Received: 12 February 2021 | Revised: 31 January 2023 | Accepted: 1 February 2023

DOI: 10.1002/jae.2980

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Nowcasting from cross-sectionally dependent panels

Jack Fosten[1,2] | **Shaoni Nandi**[1,3]

[1]King's Business School, King's College London, London, UK

[2]Data Analytics in Finance and Macro (DAFM) Research Centre, King's College London, London, UK

[3]Department of Statistics and Information Management, Reserve Bank of India, Mumbai, India

Correspondence
Shaoni Nandi, King's Business School, King's College London, WC2B 4BG London, UK.
Email: shaoni13.34@gmail.com

Summary

This paper builds a mixed-frequency panel data model for nowcasting economic variables across many countries. The model extends the mixed-frequency panel vector autoregression (MF-PVAR) to allow for heterogeneous coefficients and a multifactor error structure to model cross-sectional dependence. We propose a modified common correlated effects (CCE) estimation technique which performs well in simulations. The model is applied in two distinct settings: nowcasting gross domestic product (GDP) growth for a pool of advanced and emerging economies and nowcasting inflation across many European countries. Our method is capable of beating standard benchmark models and can produce updated nowcasts whenever data releases occur in any country in the panel.

KEYWORDS

calendar effects, common correlated effects, cross-sectional dependence, multifactor errors, mixed data sampling (MIDAS), nowcasting

1 | INTRODUCTION

Nowcasting has emerged as an important tool for timely policy-making, particularly by central banks who need to track key variables like gross domestic product (GDP) and inflation in real time. This is important as there is often a delay before the publication of economic data such as these. The main idea is to predict the variable of interest in a timely fashion leading up to its data release, using related available information from other higher-frequency variables. Nowcasting models have typically been developed and applied with single countries in mind using time series methods. On the other hand, this paper builds a panel data nowcasting model when the aim is to produce nowcasts for many countries which may include both developed and emerging economies. The use of panel data can be very important in empirical settings where the number of time series observations is too low for a meaningful forecast evaluation exercise. It has also been argued that the use of pooled panel forecasts can provide accuracy gains over the use of individual time series forecasts (Baltagi, 2008; Wang et al., 2019) which we look to develop in a nowcasting context.

The focus of this paper is to develop tools for simultaneously making nowcasts of economic series for as large a set of countries as possible, while allowing for potential heterogeneity as well as cross-country spillovers. In looking at large sets of different countries, it is often necessary to focus attention on a handful of select predictor variables which are common across all countries, especially when including developing economies. Nevertheless, in our approach, we can exploit the staggered flow of data releases (the 'ragged edge') across countries and across variables in updating our panel nowcasts. This differs from existing empirical nowcasting studies which have exploited the flow of data for a larger set of variables but only for an individual country or a very small number of similar countries (see Cascaldi-Garcia et al., (2021), and references therein). In making nowcasts for individual countries, our approach can also deliver more information than existing studies which have targeted global aggregate variables like GDP (Ferrara & Marsilli, 2019; Kindberg-Hanlon & Sokol, 2018). We can also model the interlinkages across countries which builds on existing work which finds that international variables can improve nowcast accuracy (e.g., Bragoli & Fosten, 2018).

We make three distinct contributions. First, we propose a mixed-frequency panel nowcasting setup which is new to the literature and allows for errors to be dependent over the individuals in the panel. We address the mixed-frequency issue using a mixed data sampling (MIDAS) approach, particularly the unrestricted MIDAS (UMIDAS) model (Foroni et al., 2015). We adapt this model to a potentially heterogeneous and cross-sectionally dependent (CSD) panel framework with a multifactor error structure (Chudik & Pesaran, 2015), while also allowing for different lag structures across countries based on their ragged edge of data availability. Our method allows for full parameter heterogeneity across cross-sectional units, but we can also shut down heterogeneity and pool across the panel dimension which can yield improvements in nowcast accuracy as we display in our empirical application. Among the prevalent mixed-frequency nowcasting methods (see Ghysels, (2018), for a recent review), we focus on MIDAS-type nowcasting models as they have already been extended to a panel framework with encouraging results (Fosten & Greenaway-McGrevy, 2022; Babii et al., 2020). We build upon these studies by further allowing for heterogeneous parameters to reflect diverse macrodynamics and a factor error structure to account for CSD. The resultant model is a panel extension of the observation-driven mixed-frequency vector autoregression (MFVAR) model of Ghysels (2016).

Secondly, we propose a method for obtaining feasible nowcasts given the unknown factor error structure, by suggesting a novel modification of the common correlated effects (CCE) factor estimation technique of Chudik and Pesaran (2015) which allows it to be used for nowcasting. We use a lagged CCE (LCCE) approach which estimates the factors only based on the data available at the time of making the nowcast. This moves away from the original CCE method, developed with the use of contemporaneous variables in estimating the factors, which is widely used in applied causal studies but cannot be used for forecasting or nowcasting applications. The method is simple to implement using least squares estimation and can be adapted to pooled panel least squares in cases where coefficient heterogeneity is not permitted. Simulation studies find that the LCCE method performs well in terms of estimation accuracy and out-of-sample prediction, which motivates its use for estimating panel MIDAS (PMIDAS) nowcasting models with different lag structures determined by the ragged edge.

The third contribution is to apply our method in two distinct empirical settings: nowcasting the real GDP growth of a large set of developed and emerging economies and nowcasting the inflation rate of European countries. In the first application, we construct a panel dataset of more than 30 countries' real GDP as well as some key predictors like business surveys (manufacturing and services) and industrial production. To assess how nowcasts evolve as we add information from across the panel, we perform a pseudo out-of-sample experiment making use of a doubly asynchronous calendar of macroeconomic releases: The data releases are staggered both across variables and across countries. This means that we end up with more nowcast updates than in many studies with single countries or only a few countries. The out-of-sample analysis uses a relatively short initial estimation time span which further motivates the panel approach over time series methods. We make several interesting findings. Firstly, we find that our proposed PMIDAS model performs better than a simple time series autoregressive benchmark when we pool the coefficients across countries and only allow heterogeneity through fixed effects. Secondly, we find that a single business survey variable is able to deliver as good a nowcast as when using one or more other predictors. This is potentially due to their timeliness and providing good economic signal (see also Bańbura et al., 2013; Cascaldi-Garcia et al., 2021). Finally, we find that nowcasts monotonically improve across the panel as we add information across countries and variables. This shows that findings of nowcast monotonicity also hold in the panel data context in a similar way to those seen in the time series nowcasting literature (Aastveit et al., 2014; Fosten & Gutknecht, 2018; Giannone et al., 2008; Marcellino et al., 2016). Our results also hold after investigating their robustness to the choice of evaluation sample and the addition of extra predictor variables.

Our second contrasting empirical application assesses how the PMIDAS model performs in nowcasting monthly inflation across a large set of European countries. We use weekly energy prices to provide a timely signal for tracking movements in inflation as in Modugno (2011, 2013). We therefore offer a new approach by nowcasting a panel of countries' inflation instead of single countries. This study is a useful contrast to the global GDP example as in this case; the nowcast updating does not come from the staggered release of information across countries; it only comes from the higher frequency of the predictor. This demonstrates how our method can be applied in a variety of settings. Our findings mirror those of the GDP application, showing that our proposed method is capable of nowcasting inflation well, beating a benchmark model on average across all countries in the sample.

In relating our paper more widely to the literature, the PMIDAS model we propose brings together two distinct strands of literature: mixed-frequency methods and panel data models with cross-sectional dependence. Mixed-frequency methods are widely used in macroeconometrics with various models and estimation techniques proposed (Kuzin et al., 2011; Schorfheide & Song, 2015; Ghysels, 2016). The literature on panel data methods has also grown significantly over time. In particular, the large heterogeneous panel data model with a CSD multifactor error structure has become an important tool. Pesaran (2006) introduced the CCE method of factor estimation, further developed by Chudik and Pesaran (2015) for dynamic panel models which allows for heterogeneous coefficients, cross-sectional dependence, factor error structure and feedback between target and predictor variables. We bring these aspects together in our mixed-frequency panel nowcasting model with CSD.

This paper also connects two related empirical strands of literature, namely, cross-country macroeconomic forecasting and nowcasting. Intercountry linkages have been admitted in the forecasting literature for the past few decades; see, for instance, Canova and Ciccarelli (2004), Gavin and Theodorou (2005), Garnitz et al. (2019), Chen and Ranciere (2019) for panel data; Chudik et al. (2016) for Global Vector Autoregression (GVAR); and Caselli et al. (2020) for density forecasting. Additionally, the recent empirical nowcasting literature has also recognised the importance of international data. Several studies find that the inclusion of international macrodata improves accuracy, for instance, Schumacher (2010), Eickmeier and Ng (2011), Bragoli and Fosten (2018) and Cepni et al. (2019). Separately, interlinkages have been incorporated into New Keynesian type macroeconomic models, which are now used extensively by policymakers and private institutions for nowcasting as well (Hantzsche et al., 2018). The prevalence of these studies all highlight the importance of using cross-country effects in our panel nowcasting model.

The rest of the paper is organised as follows. Section 2 introduces the main nowcasting model and the estimation technique. The Monte Carlo simulation is presented in Section 3. Sections 4 and 5 display the two different empirical applications to GDP nowcasting and inflation nowcasting. Section 6 concludes the paper. The Appendix contains some of the simulation results and other charts, and there is a separate supporting information document which houses various additional technical details as well as additional simulations and empirical results which are not included in the main paper.

2 | SETUP

In this section, we introduce the PMIDAS setup for panel nowcasting allowing for heterogeneity and CCE, using mixed-frequency data with a ragged edge. We base the model on the dynamic CSD panel data model of Chudik and Pesaran (2015) with crucial modifications for the nowcasting case as we outline below. As the model is based on unknown factors, we then set out how to obtain a feasible model which can be estimated and used for nowcasting.

2.1 | The nowcasting model

We will set up the model using the case of a quarterly target variable with monthly predictors as is the case with real GDP nowcasting. However, as we show in our simulations and empirical illustrations, our setup can easily be generalised to allow for other mixed-frequency combinations such as annual -to-quarterly or monthly-to-weekly. Suppose we have data on the quarterly target variable of interest $y_{i,t}$ for cross-sectional units $i = 1, 2, \dots, N$ and quarters $t = 1, 2, \dots, T$. We also have a vector of k predictor variables measured at a higher monthly frequency which we denote $x_{i,t}^{M}$. We follow Ghysels (2016) and stack the 3 months of quarter t into the following vector for each i:

$$X_{i,t}^{M} = \begin{pmatrix} x_{i,t}^{M} \\ x_{i,t-\frac{1}{3}}^{M} \\ x_{i,t-\frac{2}{3}}^{M} \end{pmatrix}, \tag{1}$$

which will allow us to combine the quarterly and monthly data in a MIDAS-type model. When other frequency combinations are considered, one can modify the notation and the stacked vector in Equation (1) accordingly.

In nowcasting, it is of crucial importance to take account of the ragged edge, in other words using only the recent observations available at the time of making the nowcast, which may differ across individual units, i, and across variables. Suppose we are making a nowcast on day v of the nowcast period.[1] Then we denote d_{iv} to be the latest available quarterly lag of the target variable $y_{i,t}$ for the cross-section i on the v^{th} day of the nowcast quarter.[2] Similarly, we denote m_{iv} as the latest available monthly lag (relative to the last month of quarter t) for $x^M_{i,t}$ for cross-section i on the v^{th} day of the nowcast quarter. The value $m_{iv} = 0$ corresponds to the case where all 3 months of the quarter are available for the predictor variable. In other words, the date v is varied at a daily frequency in order to capture the staggered release of new monthly and quarterly information which can be used to update the nowcasts. We therefore use this notation to allow for a fully asynchronous calendar of data releases across all entities in the cross-section. We also allow for the release to be staggered across the k variables in $x_{i,t}$ though we suppress this additional dependence of the lags on k to avoid notational clutter.

The main nowcasting model we consider uses this lag structure in a PMIDAS model with a multifactor error assumption:

$$y_{i,t} = c_{vi} + \phi_{vi} y_{i,t-d_{iv}} + \beta'_{vi} X_{i,t-\frac{m_{iv}}{3}} + u_{v,i,t}, \tag{2a}$$

$$u_{v,i,t} = \gamma'_{vi} f_t + \varepsilon_{v,i,t}, \tag{2b}$$

where c_{vi} are individual fixed effects, ϕ_{vi} is the coefficient on the autoregressive lag, and in this quarterly-to-monthly example β_{vi} is a $3k \times 1$ vector of individual-specific slope coefficients on the lag of the vector described in Equation (1).[3] The term f_t is an $m \times 1$ vector of unobserved common factors which are used to model the cross-sectional dependence in the error term $u_{v,i,t}$ and has loadings γ_{vi}. The parameters and error terms of the model depend on nowcast date v as the model variables are dependent on the lag structure determined by v. We specify the model with full heterogeneity of coefficients (across i) and note that, even with fully heterogeneous coefficients, the model still retains a panel structure through the assumed error dependence.[4] The model can be modified to have homogeneous coefficients which do not change across i. This would reduce the number of parameters to estimate and can yield forecast accuracy gains in certain scenarios (see Wang et al., 2019). This is something we will consider in our empirical study.

The model in Equation (2a) therefore builds on the original model of Chudik and Pesaran (2015) in two distinct ways. We firstly build in the mixed-frequency aspect which results in the panel equivalent of a UMIDAS model. This choice of model is motivated by Foroni et al. (2015) who conclude that UMIDAS performs better as compared with more complex nonlinear MIDAS models in the case that the difference in frequencies is not too high. The second key difference is the lag structure which is determined by the availability of the data or the ragged edge. In the supporting information, we provide step-by-step detail on how these modifications are made to the original setup of Chudik and Pesaran (2015).

Our model choice is targeted towards situations in which a relatively small number of k predictors are available in making the nowcasts. As mentioned above, this is the main focus of our first empirical application where we aim to have a large coverage of global economies for which only a few common predictors are available for a reasonable time span. Other examples of the applicability of this model include GDP nowcasting at the subnational level where relatively few usable regional predictors are typically available (see, e.g., Fosten & Greenaway-McGrevy, 2022). Our methods therefore align more closely with the small-dimensional nowcasting literature such as bridge and MIDAS models; see Schumacher (2016) for a survey. This is in contrast to studies where a larger number of predictors are available for an individual country or a handful of developed countries (see, e.g., Cascaldi-Garcia et al., 2021) where it has become common to extract factors from those variables. This may soon become applicable in our context once harmonised macro datasets become available for a large range of developed and developing countries. Additionally, the modification of heterogeneous panel data models with cross-sectional dependence to allow for high-dimensional predictors requires theoretical development and is something which we leave for further study.

[1] For more details on the notation used for nowcast updating, see Bańbura et al. (2013).

[2] As an example with two countries $i = 1, 2$, if $d_{1,25} = 1$ and $d_{2,25} = 2$, this implies that on day 25 of the nowcast period, the previous quarter's observation for $y_{i,t}$ has already been released for country 1 but not yet for country 2.

[3] We could, of course, include more lags of $y_{i,t}$ and more than three monthly lags of $x_{i,t}$ by adding additional terms to the right-hand side of Equation (2a). We do not write this down here to save introducing additional notation.

[4] These heterogeneous coefficients are assumed in Chudik and Pesaran (2015) to follow a random coefficient model with i.i.d. errors when they derive the theoretical properties of the model. In practice, we estimate these heterogeneous coefficients using an OLS regression for each cross-sectional unit.

2.2 Estimation and nowcasting

We firstly note that Equations (2a) and (2b) can be combined to write down a model for $y_{i,t}$ as follows:

$$y_{i,t} = c_{vi} + \phi_{vi} y_{i,t-d_{iv}} + \beta'_{vi} X^M_{i,t-\frac{m_{iv}}{3}} + \gamma'_{vi} f_t + \varepsilon_{v,i,t}. \tag{3}$$

However, we do not directly use Equation (3) for nowcasting due to the presence of the unobserved factors f_t which we must estimate. To do so, we propose a lagged version of the CCE estimation technique of Chudik and Pesaran (2015). We specify that the predictor variable $X_{i,t-\frac{m_{iv}}{3}}$ is also influenced by the common factor and lags of $y_{i,t}$:

$$X^M_{i,t-\frac{m_{iv}}{3}} = \kappa_{vi} + \alpha_{vi} y_{i,t-d_{iv}} + \Gamma'_{vi} f_t + \epsilon_{v,i,t}, \tag{4}$$

where, recalling from above, a value of $m_{iv} = 0$ corresponds to the last month of the current quarter. The terms κ_{vi}, α_{vi}, and $\epsilon_{v,i,t}$ are vectors, and Γ_{vi} is a matrix to match the dimensions of $X^M_{i,t}$.

The role of Equation (4) is not for use in nowcasting, as it models the high-frequency variable as a function of the low-frequency variable. Instead, it is used as a device to cast Equation (3) into a VAR form based on a stacked vector:

$$z^M_{i,t,v} = \begin{pmatrix} y_{i,t-d_{iv}} \\ X^M_{i,t-\frac{m_{iv}}{3}} \end{pmatrix}, \tag{5}$$

which resembles the approach of Ghysels (2016) where low-frequency and high-frequency variables are stacked together. In most applications, the primary focus is on the equation for the low-frequency variable which matches what is already done in traditional single-equation MIDAS models.

Once the mixed-frequency and ragged edge are accounted for in the stacked vector in Equation (5), the setup becomes similar to the original Chudik and Pesaran (2015) which also stacks the target variable with a vector of predictors. The full steps of this procedure are given in Section S1 of the supporting information and mirror those of Chudik and Pesaran (2015), which we omit here for the sake of brevity. Intuitively, the steps start by writing down a VAR for $z^M_{i,t,v}$ as a function of the unknown factors f_t. This VAR can be averaged cross-sectionally and inverted to move between the factors themselves and cross-sectional averages of the $z^M_{i,t,v}$ variable. We show how this is done below, after making some important comments about $z^M_{i,t,v}$.

We note that the vector $z^M_{i,t,v}$ in Equation (5) resembles the vector $z_{i,t}$ used in the CCE estimation method of Chudik and Pesaran (2015) except for a very important difference. In our case, $z^M_{i,t,v}$ only includes the lags of the target and predictor variables which are actually available at nowcast date v. This means that $z^M_{i,t,v}$ can be used to estimate the factors in a way which is feasible on the day the nowcast is made. The original paper of Chudik and Pesaran (2015) used contemporaneous $y_{i,t}$ and $x_{i,t}$ variables in estimating the factors which is not suitable for prediction. We therefore refer to this as LCCE estimation.

In obtaining the feasible nowcasting model for $y_{i,t}$, we use a cross-sectional (weighted) average of $z^M_{i,t,v}$ using a weight vector $w = (\omega_1, \omega_2, \ldots, \omega_N)'$. We define the cross-sectionally weighted average of Equation (5) as follows:

$$\bar{z}^M_{t,v} = \sum_{i=1}^{N} \omega_i z^M_{i,t,v}, \tag{6}$$

and we can use the following representation of Equation (3):

$$y_{i,t} = c^*_{vi} + \phi_{vi} y_{i,t-d_{iv}} + \beta'_{vi} X^M_{i,t-\frac{m_{iv}}{3}} + \delta'_{vi}(L) \bar{z}^M_{t,v} + \varepsilon_{v,i,t} + O_p(N^{-1/2}), \tag{7}$$

where $\delta_{vi}(L)$ is an infinite-order lag polynomial with a form depending on the parameters of Equations (3) and (4) above, and the $O_p(N^{-1/2})$ term is an asymptotically negligible remainder term resulting from using $\bar{z}^M_{t,v}$ in place of the factors.[5]

The final feasible nowcasting equation is based on a finite approximation of the infinite number of lags of $\bar{z}^M_{t,v}$ used in Equation (7) as follows:

$$y_{i,t} = c^*_{vi} + \phi_{vi} y_{i,t-d_{iv}} + \beta'_{vi} X^M_{i,t-\frac{m_{iv}}{3}} + \sum_{l=0}^{p_T} \delta'_{vil} \bar{z}^M_{t-l,v} + e_{v,i,t}, \tag{8}$$

where the overall error term $e_{v,i,t}$ in this feasible nowcasting equation contains the approximation from the lag truncation as well as from replacing the factors with $\bar{z}^M_{t,v}$.[6] The choice of the lag truncation is suggested to be $p_T = T^{1/3}$ by Chudik and Pesaran (2015).

We finally have a model for $y_{i,t}$ which is linear in variables which are available on day v of the nowcast period. The model can be estimated by OLS, and nowcasts can be feasibly obtained using the estimated coefficients and the latest available data. In the most general model described above with full parameter heterogeneity, OLS estimation amounts to performing one regression per cross-sectional unit.[7] However, researchers may wish to restrict the amount of allowed heterogeneity, in which case the model can be estimated by pooled panel OLS. We explore this in the empirical applications where we first obtain results under full heterogeneity using equation-by-equation OLS and then we shut down all heterogeneity except for individual-specific constants and use pooled OLS for estimation.

We can use the OLS parameter estimates to obtain a nowcast of quarter T for every cross-sectional unit i on day v of the nowcast period by estimating the conditional mean of $y_{i,t}$ given all available information on day v as follows:

$$\hat{y}_{i,T,v} = \hat{c}^*_{vi} + \hat{\phi}_{vi} y_{i,T-d_{iv}} + \hat{\beta}'_{vi} X^M_{i,T-\frac{m_{iv}}{3}} + \sum_{l=0}^{p_T} \hat{\delta}'_{vil} \bar{z}^M_{T-l,v}, \tag{9}$$

where, as written above, if one decides to impose some homogeneity on the coefficients and pool the model, then pooled panel OLS estimates can be used in Equation (9) to obtain the nowcasts.

3 | MONTE CARLO SIMULATIONS

In this section, we carry out Monte Carlo simulations using the model described above to assess the performance of our LCCE estimation strategy where we modify the CCE estimation approach of Chudik and Pesaran (2015) for use in nowcasting. Our simulations are based on the model described in Equations (3) and (4) above. However, for simplicity and tractability in the simulations, we ignore the presence of the ragged edge in the data and assume there is only a single nowcast date for which the available lags are $d_{iv} = 1$ and $m_{iv} = 0$. In other words, we assume that the previous lag is available for the target variable along with the current period for the higher-frequency predictor. We will maintain the quarterly-to-monthly frequency mix (we denote the ratio of high to low frequency as $q = 3$) in the baseline simulations, but we will also check how the results hold with a frequency mix of $q = 4$ which could represent an annual-to-quarterly or monthly-to-weekly frequency mix.

3.1 | Setup

We generate a panel dataset of dimensions $N \times T$ from model Equations (3) and (4) with $d_{iv} = 1$ and $m_{iv} = 0$.[8] We choose the parameter values to make the simulated series resemble macroeconomic growth rates, as well as being guided

[5]We note that the main interest of Chudik and Pesaran (2015) is in demonstrating the equivalence of Equations (3) and (8), and they are not *per se* concerned with consistency in estimating the 'true' factors as in studies like Bai and Ng (2002) and Stock and Watson (2002).

[6]The use of $\bar{z}^M_{t,v}$ in Equation (8) bears resemblance to 'factor-augmented' type models, where in our case the factors are estimated across countries.

[7]In the full heterogeneity case when the number of lags is large, or if we use very high-frequency data for $x_{i,t}$, one may consider using a shrinkage estimator like ridge or LASSO in obtaining the nowcasts.

[8]Equations (3) and (4) with $d_{iv} = 1$ and $m_{iv} = 0$ in fact match Equations S1a, S1b and S1c in the supporting information where we abstract from the ragged edge.

by the parameter choices used by Chudik and Pesaran (2015).[9] We fix the number of regressors to be $k = 1$ and let the regression coefficients in β_i be i.i.d. $U[0, 0.4]$ across i. The fixed-effects terms c_i and κ_i are assumed to be i.i.d. $U[-1, 1]$. The values of ϕ_i and α_i are chosen as i.i.d. $U[0, 0.4]$ and i.i.d. $U[0, 0.25]$, respectively. The error component $\varepsilon_{i,t}$ in Equation (3) (now without a v subscript due to the absence of the ragged edge in the simulations) is generated as i.i.d. $N(0, 1)$, and the unobserved common factors f_t are generated as independent stationary AR(1) processes as below:

$$f_{tl} = \rho_f f_{t-1,l} + \xi_{tl}, \tag{10}$$

where $\rho_f = 0.25$, $\xi_{tl} \sim N(0, 1)$ for $l = 1, 2, \ldots, m$ and $t = 1, 2, \ldots, T$ and we consider $m = 2$ factors here. The loadings on both factors are generated as i.i.d. $N(0.25, 0.1)$ for y and i.i.d. $N(-1, 0.1)$ and i.i.d. $N(1, 0.1)$ for x for the first and second factors, respectively. The error component $\epsilon_{i,t}$ in Equation (4) is also generated as a stationary AR(1) process as follows:

$$\epsilon_{i,t} = \rho_\varepsilon \epsilon_{i,t-\frac{1}{q}} + \eta_{i,t}, \tag{11}$$

where $\rho_\epsilon = 0.22$ and $\eta_{i,t}$ is i.i.d. $N(0, 1)$. We use an initial 100 observations as burn-in. The value of p_T, the lag truncation parameter, is selected at $T^{\frac{1}{3}}$, as recommended by Chudik and Pesaran (2015). We will look at results over a variety of sample sizes $N, T \in \{50{,}100, 150{,}200\}$. We let M denote the number of Monte Carlo replications, which is set at $M = 1000$.

We focus on two different aspects of results based on the above data generating process. We firstly assess the estimation of the parameters in the main nowcasting equation (in other words ϕ_i and β_i), where we compare our LCCE method to the original CCE method where contemporaneous $y_{i,t}$ is used in estimating the factors. This is to verify that we do not lose a lot of estimation accuracy by using LCCE rather than CCE, which we have to do in order to make nowcasting feasible. Since we allow heterogeneity in the parameters across i, we first define the following mean group parameters:

$$\phi = \frac{1}{N}\sum_{i=1}^{N}\phi_i,\ \beta^{(j)} = \frac{1}{N}\sum_{i=1}^{N}\beta_i^{(j)}, \tag{12}$$

where the superscript j on $\beta_i^{(j)}$ indexes the element of the vector β_i and we recall that the parameters no longer depend on v as we abstract from the ragged edge here. We will analyse the average (over the replications) absolute deviation of the estimated mean group parameters from the actual parameter value. The use of the absolute bias is slightly different to the criterion used in the literature on CCE estimation (see Pesaran, 2006, and Chudik & Pesaran, 2015), where the actual value of the bias is used and the exact sign is analysed. However, in this paper, the main focus is on nowcasting, so the forecast efficiency is of primary interest and the sign of any bias is not important.

The second aspect of the results we focus on is the out-of-sample performance of the PMIDAS model. In this regard, we wish to see if the additional complexity of the PMIDAS model (in terms of parameters and factors to estimate) introduces unwarranted forecast uncertainty over a simpler benchmark time series AR(1) model which may also be a good approximation for the serially dependent data we generate. To analyse out-of-sample performance, we split the dataset into two parts in the time dimension, the first being used for model estimation and the latter for out-of-sample forecasting. Let R and P denote the window length for estimation and evaluation samples, respectively. The entire time period covered by the panel is therefore split as $T = R + P$. We use a recursive window starting with R estimation observations, producing the nowcast and then increasing the estimation window by one observation at a time (see West, 1996). As in Hansen and Timmermann (2012), the split point can sometimes affect the out-of-sample results. To mitigate the issue, three different splits are considered: $P = 0.2T$, $P = 0.3T$ and $P = 0.5T$. Thus, we ensure that our forecast evaluation results are not dependent on the choice of split points.

The measure of accuracy we use is the root mean squared forecast error (RMSFE) as defined below:

$$RMSFE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{P}\sum_{t=R}^{T}\hat{e}_{i,t}^2}, \tag{13}$$

[9]We are grateful to Dr. Alexander Chudik for providing us with the Matlab replication codes.

TABLE 1 Simulation results—Absolute bias in LCCE and CCE ($q = 3$).

N/T	CCE 50	CCE 100	CCE 150	CCE 200	LCCE 50	LCCE 100	LCCE 150	LCCE 200
				ϕ				
50	0.0589	0.0266	0.0179	0.0127	0.0519	0.0243	0.0167	0.0118
100	0.0607	0.0255	0.0168	0.0120	0.0534	0.0238	0.0155	0.0112
150	0.0612	0.0272	0.0172	0.0122	0.0548	0.0251	0.0161	0.0114
200	0.0598	0.0265	0.0173	0.0121	0.0533	0.0245	0.0162	0.0113
				$\beta^{(0)}$				
50	0.0217	0.0134	0.0106	0.0085	0.0215	0.0148	0.0128	0.0109
100	0.0155	0.0093	0.0072	0.0063	0.0159	0.0108	0.0087	0.0081
150	0.0128	0.0073	0.0061	0.0049	0.0126	0.0082	0.0076	0.0062
200	0.0111	0.0067	0.0054	0.0043	0.0113	0.0072	0.0064	0.0056
				$\beta^{(1)}$				
50	0.0229	0.0136	0.0108	0.0092	0.0227	0.0156	0.0132	0.0114
100	0.0160	0.0096	0.0075	0.0065	0.0161	0.0107	0.0092	0.0082
150	0.0133	0.0080	0.0061	0.0051	0.0135	0.0090	0.0074	0.0068
200	0.0121	0.0067	0.0053	0.0045	0.0121	0.0074	0.0065	0.0058
				$\beta^{(2)}$				
50	0.0226	0.0136	0.0108	0.0087	0.0227	0.0148	0.0133	0.0113
100	0.0164	0.0096	0.0077	0.0063	0.0166	0.0109	0.0092	0.0083
150	0.0129	0.0075	0.0061	0.0051	0.0129	0.0086	0.0072	0.0065
200	0.0109	0.0069	0.0054	0.0044	0.0111	0.0078	0.0065	0.0057

Note: The numbers in this table are the absolute biases in the estimates of the key model parameters estimated using two methods, LCCE and CCE, across different sample sizes. Abbreviations: CCE, common correlated effects; LCCE, lagged CCE.

where the forecast error $\hat{e}_{i,t}$ is the difference between $y_{i,t}$ and the forecast defined in Equation (9). The statistic in Equation (13) used in this panel setting gives us a single statistic for the average RMSFE across all individuals in the panel.

3.2 | Results

Table 1 summarises the results for the absolute bias for the two estimation techniques under consideration. The figures represent the mean absolute bias. The panels from top to bottom summarise the results for ϕ followed by the individual parameters in the vector β. The results show that absolute biases in both the CCE and LCCE estimates are very small and diminish further towards zero at higher panel dimensions. For smaller panels, the bias in the autoregressive parameter ϕ is marginally higher in both CCE and LCCE when compared with those of the β coefficients.[10]

The difference in bias from the two estimation methods is negligible for all panel dimensions and converges to zero for larger panels. This is confirmed graphically by Figures 8 to 11 in the Appendix which depict the distribution of the difference in the bias between the two methods, showing that it vanishes to zero with the panel size. Overall, this means that the modification to use LCCE estimation does not have a substantial impact on the parameter estimates in the model, while the advantage of the lag structure we deploy in LCCE is that it can be used for forecasting and nowcasting. The bias results remain similar when we move from the frequency mix $q = 3$ to $q = 4$ which can be seen from Table S1 and Figures S1 to S5 in the supporting information.

Turning now to the forecast performance of the PMIDAS model estimated by LCCE. Here our aim is to verify that the estimation of the additional factors and parameters in our panel nowcasting model does not harm forecast performance relative to a smaller naïve time series AR(1) model in this simulated setup. The results are summarised in Table 2 which displays the RMSFE of the PMIDAS model relative to the time series AR(1) benchmark. Figures less than one indicate superior forecast performance of the PMIDAS model. Here we present results for the frequency mix $q = 3$ which represents the most common scenario of GDP nowcasting, as in our first empirical application, where the objective is to nowcast quarterly GDP using monthly information. However, we also have results for $q = 4$ in the supporting information, which is the frequency mix we use in our second empirical application on monthly inflation nowcasting using weekly data.

[10]We note that the bias in ϕ does not improve substantially with N, only with T, which mirrors the findings of Chudik and Pesaran (2015).

TABLE 2 Simulation results—forecast accuracy of PMIDAS relative to AR(1) benchmark—$q = 3$.

	20 %				30 %				50 %			
N/T	**50**	**100**	**150**	**200**	**50**	**100**	**150**	**200**	**50**	**100**	**150**	**200**
50	0.9143	0.8156	0.7877	0.7688	0.9442	0.8249	0.7884	0.7738	0.9864	0.8318	0.7931	0.7773
100	0.9172	0.8144	0.7854	0.7687	0.9446	0.8248	0.7875	0.7718	0.9875	0.8328	0.7922	0.7762
150	0.9216	0.8158	0.7878	0.7693	0.9489	0.8250	0.7886	0.7733	0.9911	0.8310	0.7925	0.7769
200	0.9166	0.8135	0.7866	0.7673	0.9445	0.8231	0.7883	0.7715	0.9882	0.8305	0.7926	0.7763

Note: This table represents the relative root mean squared forecast error (RMSFE) of panel mixed data sampling (PMIDAS) with mixed-frequency data, resembling a monthly-to-quarterly frequency mix. Figures less than one indicate superior performance of the PMIDAS model as compared with a time series AR(1) model. Results are shown for sample split ratios for P equal to 20%, 30% and 50% of the total sample.
Abbreviation: PMIDAS, panel mixed data sampling.

The main conclusion is that the out-of-sample RMSFE of our PMIDAS model is clearly lower than that of the time series AR(1) benchmark across sample sizes and sample splits. We also see that the gain against the benchmark grows with the sample size T, even in fairly modest sample sizes. This indicates that the estimation of the factors and additional parameters does not harm the predictions relative to a simple time series AR(1) model which might be considered a good approximation in settings such as these with serial dependence. The findings remain very similar across the different sample splits we consider, with the exception of the lowest sample size $T = 50$ and a sample split of 50% where the AR(1) has similar RMSFE, which is due to the very low number of in-sample periods used for model estimation. We also show that the results are similar when we change the frequency mix from $q = 3$ to $q = 4$ which can be seen in Table S2 in the supporting information.

4 | EMPIRICAL APPLICATION I: GLOBAL GDP NOWCASTING

In this section, we apply the panel nowcasting techniques developed in this paper to predict the quarterly real GDP growth rate of a large panel of advanced and emerging economies using timelier monthly economic activity. The main advantage of our study relative to the existing literature is that we make nowcasts for a large number of individual countries and not just of aggregate global GDP as in studies such as Ferrara and Marsilli (2019). Our study also looks at a wider spread of countries than the related study of Cascaldi-Garcia et al. (2021) which focusses on a handful of European countries.

4.1 | Data and setup

4.1.1 | Data

The target variable is the annual (year-on-year, y-o-y) growth rate of quarterly GDP in constant national prices. The y-o-y growth rate is widely used by many policymakers both in developed and emerging economies and is also useful for those countries which do not report seasonally adjusted quarterly figures. However, we will also compare our results when using quarter-on-quarter (q-o-q) growth rates which are widely used in academic studies. We predict real GDP growth using a business survey (manufacturing) index in our baseline model and will also explore the results using various combinations of other predictors including business survey (services) and an industrial production index. The focus on survey indicators is important due to their timeliness in capturing near-term economic outlook. These types of predictors are commonly used in existing nowcasting studies (e.g., Marcellino & Schumacher, (2010) and Schumacher (2016) for MIDAS and bridge equation models and Giannone et al. (2008), for dynamic factor models). These particular series are also chosen for their availability for a large number of countries.

The dataset is sourced from the OECD Main Economic Indicators (MEI), covering 37 member countries and selected nonmember partners. We consider the final vintage for the historical data as the real time data are not available for most of the countries we consider. The dataset covers a large share of global GDP, with member countries accounting for almost 50% (OECD, 2020) and includes some emerging economies with large global GDP shares such as India and China. The list of countries included in our sample can be found in Table 3.

There is some variation in the availability of data series across countries. The balanced panel database for GDP with manufacturing business surveys starts from January 2001 and ends in March 2020, which totals $T = 77$ quarters, that is, 231 months of data. There are 34 available countries consisting of 23 advanced economies (AEs) and 11 emerging market

TABLE 3 Country coverage.

Variable	N	Country names
Business surveys - manufacturing	34	Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States, Brazil, Estonia, India, Israel, Slovenia, Latvia, Lithuania
Business surveys - services	21	Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Portugal, Spain, Sweden, United Kingdom, Estonia, Slovenia, Latvia, Lithuania
Industrial production	35	Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Turkey, United Kingdom, United States, Brazil, Chile, Estonia, India, Israel, Slovenia, Latvia, Lithuania, Costa Rica

Note: The table lists the countries taken in the sample for each predictor variable.

economies (EMEs).[11] The business survey for services, however, is only available for 21 European economies (including the UK) which starts in 2003 as this enables the inclusion of 8 additional countries into the panel. Other components of the business surveys have even lower availability.[12] For industrial production, there are 35 countries, with 22 AEs and 13 EMEs. After examining all series for stationarity, the business survey variables are left in levels and industrial production is transformed using growth rates.[13]

Figure 1 summarises the distribution in GDP growth and the monthly predictors across the OECD countries for the last two decades. There is evidence of a broad common time pattern as well as some variation across countries. The dispersion of growth rates among countries increased during the global financial crisis years (2008-2010) and continued for some time. This was followed by a period of very low variation in growth among countries. For the monthly predictors, industrial production and business surveys also, we notice a dispersion within the countries along with a broadly common time path. These monthly variables seem to track the time path of GDP, which reinforces their suitability as GDP predictors.

4.1.2 | Pseudo out-of-sample setup

To evaluate the performance of the PMIDAS model using real GDP and the monthly predictors detailed above, we perform a pseudo out-of-sample experiment using a recursive estimation scheme. As in the simulation section above, we split the full sample into $T = R + P$, where the evaluation window, P, is set to be equal to $P = 0.3T$ so that 30% of the available sample is retained for evaluation. For each quarter in the evaluation sample, the nowcasts are computed for each day during a time window of 155 days from the start of the nowcast quarter. Consequently, this includes backcasting from the 91st day of the quarter onward. By the end of the window of 155 days, official GDP figures are available for the majority of the countries under consideration.

We construct a pseudo calendar to track the releases for all variables in the dataset. For the GDP and industrial production variables, this is constructed by replication of the average release day in the four quarters of 2018. Similarly, for the survey data, the approximate release date was analysed at the end of the sample and replicated for all of the years in the evaluation window. All months are assumed to have 30 days, and accordingly, the quarters consist of 90 days uniformly. Figure 12a–c in the Appendix present the average lags considered for GDP, business survey manufacturing and industrial production, respectively. The services survey data are assumed to be available uniformly on day 21 of the previous month.

[11] The classification of AE and EME is as per IMF (2021).

[12] It is challenging to obtain other high-frequency indicators harmonised across countries that are available in a timely fashion. Other variables such as building permits and commercial vehicle sales were explored. However, there are significant publication lags of more than a year in many countries and so we focus our attention on the series mentioned above.

[13] Until December 2018, the OECD used to publish seasonally adjusted figures for all series and countries. Most of the series are taken directly into the final dataset, as they are seasonally adjusted from the source entirely. Industrial production data for India and Chile are adjusted using the X-13 algorithm.

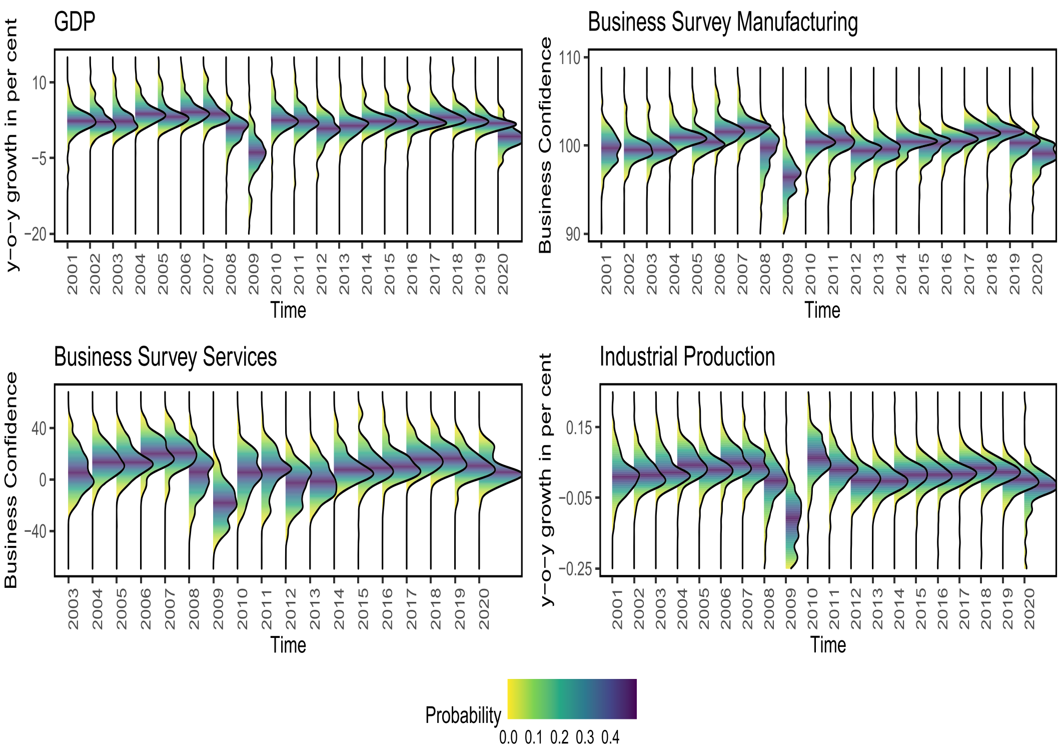


FIGURE 1 Cross-country distribution of real gross domestic product (GDP) growth and predictors across time.
Note: For each year on the horizontal axis, the cross-country distribution is displayed, with colours shaded from yellow to blue in order of low to high occurrence probability.

At each period in the nowcast evaluation exercise, we first use the pseudo calendar to assess which lags are available for every country. This determines the exact lag specification of the nowcasting model in Equation (8).[14] Then the model is estimated, and the nowcasts are obtained for every country. This is done for each of the 155 days of the nowcast and backcast period, for each of the P evaluation quarters we consider. For every day v in the nowcast period, we can obtain the RMSFE by individual country i as follows:

$$RMSFE_{v,i} = \sqrt{\frac{1}{P}\sum_{t=R}^{T}\hat{e}^2_{v,i,t}}, \tag{14}$$

where $\hat{e}_{v,i,t}$ denotes the prediction error from the PMIDAS model on nowcast day v for country i in quarter t (with a similar statistic being used for the benchmark model). We will assess the distribution of these individual RMSFEs as well as using the average across all countries in a similar way to Equation (13) from the simulation section.

4.2 | Main results

In this section, we present the nowcast performance of the PMIDAS model for both the y-o-y and q-o-q target variables using the business surveys manufacturing variable as the baseline case. Although we allow for possibly heterogeneous coefficients in the nowcasting model in Equation (8), an interesting empirical question is whether pooling can produce better nowcasts. We therefore also produce results where we impose homogeneity on all of the slope coefficients in

[14]The number of lags is also as in Equation (8) where we do not consider further lags of $y_{i,t-d_{iv}}$ and $X^M_{i,t-\frac{m_{iv}}{3}}$ and where the lag truncation of the factors is $p_T = T^{1/3}$ as detailed above. Although some methods have proposed panel forecast lag selection methods in the presence of fixed effects nuisance parameters (Lee & Phillips, 2015) and cross-sectional dependence (Greenaway-McGrevy, 2019), these are not applicable in the current context with potential parameter heterogeneity and factors. In previous versions of the paper, we also experimented with the use of machine learning methods like LASSO and the elastic net in order to perform shrinkage and lag selection, motivated by other studies using this in the MIDAS context (Babii et al., 2020; Siliverstovs, 2017; Xu et al., 2018).

TABLE 4 GDP nowcast RMSFE by quantile—year-on-year.

	PMIDAS						Benchmark		
	Pooled			Not Pooled			TS AR		
Days	25%	50%	75%	25%	50%	75%	25%	50%	75%
1	0.7850	1.0568	1.3127	1.0785	1.3658	1.6879	0.9865	1.1307	1.4636
16	0.7956	1.0440	1.3373	1.0785	1.3658	1.6879	0.9865	1.1307	1.4636
31	0.7265	1.0124	1.3647	1.0657	1.3525	1.5683	0.9082	1.0642	1.3018
46	0.6570	0.9127	1.1834	1.0088	1.2610	1.4927	0.8866	1.0487	1.2867
61	0.6146	0.8146	1.0767	0.9293	1.1573	1.4804	0.8115	0.9249	1.2867
76	0.5597	0.7427	1.0269	0.7460	0.9618	1.2776	0.7627	0.9077	1.2550
91	0.5512	0.7323	1.0013	0.7376	1.0246	1.2659	0.7327	0.8952	1.2550
106	0.5443	0.7293	1.0273	0.7376	1.0246	1.2659	0.7327	0.8952	1.2550
121	0.4043	0.6249	0.9947	0.4600	0.7054	1.0614	0.4745	0.7627	1.0487
136	0.0000	0.0000	0.7990	0.0000	0.0000	1.0470	0.0000	0.0000	0.8324
151	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: The table reports the quantiles across countries of RMSFE for each method. We display two nowcasts per month of the nowcast period. The single predictor variable used is the business surveys manufacturing indicator. The RMSFE drops to zero after a country's GDP data are released, so all displayed quantiles have a value of zero on day 151 of the nowcast period.
Abbreviations: GDP, gross domestic product; PMIDAS, panel mixed data sampling; RMSFE, root mean squared forecast error.

TABLE 5 GDP nowcast RMSFE by quantile–quarter-on-quarter.

	PMIDAS						Benchmark		
	Pooled			Not Pooled			TS AR		
Days	25%	50%	75%	25%	50%	75%	25%	50%	75%
1	0.4310	0.5448	0.7014	0.7127	0.8975	1.2539	0.5950	0.7355	1.0317
16	0.4386	0.5425	0.7223	0.7127	0.8975	1.2539	0.5950	0.7355	1.0317
31	0.4457	0.5446	0.7284	0.7317	0.8897	1.2474	0.5950	0.7323	1.0317
46	0.3966	0.5494	0.6674	0.6824	0.8863	1.2466	0.5885	0.7323	0.9823
61	0.4006	0.5463	0.6619	0.7200	0.9313	1.2460	0.5885	0.7394	0.9823
76	0.3816	0.5416	0.6728	0.6466	0.7207	1.1244	0.5885	0.7416	0.9823
91	0.4215	0.5777	0.7000	0.6327	0.7477	1.0849	0.5885	0.7416	0.9823
106	0.4239	0.5713	0.7007	0.6327	0.7477	1.0849	0.5885	0.7416	0.9823
121	0.3166	0.5137	0.6673	0.3224	0.6433	0.8061	0.3797	0.6138	0.8046
136	0.0000	0.0000	0.5729	0.0000	0.0000	0.7314	0.0000	0.0000	0.7416
151	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: See notes for Table 4, above.

Equation (8) while we still have individual-specific fixed effects to allow heterogeneity. We will compare the results of these methods to the time series AR(1) benchmark model. As mentioned above, we compute the nowcasts on a daily basis for 155 days from the beginning of the nowcast period, and we will track how the model performance changes as we add new information.

The main results using the business survey manufacturing data are presented in Tables 4 and 5 which display the y-o-y and q-o-q results. The numbers represent the quantiles (across countries) of the RMSFE of each model on different nowcast days. Starting with the y-o-y results in Table 4, the most striking finding is that the model which has uniformly lowest RMSFE across all nowcast dates is the proposed PMIDAS model when pooling is used with equal slopes across countries. In particular, we see gains relative to the time series AR model, which holds across the 25%, 50% and 75% quantiles. On the other hand, the fully heterogeneous model does not perform as well as the AR model. This indicates that the question of 'to pool or not to pool', as put by Wang et al. (2019), is that nowcast performance is improved when pooling across this sample of global economies. This result is mirrored when looking at the q-o-q results in Table 5 where we see that the best method across nowcast days and quantiles is the pooled PMIDAS model, whereas allowing full parameter heterogeneity leads to worsening even relative to an AR benchmark. Overall, these results lend evidence in favour of the mixed-frequency panel data approach, in a similar way to findings of Fosten and Greenaway-McGrevy (2022) and Babii et al. (2020) although with different applications. The findings are also in favour of the use of panel models for forecasting in general, see Baltagi (2008), where in this case the panel dimension is especially useful when the model is pooled across countries.

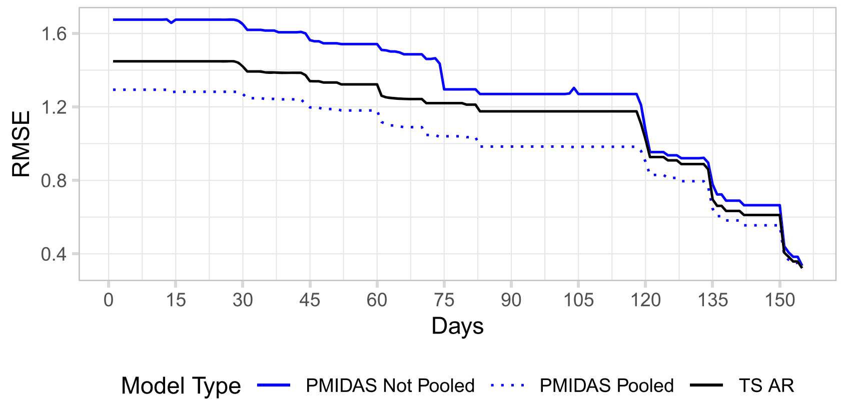


FIGURE 2 Gross domestic product (GDP) nowcast average root mean squared forecast error (RMSFE)—year-on-year.
Note: The figure plots the mean of RMSFE across all countries on each of the 155 days in the nowcast period. The single predictor variable used is the business surveys manufacturing indicator.

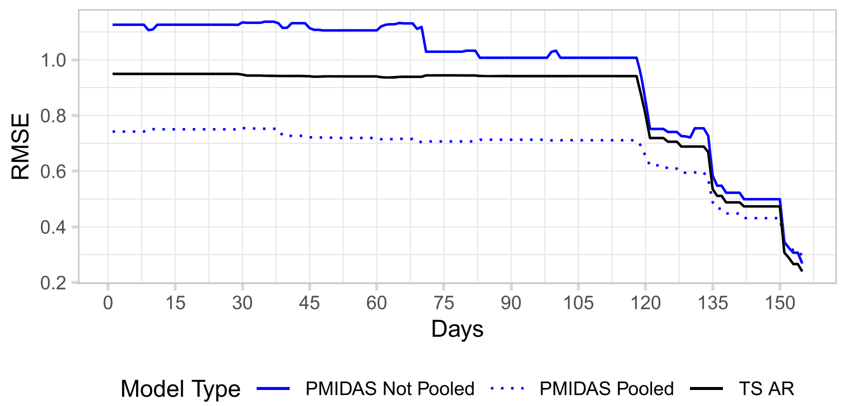


FIGURE 3 Gross domestic product (GDP) nowcast average root mean squared forecast error (RMSFE) - quarter-on-quarter.
Note: See notes for Figure 2.

In order to visualise the behaviour of the RMSFE across all days of the nowcast period, Figures 2 and 3 plot the mean of the RMSFE across all countries for each of the 155 nowcast days, for the same models as in the table above. These results confirm that, on average, the pooled version of the PMIDAS model has better performance than the AR model and the version of the model with fully heterogeneous coefficients. Importantly, the plots help to reveal how the methods behave as we sequentially add more information across countries and variables. Indeed, from Figure 2, it seems that the average RMSFE for the y-o-y GDP growth target is monotonically falling as we add information. Therefore, this panel model, like with the time series studies mentioned above, improves as we take into account more information as it becomes available during and beyond the nowcast quarter. For the q-o-q target, the results are slightly weaker, showing only moderate improvements in the pooled PMIDAS approach at the beginning of the nowcast period before RMSFE flattens until near the end of the nowcast period.

4.3 | Further results

The results in the previous section focussed on the PMIDAS model when using the business surveys manufacturing variable as the sole predictor. It is important to assess how the PMIDAS model performs when changing to use other monthly predictors, and to allow for multiple predictor variables. To do this, we employ the two other variables mentioned above: business surveys services and industrial production. In order to obtain a common sample across all variables, we have to drop the number of countries to 19 so the results are based on a smaller sample than those in the previous section. We obtain results for the single-variable models as well as the other combinations of two and three variables. Figures 4 and 5 display the RMSFE for the pooled PMIDAS model across the various combinations of variables in the model. The results show that, while the three single-variable models all perform quite similarly, as soon as the number of

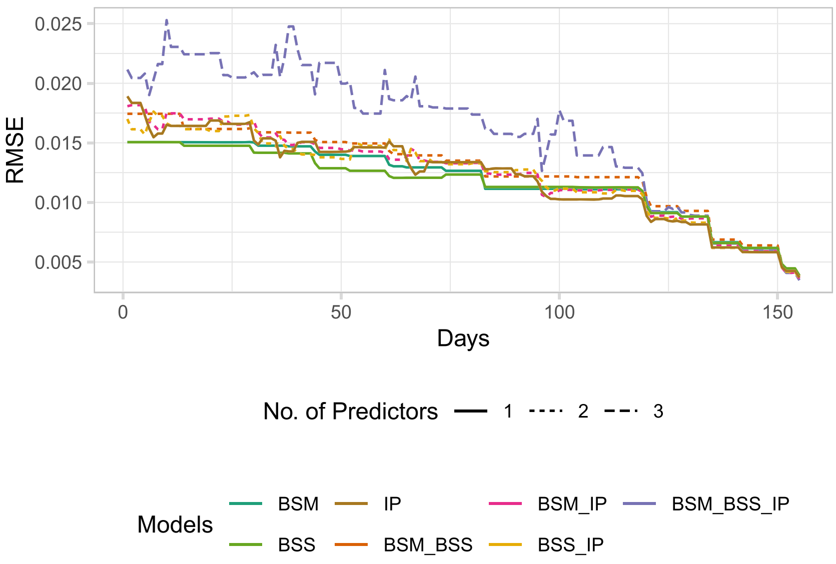


FIGURE 4 Gross domestic product (GDP) nowcast average root mean squared forecast error (RMSFE)—year-on-year— Additional predictors.
Note: The figure plots the mean of RMSFE across all countries on each of the 155 days in the nowcast period. The pooled panel mixed data sampling (PMIDAS) model results are displayed for various combinations of business surveys manufacturing and services ('BSM' and 'BSS') and industrial production ('IP').

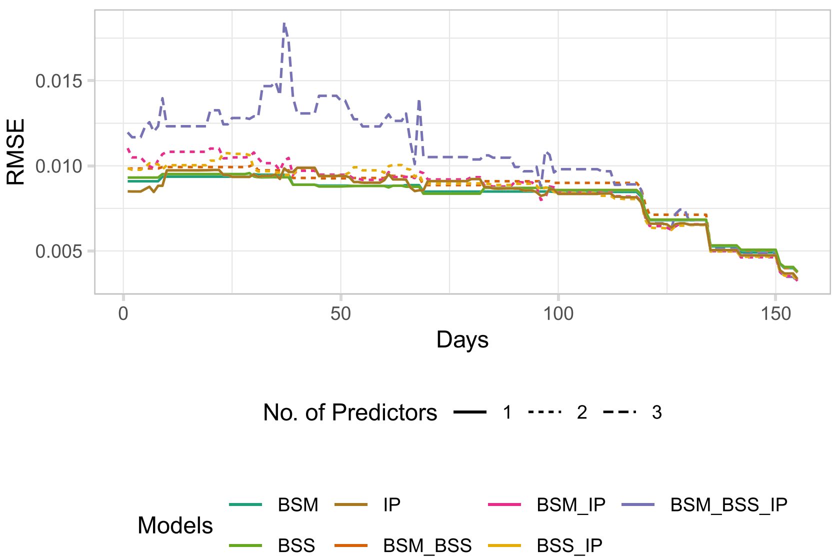


FIGURE 5 Gross domestic product (GDP) nowcast average RMSFE—quarter-on-quarter —Additional predictors.
Note: See notes for Figure 4.

variables included in the model increases, the nowcast performance worsens.[15] This is likely due to the additional burden of parameter estimation.

To further explore the robustness of the results to our chosen set of predictor variables, we also experimented with adding further surveys available in the OECD MEI database (additional business surveys from retail, trade and construction as well as a consumer confidence survey). This further reduces the number of countries in the sample to 18. However, in searching over many possible combinations of models up to three variables, we find that it is always the case that RMSFE is increasing in the number of variables. The results can be found in Figures S6 and S7 in the supporting information. We therefore conclude that a single well-chosen predictor variable can dominate larger models in this PMIDAS context.[16]

We also explore the robustness to the choice of sample split in terms of the in-sample and out-of-sample observations. The results in Figures S8 to S11 in the supporting information show that the results are qualitatively similar when we vary the sample split from $P = 0.2T$ to $P = 0.4T$. While the magnitude of the RMSFE changes slightly as we alter the

[15]When looking at individual countries there is some evidence that the performance does change with different predictors. The fact that there is some difference in accuracy gains of MIDAS models when using different variables has also been documented in earlier studies (see Clements & Galvão, 2008, 2009; Foroni et al., 2015; Schumacher, 2016).

[16]An alternative way to include additional predictors would be to use a dynamic factor model as in Cascaldi-Garcia et al. (2021) which we do not explore in this paper.

TABLE 6 Inflation nowcast RMSFE by quantile—year-on-year.

	PMIDAS						Benchmark		
	Pooled			Not pooled			TS AR		
Days	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
1	0.3684	0.3920	0.4424	0.3657	0.4912	0.6628	0.3552	0.4272	0.5734
7	0.3436	0.3841	0.4428	0.3602	0.4789	0.6702	0.3552	0.4272	0.5734
14	0.3090	0.3706	0.4198	0.3699	0.4540	0.6094	0.3552	0.4272	0.5734
21	0.2149	0.2721	0.3297	0.2900	0.3576	0.4286	0.2645	0.3277	0.4013
28	0.2103	0.2830	0.3366	0.2762	0.3446	0.4819	0.2645	0.3277	0.4013

Note: The table reports the quantiles across countries of RMSFE for each method, for each of the five nowcast days under consideration. The RMSFEs have been scaled up by 100 from the log difference transformation.
Abbreviations: PMIDAS, panel mixed data sampling; RMSFE, root mean squared forecast error.

TABLE 7 Inflation nowcast RMSFE by quantile—month-on-month.

	PMIDAS						Benchmark		
	Pooled			Not pooled			TS AR		
Days	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
1	0.0756	0.0828	0.1178	0.0869	0.1212	0.1556	0.0781	0.0886	0.1256
7	0.0697	0.0795	0.1083	0.0826	0.1085	0.1390	0.0781	0.0886	0.1256
14	0.0688	0.0804	0.1039	0.0844	0.1122	0.1434	0.0781	0.0886	0.1256
21	0.0451	0.0514	0.0644	0.0405	0.0676	0.1169	0.0474	0.0556	0.0747
28	0.0430	0.0502	0.0663	0.0433	0.0757	0.1078	0.0474	0.0556	0.0747

Note: Please see the notes for Table 6, above.

split fraction, the ranking of the models remains very stable across all of the nowcast dates. We therefore believe that the results are not affected by this choice.

5 | EMPIRICAL APPLICATION II: EURO AREA INFLATION NOWCASTING

In this section, we present an additional contrasting empirical application to that above, where we predict the monthly inflation rate of a set of European countries. Although inflation data are monthly, they are published with a two to three week delay which makes timely nowcasts important to short-term policymakers and market participants. We will exploit data from weekly energy prices which follows the approach of Modugno (2013).[17] The setup differs from the previous section on global GDP nowcasting as the data here are released at the same time across countries with the timeliness coming from the use of high-frequency weekly data, whereas in the GDP context, the data releases were staggered across countries. There are few, if any, studies looking to use panel approaches to nowcast inflation so this application may be of stand-alone interest.

5.1 | Data and setup

We will target the annual (y-o-y) growth rate of inflation as this is what tends to be monitored most closely by market participants and news agencies. However, as in the previous section, we will also present results for the month-on-month (m-o-m) inflation target which is also of interest. The data we use are the Eurostat harmonised index of consumer prices (HICP) for which monthly data are available around three weeks after the end of the reference month. We transform the HICP data as annual and monthly log differences (for y-o-y and m-o-m, respectively). As a predictor variable, we use consumer prices of petroleum products, net of duties and taxes, which are taken from the European Commission's Weekly Oil Bulletin (WOB). We use data for automotive gas oil, heating gas oil and Euro Super 95 gasoline, which we average together as in Modugno (2011). We use the same transformations as in the case of inflation, that is, the annual (52 week) and weekly log differences (for y-o-y and m-o-m, respectively).

Our dataset runs from July 2004 and ends in December 2019 which gives us a total of $T = 186$ months and 812 weeks of data. The dataset covers all major countries in the European Union, including the largest Euro area countries Germany,

[17]The earlier version of the paper (Modugno, 2011) analysed both US and Euro area inflation, whereas Modugno (2013) focusses only on US inflation.

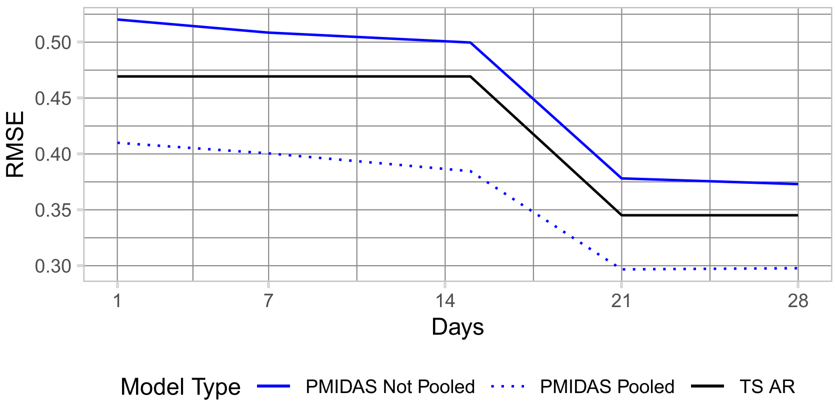


FIGURE 6 Inflation nowcast average RMSFE—year-on-year.
Note: The figure plots the mean of root mean squared forecast error (RMSFE) across all countries on each of the 5 dates in the nowcast period. The single predictor variable used is the oil price.

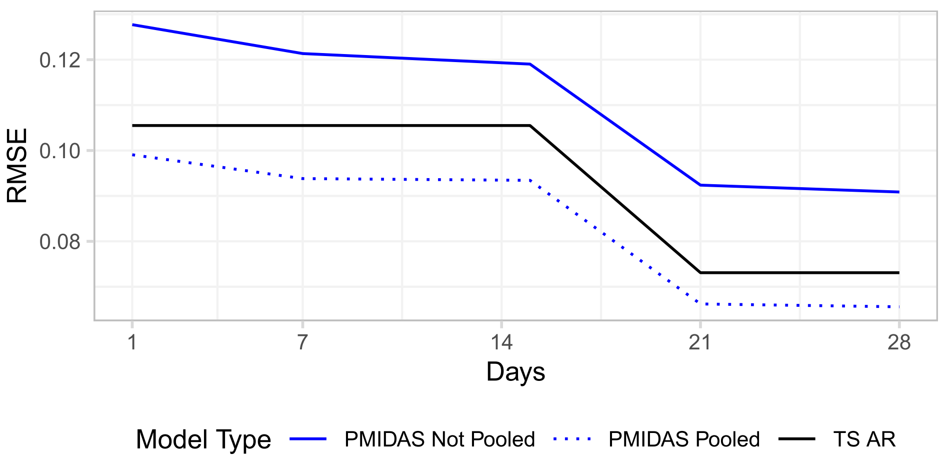


FIGURE 7 Inflation nowcast average root mean squared forecast error (RMSFE)—month-on-month.
Note: See notes for Figure 6.

France and Italy.[18] In performing the out-of-sample evaluation, as in the previous empirical application, we will retain 30% of the sample for evaluating the nowcasts, so $P = 0.3T$. This means that we start nowcasting in May 2015 and continue until we reach the end of the sample. We will evaluate the performance of the PMIDAS model with four weekly lags (both with pooled and nonpooled coefficients) and compare the performance to a time series AR benchmark.

We will make a sequence of inflation nowcasts on different dates, v, throughout the month. We start on day 1 of the reference month and then make four subsequent nowcasts on days 7, 14, 21 and 28. Using information on the HICP release schedule, we always attribute the inflation release to occur when we update the model on day 21. This means that at the beginning of the nowcast month, we do not have the past month's inflation data; this only becomes available when we update the model on day 21. Regarding the weekly WOB data, on each nowcast date, we will use the most recent weekly data point which has been released before the nowcast date. We always use the four most recent weeks' data to make the nowcast. In the same way as the previous empirical application, we will summarise the nowcast performance for each of the countries on each nowcast date using the quantiles and average of the $RMSFE_{v,i}$ statistic described in Equation (14).

5.2 | Results

We now present the nowcast evaluation results for the y-o-y and m-o-m inflation nowcasting exercise. In a similar way as before, Tables 6 and 7 display the quantiles of the RMSFE across countries, for the pooled and nonpooled PMIDAS model and the time series AR benchmark. The results are similar to the previous empirical application in the sense that we find that the pooled PMIDAS model outperforms the time series AR benchmark for the reported quantiles for both y-o-y and m-o-m targets.

[18]The full set of countries is as follows: Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Slovenia, Spain and Sweden.

Figures 6 and 7 graphically display the nowcast performance on average throughout the nowcast period on the five different nowcast dates. Likewise, from the figures, we can again validate the superiority of the pooled PMIDAS model relative to the benchmark. We also see that the nowcasts tend to improve as the weekly information is added, especially when we nowcast the y-o-y inflation rate. We note that, as is common in nowcasting studies involving an autoregressive term, there is a sharper drop in the RMSFE on the date when the previous period's inflation is released, in other words on day 21 of the nowcast period.

Overall, we find encouraging results for the use of PMIDAS-type models in the context of nowcasting a panel of European countries' inflation. Coupled with the previous section's results on global GDP nowcasting, there is evidence that this method can usefully be applied in a variety of different settings.

6 | CONCLUSION

In this paper, we build a mixed-frequency panel data nowcasting model that can simultaneously make predictions of a large number of countries, regions or sectors. Our approach is based on a panel version of a UMIDAS-type nowcasting model, which we extend to allow for heterogeneous coefficients and CSD errors with a factor structure. We base our estimation approach on the CCE estimation method of Chudik and Pesaran (2015), which must be adapted to the nowcasting setting. This requires us to use only the lags of the data which are available on the date which we make the nowcast, unlike existing CCE approaches which use contemporaneous variable for estimation which is suitable for causal studies but not for forecasting.

We provide two contrasting empirical applications of our methodology: nowcasting a large amount of global countries' GDP and nowcasting European countries' inflation. The first main conclusion from both of our empirical studies is that our proposed PMIDAS approach is capable of beating a simple benchmark model, when we switch off heterogeneity and pool the coefficients of the model with heterogeneity only coming through the fixed effects. The results imply evidence in favour of pooling in the debate of whether 'to pool or not to pool' (Wang et al., 2019), though our model can flexibly allow for more heterogeneity if required in other empirical settings. From the contrasting nature of our applications, we conclude that our method has the potential to work well in nowcasting other types of economic variables. Finally, our results also show that adding new releases of data across variables and countries is able to improve nowcast accuracy in a roughly monotonic fashion. From this, we conclude that, although existing studies typically assess nowcast performance as new data arrives for a single country across several variables, there is also benefit in incorporating timely data releases which occur across different countries in a panel data context.

ACKNOWLEDGEMENTS

The views in this paper are of the authors and not of any affiliated institution. We are grateful for helpful comments by Martin Weale, Michele Piffer and participants in the seminar at King's College London; Kirstin Hubrich, Tatevik Sekhposyan, Danilo Cascaldi-Garcia, John Zito, Andrew Martinez and other participants of the NBER-NSF Time Series Conference; and to Raffaella Giacomini and other participants of the Econometric Society Winter Meeting 2021.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://doi.org/10.15456/jae.2023062.1054260926.

DATA AVAILABILITY STATEMENT

All data are available in the public domain.

REFERENCES

Aastveit, K. A., Gerdrup, K. R., Jore, A. S., & Thorsrud, L. A. (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business and Economic Statistics*, *32*(1), 48–68.

Babii, A., Ball, R. T., Ghysels, E., & Striaukas, J. (2020). Machine learning panel data regressions with an application to nowcasting price earnings ratios. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3670847

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

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting*, *27*(2), 153–173.

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. *ECB Working Papers*, *2*(15), 195–237.

Bragoli, D., & Fosten, J. (2018). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics*, *80*(2), 259–282.

Canova, F., & Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics*, *120*(2), 327–359.

Cascaldi-Garcia, D., RT Ferreira, T., Giannone, D., & Modugno, M. (2021). Back to the present: Learning about the euro area through a now-casting model. *International Finance Discussion Paper*. (1313).

Caselli, F., Grigoli, F., Lafarguette, R., & Wang, C. (2020). Predictive density aggregation: A model for global GDP growth. *IMF Working Papers*, *20*(78).

Cepni, O., Güney, I. E., & Swanson, N. R. (2019). Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. *International Journal of Forecasting*, *35*(2), 555–572.

Chen, S., & Ranciere, R. (2019). Financial information and macroeconomic forecasts. *International Journal of Forecasting*, *35*(3), 1160–1174.

Chudik, A., Grossman, V., & Pesaran, M. H. (2016). A multi-country approach to forecasting output growth using PMIS. *Journal of Econometrics*, *192*(2), 349–365. Innovations in Multiple Time Series Analysis.

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, *188*(2), 393–420.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, *26*(4), 546–554.

Clements, M. P., & Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, *24*(7), 1187–1206.

Eickmeier, S., & Ng, T. (2011). Forecasting national activity using lots of international predictors: An application to New Zealand. *International Journal of Forecasting*, *27*(2), 496–511.

Ferrara, L., & Marsilli, C. (2019). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *World Economy*, *42*(3), 846–875.

Foroni, C., Marcellino, M., & Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *178*(1), 57–82.

Fosten, J., & Greenaway-McGrevy, R. (2022). Panel data nowcasting. *Econometric Reviews*, *41*(7), 675–696.

Fosten, J., & Gutknecht, D. (2018). Testing nowcast monotonicity with estimated factors. *Journal of Business and Economic Statistics*, *38*(1), 107–123.

Garnitz, J., Lehmann, R., & Wohlrabe, K. (2019). Forecasting GDP all over the world using leading indicators based on comprehensive survey data. *Applied Economics*, *51*(54), 5802–5816.

Gavin, W. T., & Theodorou, A. T. (2005). A common model approach to macroeconomics: Using panel data to reduce sampling error. *Journal of Forecasting*, *24*(3), 203–219.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, *193*(2), 294–314.

Ghysels, E. (2018). Mixed frequency models. *Oxford Research Encyclopedia of Economics and Finance*.

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

Greenaway-McGrevy, R. (2019). Asymptotically efficient model selection for panel data forecasting. *Econometric Theory*, *35*(4), 842–899.

Hansen, P., & Timmermann, A. (2012). Choice of sample split in out-of-sample forecast evaluation. Economics Working Papers ECO2012/10, European University Institute.

Hantzsche, A., Lopresto, M., & Young, G. (2018). Using NiGEM in uncertain times: Introduction and overview of NiGEM. *National Institute Economic Review*, *244*, R1–R14.

IMF (2021). World economic outlook recovery during a pandemic health concerns, supply disruptions, and price pressures.

Kindberg-Hanlon, G., & Sokol, A. (2018). Gauging the globe: The bank's approach to nowcasting world GDP. *Bank of England Quarterly Bulletin*, *58*(3), 21–30.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529–542.

Lee, Y., & Phillips, P. C. B. (2015). Model selection in the presence of incidental parameters. *Journal of Econometrics*, *188*(2), 474–489.

Marcellino, M., Porqueddu, M., & Venditti, F. (2016). Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business and Economic Statistics*, *34*(1), 118–127.

Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550.

Modugno, M. (2011). Nowcasting inflation using high frequency data. *ECB Working Paper 1324*.

Modugno, M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, *29*(4), 664–675.

OECD (2020). OECD share in world GDP stable at around 50% in PPP terms in 2017. https://www.oecd.org/sdd/prices-ppp/oecd-share-in-world-gdp-stable-at-around-50-per-cent-in-ppp-terms-in-2017.htm

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012.

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

Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters*, *107*(2), 95–98.

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, *32*(2), 257–270.

Siliverstovs, B. (2017). Short-term forecasting with mixed-frequency data: A MIDASSO approach. *Applied Economics*, *49*(13), 1326–1343.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Wang, W., Zhang, X., & Paap, R. (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions?. *Journal of Applied Econometrics*, *34*(5), 724–745.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, *64*(5), 1067–1084.

Xu, Q., Zhuo, X., Jiang, C., Liu, X., & Liu, Y. (2018). Group penalized unrestricted mixed data sampling model with application to forecasting US GDP growth. *Economic Modelling*, *75*(May), 221–236.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of the article.

How to cite this article: Fosten, J., & Nandi, S. (2023). Nowcasting from cross-sectionally dependent panels. *Journal of Applied Econometrics*, *38*(6), 898–919. https://doi.org/10.1002/jae.2980

APPENDIX A: ADDITIONAL SIMULATION RESULTS

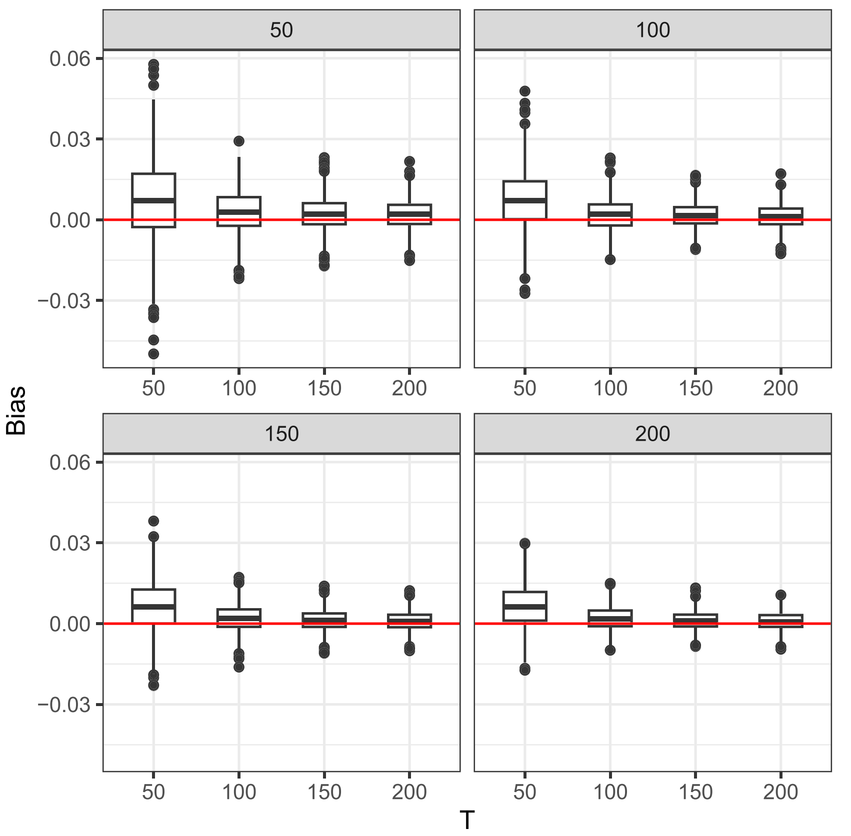


FIGURE 8 Bias in ϕ; $q = 3$.
Note: The figures display the distribution of the difference in absolute biases of the parameter ϕ for the estimation methods lagged common correlated effects (LCCE) relative to CCE. Figures lower than zero mean that LCCE has lower absolute bias than CCE. The panel header shows the number of cross-sections.

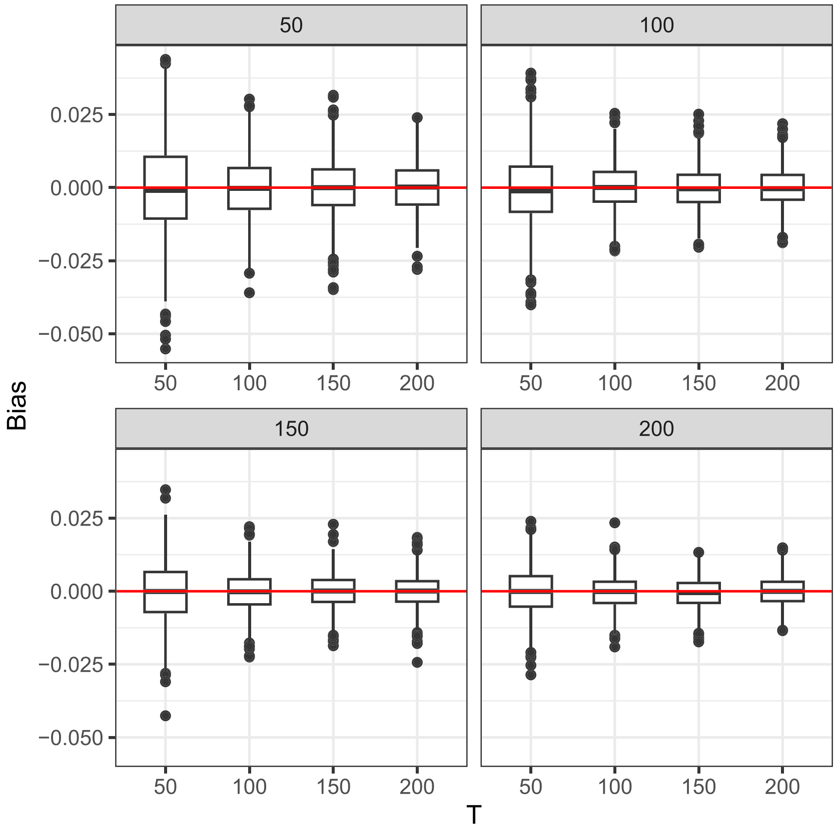


FIGURE 9 Bias comparison in $\beta^{(0)}$; $q = 3$. *Note*: The figures display the distribution of the difference in absolute biases of the parameter $\beta^{(0)}$ for the estimation methods lagged common correlated effects (LCCE) relative to common correlated effects (CCE). Figures lower than zero mean that LCCE has lower absolute bias than CCE. The panel header shows the number of cross-sections.

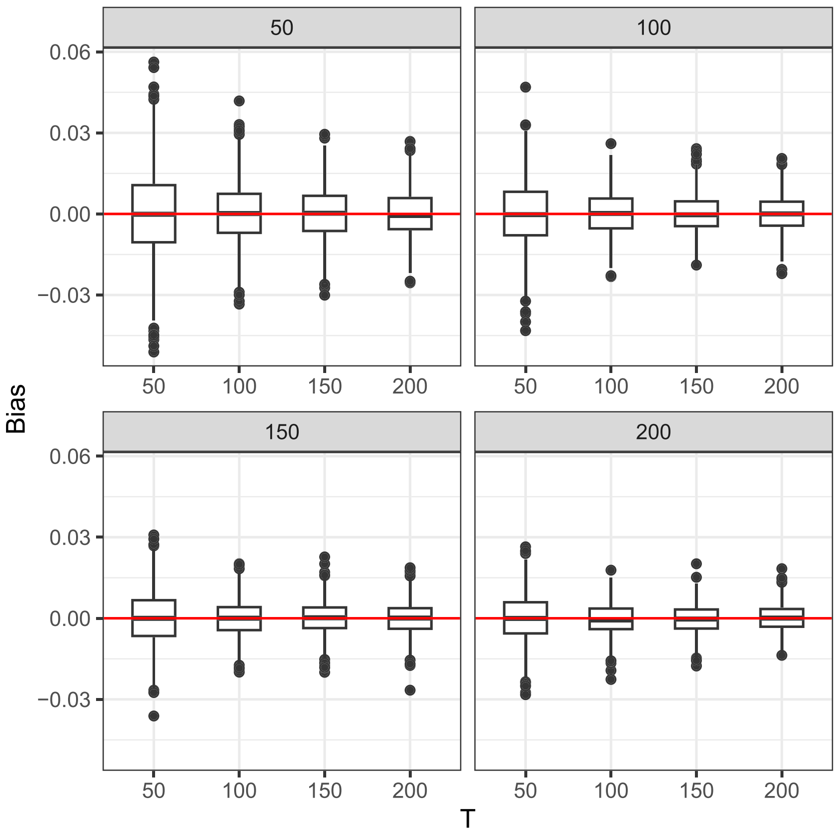
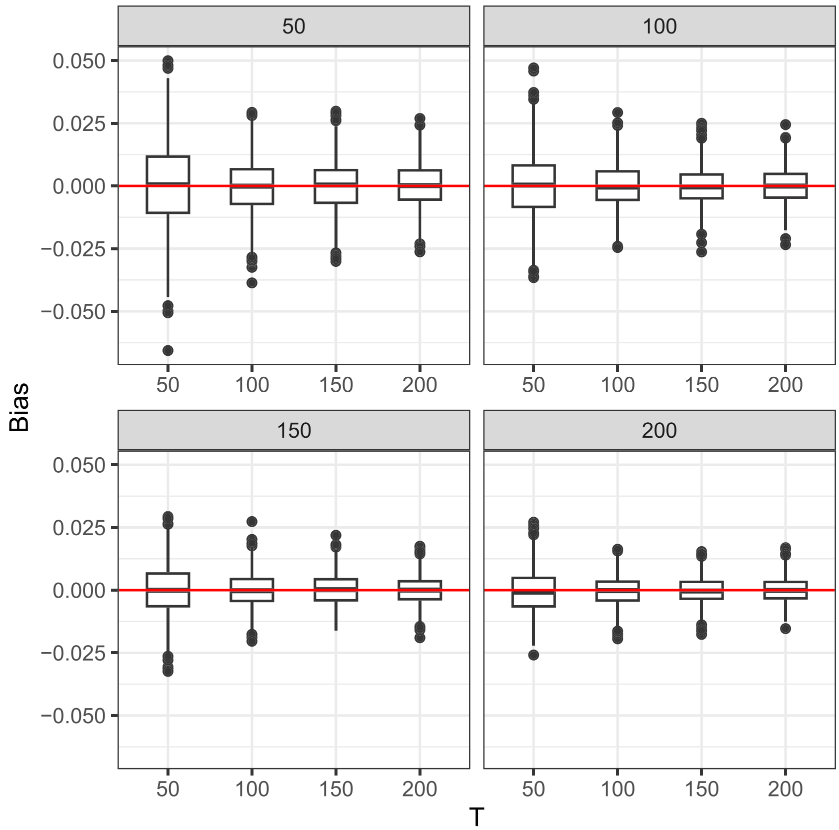


FIGURE 10 Bias comparison in $\beta^{(1)}$; $q = 3$. *Note*: The figures display the distribution of the difference in absolute biases of the parameter $\beta^{(1)}$ for the estimation methods lagged common correlated effects (LCCE) relative to common correlated effects (CCE). Figures lower than zero mean that LCCE has lower absolute bias than CCE. The panel header shows the number of cross-sections.

FIGURE 11 Bias comparison in $\beta^{(2)}$; $q = 3$.
Note: The figures display the distribution of the difference in absolute biases of the parameter $\beta^{(2)}$ for the estimation methods lagged common correlated effects (LCCE) relative to common correlated effects (CCE). Figures lower than zero mean that LCCE has lower absolute bias than CCE.The panel header shows the number of cross-sections.



APPENDIX B: EMPIRICAL APPLICATION I: ADDITIONAL PLOTS

B.1 | Release calendars

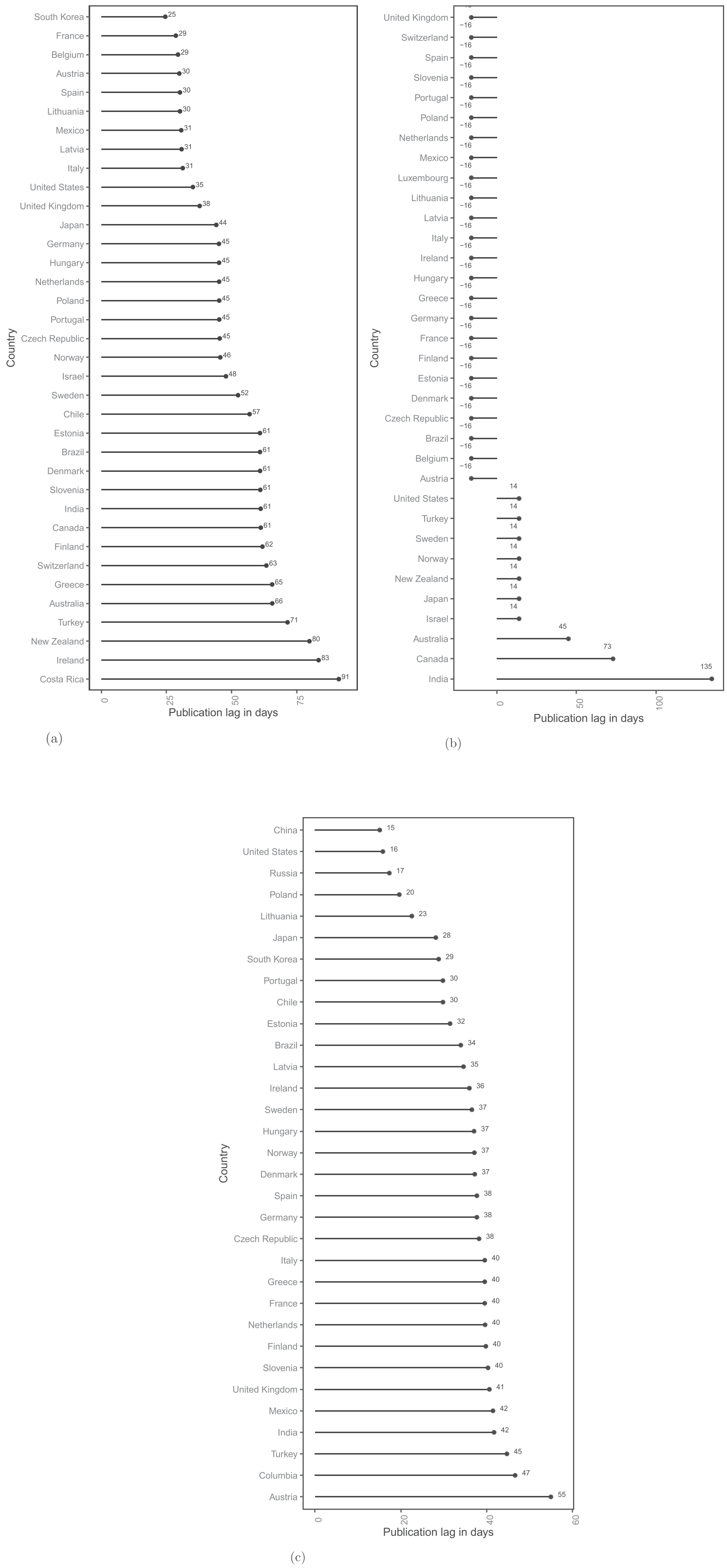


FIGURE 12 Publication lag across different countries. (a) Pseudo calendar for the first release of quarterly gross domestic product (GDP) (*Source*: Bloomberg Finance L.P). (b) Pseudo calendar for Business Survey Manufacturing (*Source*: Macrobond)*. (c) Pseudo calendar for Industrial Production (*Source*: Bloomberg Finance L.P).

*Negative lag indicates the data are available before the start of the month.