of the three COVID-19 intervention scenarios, assuming only the 2/3 lag of composite PMI is relevant. Our results suggest a 4.5% difference in 2021 Q1 GDP growth and a 4.8% difference in 2021 Q2 GDP growth between the easing and tightening scenarios. Table 5's bootstrap prediction intervals take into account both the uncertainties in the GDP regression model and the underlying PMI forecasts.

Lending support to the strongly positive GDP growth path in our easing restriction scenario, the most recent (at the time of writing) Bank of England (2021) Monetary

Table 5. Forecasts for UK quarterly real GDP growth (%) based on PMI forecasts for the tightening, constant and easing intervention measure scenarios. 95% prediction intervals in brackets.

Quarter	Tightening	Constant	Easing
2020 Q4	-0.1 (-4.1, 3.8)	-0.1 (-4.1, 3.8)	-0.1 (-4.1, 3.8)
2021 Q1	-1.2 (-5.8, 2.9)	0.3 (-4.1, 4.4)	3.3 (-1.3, 7.4)
2021 Q2	-1.6 (-7.1, 2.2)	-0.1 (-5.7, 3.2)	3.2 (-2.0, 7.1)

Policy Report forecasts a material recovery in UK GDP in the event of an easing in Covid-related restrictions. The authors suggest that the economic impact of such easing would be driven by a substantial increase in consumer spending, and a rise in business investment to a lesser extent as sales rise and uncertainty diminishes.

5.   Discussion

We showed how to apply network time series models to multivariate PMI time series for many countries and how forecast performance could be improved by incorporating network information. We introduced the GNARX model, which permits exogenous variables to be incorporated into GNAR models for the first time and improves forecasting results further still by using COVID-19 intervention stringency indices and COVID-19 death rates. Parameter estimates suggest a highly significant negative impact of harsher intervention measures and COVID-19 deaths on composite PMI and suggests that this effect can be transmitted to neighbours through the export channel.

GNAR(X) model order choice using BIC selects highly parsimonious models that result in excellent forecasting performance, considering the inherent difficulties in forecasting economic indicators. On the other hand, the BIC-chosen VAR model appears to overfit, leading to relatively high forecast errors. Future work might compare the GNAR forecasting performance to that of BVAR models (Giannone et al., 2015).

Our network was constructed by considering what looked 'sensible' rather than optimality with respect to forecasting performance and further development work is required. Some preliminary ideas have been suggested relating to network construction in different fields, see Leeming (2019); Knight et al. (2020) and references therein.

Our three different scenarios earlier all operated under the simplifying assumption that policy interventions would be modified independently of the COVID-19 death rate. Future work may relax this unrealistic assumption, and incorporate other variables to capture attempts by governments to mitigate the negative impact of containment measures through fiscal stimulus.

We have used mean-squared forecasting error throughout our work as it is a valid measure, well-understood and meaningful. There are, of course, many other forecast quality metrics that could be used. For example, the actual proportion of forecasts whose out-turn performed well with respect to (or within) previously established forecasting intervals, see Ehm et al. (2016) for more examples and discussion.

Network time series modelling is young and much remains unexplored. For GNARX models, an obvious extension might permit exogenous regressor coefficients to vary by node. Regarding the COVID-19 experiments discussed above, this would allow

economies to react differently to a given stringency level of policy interventions. Intuitively, economies where a larger proportion of employees are able to work remotely should be less vulnerable to strict lockdown measures (Papanikolaou and Schmidt, 2020).

We are also conducting research where more than one network time series interacts. So, for example, one can already see suggestions in the media on how countries' lockdown policies influence each other. For example, the second UK lockdown on 5th November was made politically easier because of recent lockdowns in major European neighbours. Hence, the stringency index time series (and the COVID-19 deaths) could also be represented by network time series and there seems to be no reason why the PMI and stringency index networks would have to be the same. Finally, further extensions to network time series might be considered, such as time-varying parameters that can be used to create a locally stationary network time series process, or permitting edge-distances to change and be modelled (e.g. as exports vary between countries over time).

Acknowledgements

References

Bańbura, M., Giannone, D., and Lenza, M. (2015) Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections, *Int. J. Forecasting*, **31**, 739–756.

Bank of England (2021) Monetary Policy Report, `https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2021/february/monetary-policy-report-february-2021.pdf`, February 2021.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005) Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, *Quart. J. Econ.*, **120**, 387–422.

D'Agostino, A. and Schnatz, B. (2012) Survey-based nowcasting of US growth: a real-time forecast comparison over more than 40 years, Technical Report 1455, ECB, Frankfurt, Germany.

Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016) Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings, *J. R. Statist. Soc.* B, **78**, 505–562.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004) The MIDAS touch: Mixed data sampling regression models, Technical Report qt9mf223rs, UCLA, Anderson GSM.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015) Prior selection for vector autore-gressions, *Rev. Econ. Stat.*, **97**, 436–451.

Hale, T., Webster, S., Petherick, A., Phillips, T., and Kira, B. (2020) Oxford COVID-19 Government Response Tracker, `https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker`.

International Monetary Fund (2020) World Economic Outlook: A Long and Difficult Ascent, October.

Jansen, M., Nason, G. P., and Silverman, B. W. (2009) Multiscale methods for data on graphs and irregular multidimensional situations, *J. R. Statist. Soc.* B, **71**, 97–125.

Knight, M., Leeming, K., Nason, G., and Nunes, M. (2020) Generalised network autore-gressive processes and the GNAR package, *J. Statist. Soft.*, **96**, 1–36.

Knight, M. I., Nunes, M. A., and Nason, G. P. (2016) Modelling, Detrending and Decor-relation of Network Time Series, *ArXiv e-prints*, **1603.03221**.

Leeming, K. A. (2019) *New Methods in Time Series Analysis: Univariate Testing & Network Autoregression Modelling*, Ph.D. thesis, School of Maths, Univ. Bristol, U.K.

Lütkepohl, H. (2005) *New introduction to multiple time series analysis*, Springer Science & Business Media, Berlin.

MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, *J. Econometrics*, **29**, 305–325.

Nason, G. P. (2013) A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series., *J. R. Statist. Soc.* B, **75**, 879–904.

Nason, G. P. (2016) *locits: Tests of stationarity and localized autocovariance*, R package version 1.7.3.

Papanikolaou, D. and Schmidt, L. D. W. (2020) Working remotely and the supply-side impact of Covid-19, Technical Report 27330, NBER.

Roser, M., Ritchie, H., Ortiz-Ospina, E., and J., H. (2020) Coronavirus pandemic (COVID-19), `https://ourworldindata.org/coronavirus`.

Tsay, R. S. (2013) *Multivariate time series analysis: with R and financial applications*, Wiley, Hoboken, N.J.

Waggoner, D. F. and Zha, T. (1999) Conditional forecasts in dynamic multivariate mod-els, *The Review of Economics and Statistics*, **81**, 639–651.

Zarnowitz, V. (1991) What is a business cycle?, Technical Report 3863, NBER.

Zeileis, A. (2004) Econometric computing with HC and HAC covariance matrix estima-tors, *J. Statist. Soft.*, **11**, Issue 10.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017) Network vector autoregression, *Ann. Statist.*, **45**, 1096–1123.