KING'S BUSINESS SCHOOL

Data Analytics for Finance & Macro Research Centre

KING'S College LONDON

Nowcasting U.S. State-level CO_2 Emissions and Energy Consumption

J. Fosten, S. Nandi

Working paper No. 2022/2 | May 2022 | ISSN 2516-5933

Nowcasting U.S. State-level CO_2 Emissions and Energy Consumption*

Jack Fosten[†‡] Shaoni Nandi[§¶]

November 8, 2022

Abstract

This paper proposes panel nowcasting methods to obtain timely predictions of CO_2 emissions and energy consumption growth across all U.S. states. This is of crucial importance not least because of the increasing role of sub-national carbon abatement policies but also due to the very delayed publication of the data. Since the state-level CO_2 data are constructed from energy consumption data, we propose a new panel bridge equation method. We use a mixed frequency set-up where economic data are first used to predict energy consumption growth. This is then used to predict CO_2 emissions growth while also allowing for cross-sectional dependence across states using estimated factors. We evaluate the models' performance using an out-of-sample forecasting study, finding gains in using timely economic data to nowcast and backcast state-level energy consumption growth. These gains are sizeable in many states, even around two years before the data are eventually released. In predicting CO_2 emissions growth, nowcast accuracy gains are more focussed on a few states although accurate nowcasts can be obtained across all states if they are made after the current year's energy consumption data are released.

JEL Classification: C23, C53, Q47, Q50, R10

Keywords: Panel Data, Nowcasting, CO_2 Emissions, Energy Consumption, Environmental Degradation

*The views in this paper are those of the authors and not of any affiliated institution. All data are available in the public domain.

†Corresponding Author. King's Business School, King's College London, WC2B 4BG London, UK, E-mail: jack.fosten@kcl.ac.uk
‡Data Analytics in Finance and Macro (DAFM) Research Centre, King's College London, UK
§King's Business School, King's College London, WC2B 4BG London, UK, E-mail: shaoni.nandi@kcl.ac.uk
¶Department of Statistics & Information Management, Reserve Bank of India, Mumbai, India

1 Introduction

The growing climate emergency has rapidly expanded the need for policies on abating CO_2 emissions due to fossil fuel energy production and consumption. The importance of using environmental variables in economic modelling is now well accepted since the seminal DICE model of Nordhaus (1992). This has led to significant recent debate amongst economic policymakers on tracking the social cost of carbon (Rennert et al., 2021) as well as the widespread use of environment-economic models by international institutions such as the OECD and the United Nations.[1] In turn, this has placed increasing importance on the ability to forecast and monitor both short-term and long-term energy consumption and CO_2 emissions. Our focus will be on near-term prediction, or "nowcasting" of these environmental variables, which has only recently received attention by Bennedsen et al. (2021) in the context of nowcasting national U.S. CO_2 emissions.

In this paper, we propose new models for jointly nowcasting multiple regions' energy consumption and CO_2 emissions, specifically for states in the U.S., which has not yet been studied in the existing literature. This improves upon studies which look only at the national context by allowing a more granular overview of regional environmental degradation. The focus on sub-national variables is important for several reasons. Firstly, there is growing evidence that sub-national efforts to reduce emissions can accelerate the achievement of national abatement targets (see Hultman et al., 2020 and references therein). Secondly, the discussion of local-level environmental action has gained a stage in the largest climate meetings, such as the dedicated *"Cities, Regions and the Built Environment"* day at COP26. Finally, there are already many sub-national environmental initiatives in the U.S., where around half of all U.S. states currently have greenhouse gas emissions targets,[2] and more than ten states which participate in the Regional Greenhouse Gas Initiative (RGGI), a market-based program to reduce emissions. For these reasons, it is crucial that policymakers have access to up-to-date data on regional CO_2 emissions and energy consumption. However, it is very challenging to monitor the movements in these variables in real time as the data are only available annually and with very long publication lags. This challenge has not been addressed by existing academic studies.

This paper aims to fill this gap in the literature by providing a novel nowcasting methodology for U.S. state-level energy consumption and CO_2 emissions growth. This allows us to obtain timely predictions of these variables before the data are published. This builds on existing academic studies in several ways. Firstly, our study is unique in nowcasting state-level energy consumption and CO_2 emissions, where only the recent study of Bennedsen et al. (2021) looks at nowcasting national CO_2 emissions and not at state level. Secondly, our paper provides a novel application of recently-emerging panel data nowcasting methods which have typically been used only for predicting macroeconomic variables like real GDP (Fosten and Greenaway-McGrevy, 2022) and not environmental variables. More broadly, panel data nowcasting is a relatively new and increasing field (Koop et al., 2020; Babii et al., 2020;

[1] See: https://www.oecd.org/environment/indicators-modelling-outlooks/modelling.htm and https://www.unep.org/explore-topics/green-economy/what-we-do/economic-and-trade-policy/green-economy-modelling [Last accessed: 01/09/2022]

[2] See: https://www.c2es.org/content/state-climate-policy/ [Last accessed: 29/03/2022]

Larson and Sinclair, 2022) relative to the long history of time series nowcasting (see the surveys of Banbura et al., 2013; Bok et al., 2018). Finally, our paper is different from traditional nowcasting studies of real GDP where publication lags may be only one or two months. In our setting, there is even stronger motivation for the use of nowcasting due to the annual frequency and the abnormally large publication lags in the U.S. state-level energy consumption and emissions data. The CO_2 data are only available over two years after the end of the relevant year, while energy consumption data have a delay of around a year and a half. These publication lags make the problem more interesting than existing studies and require methods which are capable not just of nowcasting but also backcasting.

The first contribution of the paper is to propose a panel data nowcasting methodology for state-level energy consumption and CO_2 emissions growth. Motivated by the fact that the emissions data are calculated directly from energy consumption data, we propose a two-step bridge equation approach adapted to the case of panel data. We first use a mixed-frequency panel MIDAS model to obtain nowcasts of annual state-level energy consumption growth using higher frequency quarterly economic activity data. This model we use is adapted from the mixed frequency approach of Ghysels (2016), which we extend from the time series to the panel data context, and the model's predictions can be updated every time new information arrives. We then employ a panel bridge equation approach to transform the nowcasts of energy consumption growth into nowcasts of CO_2 emissions growth. In doing so, we use a multi-factor error structure to allow for cross-sectional dependence across states in the style of Chudik and Pesaran (2015). Our panel bridge equation model is similar to the well-known time series bridge equation approach (see for example Baffigi et al., 2004; Foroni and Marcellino, 2014; Schumacher, 2016) with the difference that we extend this to allow the modelling of panel data, which is an improvement in contexts where regional data are available. The cross-sectional dependence structure we use is similar to the recent panel nowcasting approach of Fosten and Nandi (2021), which in this paper we adapt to the case of bridge equation models.

The second contribution of the paper is the empirical part where we perform a detailed pseudo out-of-sample forecasting study using our models to predict energy consumption and CO_2 emissions growth over a period of history. We mimic the release schedule of the variables in real time and make multiple nowcasts and backcasts for every period under consideration. This allows us to assess how the performance of these methods changes as we add new information into the nowcasting model, as is commonly done in empirical nowcasting studies (see, for instance, Giannone et al., 2008; Banbura et al., 2013; Bok et al., 2018). For the predictions of energy consumption growth, we use real GDP or real personal income growth. Since these economic series have a much lower publication lag, we first of all consider restricting the data flow to only use the year-end annual growth rate of these series in predicting annual energy consumption growth, before turning to assess whether incorporating the mixed frequency quarterly data can make further improvements. We finally use the bridge equation method to feed in these energy consumption predictions and arrive at predictions of CO_2 emissions growth.

We make several interesting findings. We find that the predictions of energy consumption growth improve across states when current economic data are used for nowcasting and backcasting, relative to a naïve benchmark. There are particularly sizeable gains in several states, which we assess by looking at the across-state distribution of the gain in predictive accuracy of our model relative to the benchmark. Even when using the economic data at the annual frequency, gains in predictive accuracy occur around a year ahead of the release of the energy consumption data. This highlights the gain from using timely information in prediction, even if there is no difference in the frequency of the series. Furthermore, when we increase the frequency to use quarterly economic data we find that nowcast improvements are possible even within the nowcast year itself, around two years before the release of the data for the target variable. With regards to the CO_2 predictions, the gains are less notable when adding economic data although still sizeable in some states, and the biggest gains come when we add in the current year's energy consumption. This is still important as these accurate predictions come many months before the release of the data by the statistical authorities, and we use a much simpler methodology than that used in constructing the data. We find some additional but marginal gain from using factors estimated to pick up common correlated effects in the CO_2 bridge equation method. We also provide various robustness checks such as the use of per capita energy consumption and emissions growth as target variables.

Our empirical study builds on an increasing body of empirical work in nowcasting. While only the aforementioned study of Bennedsen et al. (2021) looks at nowcasting environmental variables, there have been a vast amount of studies using nowcasting for macroeconomic monitoring. The majority of studies look at nowcasting real GDP and have done so in a variety of different contexts: developed economies (Bok et al., 2018; Anesti et al., 2022), emerging economies (Bragoli and Fosten, 2018; Dahlhaus et al., 2017), global GDP (Ferrara and Marsilli, 2019) and so on. Nowcasting has also been applied to several other macroeconomic series such as the GDP components (Fosten and Gutknecht, 2020), inflation (Modugno, 2013; Knotek and Zaman, 2017) and unemployment claims (Larson and Sinclair, 2022). Our paper helps to shift this focus from macroeconomic to environmental nowcasting, which we believe will be a fruitful area of future research.

The rest of the paper is organised as follows. Section 2 describes the data sources used in the study. Section 3 describes the models we propose and Section 4 details the pseudo out-of-sample methodology we use in evaluating these models. Section 5 discusses the results of the pseudo out-of-sample experiment and Section 6 concludes the paper. The Appendix houses additional sets of results not included in the main text.

2 Data

2.1 CO_2 Emissions

State-level CO_2 emissions data are available from the U.S. Energy Information Administration (EIA).[3] The data are available on an annual basis with observations from 1980 onwards. The data cover the CO_2 emissions from

[3]See: https://www.eia.gov/environment/emissions/state/ [Last accessed: 11/11/2021]

direct fuel use across various sectors: commercial, industrial, residential and transportation. We focus on the total emissions by state but we will also consider per-capita CO_2 emissions as this has been the target variable of other studies (Bennedsen et al., 2021). Of crucial importance to this study is that the publication lag for the CO_2 emissions data is very large, around two years and three months after the end of the reference year. For instance, the data for 2018 were released at the beginning of March 2021. This lag is considerably larger than other types of state-level data such as the economic variables mentioned below. This lack of timeliness will mean that both nowcasting and backcasting are appropriate.

In producing the data, the EIA estimate state-level CO_2 emissions based on underlying energy consumption data from the State Energy Data System (SEDS).[4] Knowing this aspect of the data construction is what motivates the use of a bridge equation where total state-level CO_2 emissions data are directly linked to total state-level energy consumption data.[5] We note that this approach will be like an approximation to the more disaggregated way in which the EIA computes the state-level CO_2 data. To be more precise, according to the EIA's methodology documentation,[6] the conversion to CO_2 emissions from energy consumption is first made at a very granular level by fuel type and sector, using different emissions factors and proportions of fuel used in fuel combustion. After conversion, the total CO_2 emissions are summed up from the disaggregates. An alternative approach to ours would be a bottom-up approach to mimic the EIA's calculation by nowcasting the disaggregate energy consumption series, converting them and then aggregating them afterwards. However, we do not pursue this approach as it would entail a large amount of additional nowcast uncertainty: (i) the nowcast errors from a large number of individual disaggregates summed up to get the total, (ii) the errors from predicting the emissions factors which are themselves estimated and would require nowcasting, (iii) some estimation of the proportions of each fuel type that is used in combustion, which the EIA bases on various sources. We prefer a direct top-level approach, much in the same way that GDP nowcasters target the aggregate GDP series and not the very granular disaggregated output series which are also available.

2.2 Energy Consumption

The data for state-level energy consumption are also available on an annual basis. The data are available from the SEDS, mentioned above, also produced by the EIA. The annual time series for each state are available from 1960 onwards. As with CO_2 emissions, we will consider both the raw and per-capita energy consumption in our analysis. Regarding the timeliness of the data, although the data frequency is the same as that of CO_2 emissions, the SEDS data are published in a more timely fashion. Here, the publication lag is around one year and six months, which is roughly nine months quicker than for the CO_2 data. For instance, the data for 2019 were published at the end of June 2021. Although the data are more timely, if we want to use the current year's energy consumption

[4]See: https://www.eia.gov/state/seds/ [Last accessed: 11/11/2021]]

[5]This is instead of modelling CO_2 emissions directly as a function of, say, economic variables. We tried this latter approach in our empirical investigations but found it to perform worse than modelling using energy consumption.

[6]See: https://www.eia.gov/environment/emissions/state/pdf/statemethod.pdf [Last accessed: 31/08/22]

in predicting CO_2 emissions, this would constitute a backcast and not a nowcast. In order to obtain nowcasts of energy consumption and therefore CO_2 emissions, we require data which are available in a much more timely fashion, such as the economic indicators outlined next.

2.3 Economic Indicators

Since the aim is to produce state-level energy consumption nowcasts, it is natural to use state-level economic indicators. We consider two different variables: real GDP and real personal income (PI). Both of these series are available from the Bureau of Economic Analysis (BEA).[7] The quarterly PI data are available at a quarterly frequency for all states from 1950 onwards, which we deflate by the GDP deflator for the U.S. to obtain real figures. The real GDP data have a much shorter history than PI. Annual data are available from 1997, and are published separately from the quarterly data which are only available from 2005. We will therefore consider both annual and quarterly versions so that we can compare GDP and PI as predictors in the annual case. In the quarterly case we will focus on PI data only as the time span of the quarterly GDP data is not long enough for a meaningful pseudo out-of-sample reconstruction.

There are two factors which make these economic series appropriate for nowcasting energy consumption and therefore CO_2 emissions. Firstly, their quarterly frequency makes them anyway more timely than the annual data. Secondly, for both PI and GDP, the publication lag is around three months after the end of the reference quarter.[8]. This implies that already in the middle of the nowcast year, the first quarter of that year's economic data are available for making predictions of energy consumption for that same year.

It is difficult to expand on the set of economic predictor variables we use due to the limited availability of state-level data. For instance, Bennedsen et al. (2021) note that the Industrial Production (IP) index is useful in nowcasting national CO_2, but IP data are not available by state. However, we will instead show some additional results using the Philly Fed's State Coincident Indexes.[9] These indices are available in a timely fashion at the monthly level and are constructed using a dynamic factor model on four state-level employment type series, which bears resemblance to the factor model methods used in nowcasting with many predictors.

3 Panel MIDAS and Bridge Equation Methodology

In this section we describe the models we use to predict the annual growth of energy consumption ("EC" hereafter) and subsequently of CO_2 emissions growth.[10] As mentioned above, the CO_2 data are released in March over two

[7]See: https://www.bea.gov/data/gdp/gdp-state and https://www.bea.gov/data/income-saving/personal-income-by-state [Last accessed: 12/02/2022]

[8]We will assume the same publication lags for GDP and PI, as these data are generally released in the same month, often on the same date. See: https://www.bea.gov/news/schedule/full

[9]See: https://www.philadelphiafed.org/surveys-and-data/regional-economic-analysis/state-coincident-indexes

[10]We focus on the growth rates of these series as is standard in the macroeconomic nowcasting literature when analysing trending unit root processes. Since there is little existing evidence on unit roots in the state-level energy consumption and CO_2 emissions data we performed a battery of panel unit root tests (the Levin et al. (2002) (LLC) test, the Im et al. (2003) test (IPS) and the Choi (2001)

years after the reference year, whereas the EC data are published in June each year, a year and a half after the reference year. The economic data are available in a more timely fashion. Our approach is therefore to use a bridge equation to compute predictions of CO_2 emissions growth for the target year by first obtaining predictions of EC using economic indicators. Therefore, while CO_2 emissions are the 'target' variable of the bridge equation, we also obtain timely predictions of EC which is of separate interest in itself.

We differ from the prevalent bridge equation models (see Foroni and Marcellino, 2014; Schumacher, 2016, and the references therein) in several important ways. Firstly, we use a panel data set-up instead of a time-series approach that is common in economic nowcasting. Secondly, the EC variable we predict in the first step is not available at a higher frequency but has lesser publication lags as compared to our final target variable, CO_2 emissions. Lastly, we do not restrict ourselves to AR models for predicting EC as is typical of economic bridge equation set-ups. Instead, we also use panel data models and incorporate mixed frequencies to use higher frequency quarterly PI or real GDP growth.

3.1 Panel MIDAS Model for Energy Consumption

We now describe the panel model for nowcasting energy consumption growth using economic data. Since the economic data are both more timely and available at a higher frequency, we will try out two different approaches. In the first approach we simply use annual data for both the energy consumption and economic variables, in order to assess whether the timeliness of economic data is useful even when using annual data. In the second approach we check whether inserting quarterly economic data in a mixed frequency approach brings further benefits. This is also motivated by the data constraint mentioned above, that the quarterly real GDP figures do not have sufficient history to be used in a pseudo out-of-sample experiment whereas the annual GDP data do have sufficient history. The annual version will therefore give results both when using GDP and PI whereas the quarterly version will only yield results when using PI.

Annual Frequency Model

We start out with the model which predicts EC using the available autoregressive lags on day v of the nowcast period as well as the annual lags of the economic indicator:

$$c_{i,t} = \alpha_{vi} + \phi_v c_{i,t-d_{cv}} + \beta_v X_{i,t-d_{Xv}} + u_{v,i,t} \tag{1}$$

where t denotes the annual time index and $c_{i,t}$ is a generic notation indicating the annual growth rate in energy consumption. In the main results this is simply the percentage change in actual energy consumption for state i in year t, in other words the growth rate of $EC_{i,t}$. Alternatively, we also explore the results where $c_{i,t}$ is the growth

test). As expected, these tests confirm non-stationarity in levels and stationarity in growth rates. We do not present the results in the text for the sake of brevity.

rate of per capita consumption, in other words the growth rate of $\frac{EC_{i,t}}{pop_{i,t}}$, where $pop_{i,t}$ is the state population. In a similar way, $X_{i,t}$ is a generic notation for the annual growth rate of the economic indicator, either GDP or PI, and could be actual or per-capita according to the target variable.

The model in equation (1) takes account of the ragged edge problem in the following way. Denoting v to be the date of prediction, we define d_{cv} as the available lag of $c_{i,t}$ at the time of prediction, based on its publication lag. Similarly, d_{Xv} is used to denote the available lag of $X_{i,t}$ used in the model at time v. As we change the nowcast date v, the available lags of each variable may change and the model lag structure is updated to accommodate new information. Since the model variables change on each date, v, the parameters of the model and the error term are also indexed by v. To give an example, in nowcasting year t, if v is the start of year t, based on the data flow described in the Data section above, the model will use $c_{i,t-3}$ and $X_{i,t-2}$. After March of year t, the economic data are updated and $X_{i,t-2}$ is replaced with $X_{i,t-1}$ and so on. The full updating procedure will be described later when we introduce the pseudo out-of-sample set-up.

Mixed Frequency Model with Quarterly Data

We now re-state equation (1) so that the quarterly frequency of the economic data is fully utilised in a mixed-frequency model. This model is a panel version of the unrestricted MIDAS (UMIDAS) model (see Foroni et al., 2015; Schumacher, 2016) which takes on the following form:

$$c_{i,t} = \alpha_{vi} + \phi_v c_{i,t-d_{cv}} + \beta_v^{(m)\prime} \boldsymbol{x}_{i,t-\frac{q_v}{4}} + u_{v,i,t} \tag{2}$$

where $\boldsymbol{x}_{i,t-\frac{q_v}{4}} = \begin{pmatrix} x_{i,t-q_v/4}, & x_{i,t-(q_v-1)/4}, & x_{i,t-(q_v-2)/4}, & x_{i,t-(q_v-3)/4} \end{pmatrix}'$ denotes the generic stacked skip-sampled PI or GDP growth which is inserted into the model with a quarterly lag of q_v at nowcast date v. Note that a lag of one quarter is denoted in annual terms as $t - \frac{1}{4}$. In equation 2, the slope coefficient $\beta_v^{(m)}$ is a vector of length four, corresponding to the stacked skip-sampled process $\boldsymbol{x}_{i,t-\frac{q}{4}}$ consisting of the four quarters in a year. The nowcast updating works in the same way as for equation (1) above. When we change the nowcast date, v, we update the lag structure to incorporate any newly-available annual data for $c_{i,t}$ and quarterly data for $\boldsymbol{x}_{i,t}$.

Equations 1 and 2 are panel versions of the ARX model (AR with an exogenous regressor) and we refer to it as the ARX model subsequently. We will also use a naïve benchmark method to compare with the predictions from the panel ARX model. For this benchmark we will use a simple historic mean prediction using all available data at the time of making the nowcast.[11] Later on, we use the EC predictions from both the panel ARX and the benchmark model to predict CO_2 emissions and compare the results.

[11] In previous version of the paper we also considered using an autoregressive benchmark but the results are qualitatively similar.

3.2 Bridge Equation for CO_2 Emissions

Here we describe the main nowcasting bridge equation for CO_2 emissions growth, where we plug in the predictions for EC obtained from the previous equations 1 or 2. Define $\hat{c}_{v,i,t}$ generically as the predicted value of $c_{i,t}$ on day v of the nowcast period. The main equation is a panel bridge equation model with a multi-factor error structure:

$$e_{i,t} = \theta_{vi} + \rho_v e_{i,t-d_{ev}} + \delta_v \hat{c}_{v,i,t} + \lambda_v f_t + \varepsilon_{v,i,t} \tag{3}$$

where we define emissions growth, $e_{i,t}$, which either represents the growth of $CO2_{i,t}$, the CO_2 emissions in state i in year t, or the growth of per-capita emissions $E_{i,t} = \frac{CO2_{i,t}}{pop_{i,t}}$. In a similar way to before, the autoregressive lags included in the model depend on the publication lag, which at prediction time v is denoted by d_{ev}. As above, the parameters and error term in equation 3 also depend on v as the model variables change with v.

The variable f_t denotes unknown factors with loadings λ_v which are common across all states and are used to model the cross-sectional dependence in the error terms. In order to estimate these factors, in a similar way to Chudik and Pesaran (2015) they are also assumed to influence the $\hat{c}_{v,i,t}$ in the following way:

$$\hat{c}_{v,i,t} = \zeta_{vi} + \kappa_v e_{i,t-d_{ev}} + \Gamma_v f_t + \epsilon_{v,i,t} \tag{4}$$

We note that equations (3) and (4) assume away heterogeneity (across i) in the factor loadings λ and Γ, which was permitted in the original paper of Chudik and Pesaran (2015). This is partly because pooling coefficients is often seen to be preferable to heterogeneous coefficients in panel forecasting (Wang et al., 2019), and also because our relatively small number of annual time periods makes it less desirable to add coefficient heterogeneity. Thus, the common factors f_t could also be regarded as time fixed-effects (see Pesaran, 2016, Ch. 31, p. 833).

Equations 3 and 4 jointly create a set-up that can be estimated through the Common Correlated Effects (CCE) method. Since the original method of Chudik and Pesaran (2015) was not designed to use for forecasting, we use the lagged common correlated effects (LCCE) approach developed in Fosten and Nandi (2021) which ensures that only the available lags of the predictor variables are used in estimating the factors. In this way, the final prediction equation replaces the unknown factors in equation 3 as follows:

$$e_{i,t} = \theta_{vi} + \rho_v e_{i,t-d_{ev}} + \delta_v \hat{c}_{v,i,t} + \sum_{l=0}^{p_T} \gamma'_{vl} \bar{z}_{v,i,t-l} + \varepsilon_{v,i,t} + O_p(N^{-\frac{1}{2}}) \tag{5}$$

where $\bar{z}_{v,i,t}$ are the factor estimates used to pick up CCE in the errors and p_T is a lag truncation parameter. The factor estimates are obtained by taking a state-weighted average of the vector $z_{v,i,t} = [e_{i,t-d_{ev}}, \hat{c}_{v,i,t}]'$. Chudik and Pesaran (2015) and Fosten and Nandi (2021) discuss the equivalence of least squares estimation of equation 5 and the system of equations 3 and 4. We therefore use panel least squares estimation of equation (5) in our out-of-sample forecasting exercise.

We will compare the results with those from a simple panel ARX model, where we simply estimate equation 3 without the factors f_t. This will allows us to observe any effects from allowing cross-sectional dependence. As a naïve benchmark, in the same way as above, we will use the historic mean using the data available at the time of making the nowcast.

4 Pseudo Out-of-Sample Set-up

We perform pseudo-out-of-sample experiments for nowcasting annual EC and CO_2 emissions growth across the $N = 51$ individual states plus the District of Columbia. We start our out-of-sample nowcasts in 2009 and finish in 2018. As is common in the nowcasting literature (dating back to Giannone et al., 2008) we will make multiple nowcast and backcast updates at different dates, v, for every year in the out-of-sample evaluation period. We do this to replicate the ragged edge in the data using a calendar of releases as they would have occurred in real time. This allows us to see how the nowcasts and backcasts behave, on average, as we add more information whenever it becomes available. For every data release we take into account the new lag of data available, adjust the model lag structure as detailed above, re-estimate the models and obtain first the EC predictions and then the CO_2 predictions from the bridge equation in (5). Once we have finished making nowcasts and backcasts of a given year, we move on to the next year by expanding the information set as in the recursive out-of-sample scheme of West (1996).

To be more specific on the nowcast updating procedure, we will start by making a nowcast at the beginning of the reference year, at the end of January. This can be seen as the first date in Tables 1 and 2 which detail the release calendar in the annual and quarterly data set-up. We then move through the nowcast year, updating in March and then June in both the annual and quarterly set-ups. In the quarterly set-up there are two further releases in the nowcast year in September and December as additional quarterly lags of the economic data become available, as seen in Table 2. This gives a total of three nowcasts in the annual set-up and five in the quarterly set-up. We then move into the next year and start backcasting. For EC this gives a further two updates in both the annual and quarterly set-ups, as we stop updating the economic data after the observation for the target nowcast year has been released (in other words we do not use 'future' economic data to predict current EC). This gives a total of five predictions (three nowcasts and two backcasts) for EC in the annual set-up and seven (five nowcasts and two backcasts) in the quarterly set-up. When it comes to making the CO_2 predictions, we have the same number of predictions made as in the case of EC but there are two additional updates: in March of the second backcast year when the first lag of CO_2 data is released, and in June when the current year's EC data is released. In other words, the last bridge equation nowcast we make of CO_2 will replace the predicted EC with its actual realised value.

We will therefore have multiple nowcasts and backcasts made per year for a total of nine evaluation years from 2009 to 2018. Since all of the data series have slightly different sample sizes, it is useful to consider the proportion

		Month	Year	EC	GDP/PI	CO_2
Nowcast	1	January	0	3	2	4
	2	March	0	3	1	3
	3	June	0	2	1	3
Backcast	4	March	1	2	0	2
	5	June	1	1	0	2
	6	March	2	1	0	1
	7	June	2	0	0	1

Notes: Month and Year denote when the prediction is made, with Year being the number of years after the nowcast year (so Year 0 is the nowcast year itself).
The columns EC, GDP/PI and CO_2 display the available lags of that variable in years, relative to the nowcast year.
The horizontal line after release 5 denotes the point at which we stop predicting EC in the annual set-up. Releases 6 and 7 are only used for predicting CO_2.

Table 1: Release Calendar for the Annual Set-up

		Month	Year	EC	GDP/PI	CO_2
Nowcast	1	January	0	3	5*	4
	2	March	0	3	4*	3
	3	June	0	2	3*	3
	4	September	0	2	2*	3
	5	December	0	2	1*	3
Backcast	6	March	1	2	0*	2
	7	June	1	1	0*	2
	8	March	2	1	0*	1
	9	June	2	0	0*	1

Notes: Same as for Table 1 except the publication lags for GDP/PI (*) are in quarters relative to the last quarter of the nowcast year. A value of 0* means that all quarters of the nowcast year are already available).
The horizontal line after release 7 denotes the point at which we stop predicting EC in the quarterly set-up. Releases 8 and 9 are only used for predicting CO_2.

Table 2: Release Calendar for the Quarterly Set-up

of the sample which is being used for evaluation. In predicting EC, given that annual state-level GDP data begins only in 1997, starting our evaluation in 2009 implies we use around a half of the sample for evaluation of the EC predictions using GDP. Since the data span for PI is much longer we use about 17 per cent of the sample for evaluating the EC predictions using PI. In predicting CO_2, given that the data runs from 1980 to 2018, we assess the accuracy of our predictions for about a quarter of the total length of our time sample.

To compare the accuracy of the predictions from the various competing methods, we will use the average root mean squared forecast error (RMSFE) as the criterion. This will be the square root of the time-averaged squared prediction errors, averaged across all states $i = 1, ..., N$. The RMSFE will be tracked across multiple nowcast dates, v, and is defined as follows, denoting that T is the last year in the sample and we have P out-of-sample predictions made:

$$RMSFE_v = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{P}\sum_{t=T-P+1}^{T}\widehat{\varepsilon}_{v,i,t}^2} \quad (6)$$

where $\widehat{\varepsilon}_{v,i,t}$ generically stands for the prediction error of a model on nowcast date v for state i and year t.

We will also perform some analysis of the RMSFE for each state, where we do not average over the states. In other words we take the RMSFE for state i on nowcast date v as:

$$RMSFE_{vi} = \sqrt{\frac{1}{P}\sum_{t=T-P+1}^{T}\widehat{\varepsilon}_{v,i,t}^2} \quad (7)$$

where, of course, these results are only indicative as they are based on rather a small time series sample size and will be treated with some caution.

5 Results

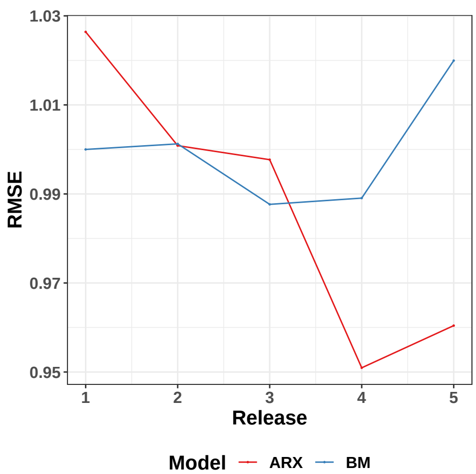
In this section, we discuss the results of the pseudo-out-of-sample experiment described in the previous section. We first discuss the accuracy of the EC predictions before then turning to the accuracy of the bridge equation method results for CO_2 emissions. For these accuracy assessments for EC and CO_2, we analyse both the annual data set-up and the quarterly data set-up as described above. We present results only for the original EC and CO_2 growth series, with the per-capita growth being reported in the Appendix.[12] The findings are very similar between the main results and the per-capita results.

[12] In arriving at the per-capita figures for the quarterly series, the population is assumed to remain constant for all four quarters of any year and is equal to the annual number.

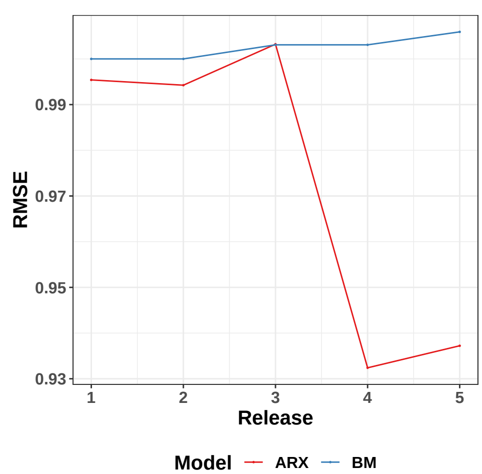
5.1 Energy Consumption Results

5.1.1 Energy Consumption Predictions with the Annual Data Flow

Figure 1 displays the RMSFEs obtained from predicting the growth rates of EC according to the release schedule in Table 1, where the economic data are used at the annual frequency. In all figures, the RMSFEs have been normalised by the RMSFE of the benchmark in the first nowcast period so that any figures lower than 1 are gains relative to the benchmark in the first period. These results show that, on average across all states, there is a drop in the RMSFE from the ARX model when the model also includes current economic data, both for GDP and PI. From Table 1, we noted earlier that there are only two annual economic data releases, which occur in releases two and four. While release two, corresponding to the year lagged economic data, is not able to improve the RMSFE of the ARX model in comparison to the benchmark, release four shows that the ARX model improves over the benchmark on the release of the up-to-date economic data. We see a sharper drop in the RMSFE based on the PI data relative to the GDP data. We note that the GDP data only have a relatively short history, starting in 1997, so the results based on PI appear to be more reliable.



(a) Predictor: GDP



(b) Predictor: PI

Figure 1: RMSFE - Energy Consumption - Annual Data Flow

Notes: The RMSFE figures are normalised by the benchmark ("BM") at the first release date. Therefore any points below 1 indicate that the RMSFE is lower than that of the benchmark in the first nowcast period.

While the average RMSFE results across states show a quantitatively modest improvement over the benchmark after economic data have been released (gains of 5-6% in both cases), when we dig into the individual states we see much more substantial improvements of our method in some of the larger states such as Florida, with gains of up to 30%. To summarise the results across states, Table 3 presents the quantiles of the state-specific RMSFEs for the ARX model relative to the benchmark model. In general the table confirms what is seen in Figure 1 and we see that at releases 4 and 5 there are gains from the ARX model relative to the benchmark across the majority of states. Additionally, we see that gains at the later nowcast updates are as large as 20% over the benchmark in several states at the 10th percentile, both for the GDP and PI models. The gain is in the region of 10% to 15% at the 25th percentile.

Release	**10%**	**25%**	**50%**	**75%**	**90%**
1	0.9767	0.9949	1.0226	1.0377	1.1139
2	0.9537	0.9678	1.0009	1.0218	1.0494
3	0.9534	0.9786	1.0129	1.0367	1.0690
4	0.8556	0.9106	0.9485	1.0070	1.0905
5	0.8125	0.8743	0.9353	0.9961	1.0908

(a) Predictor - GDP

Release	**10%**	**25%**	**50%**	**75%**	**90%**
1	0.9951	1.0028	1.0107	1.0181	1.0312
2	0.9614	0.9978	1.0218	1.0388	1.0519
3	0.9575	0.9876	1.0167	1.0317	1.0465
4	0.8145	0.9038	0.9429	0.9914	1.0240
5	0.8334	0.9044	0.9403	0.9923	1.0372

(b) Predictor - PI

Table 3: Distribution of Relative RMSFE Across States - Energy Consumption - Annual Data Flow

Notes: The numbers represent the quantiles of the distribution of relative RMSFE across states, where we take the RMSFE of the ARX model relative to the benchmark. Figures lower than 1 indicate that the RMSFE of the ARX model was lower than that of the benchmark for all of the countries below the relevant quantile.

The naïve benchmark method, on the other hand, does not improve even as newer relevant information is added in calculating the historic mean. If anything, the results seem to worsen as the data for EC gets released and is included in the predictions. This is more evident from Figure 1b where we use the entire available history of EC growth rates since 1961.

In summary, we find that releases of current economic data yield improvements in predicting growth rates of EC. The improvement is modest on average across all states, and rather large in some of the most energy-consuming states. Looking at the performance across all states indicates that the proposed method is capable of delivering nowcast accuracy gains in a non-trivial number of states once relevant economic data are included in the model. We find that the backcast made in March of the year after the reference year (release four) is of particular use. This is available well over a year in advance of the release of the EC data, and so we are able to make timeliness gains even using this example with annual economic data.

5.1.2 Energy Consumption Predictions with the Quarterly Data Flow

Now we present an assessment of the nowcast and backcast predictions of EC growth using the mixed frequency version of the model in equation 2. In this case we update the dataset following Table 2 using quarterly frequency PI data. The quarterly state-level GDP data starts only in 2005 and hence leaves us with too few observations for estimating and evaluating the models. Therefore, we do not include state-level GDP in the mixed-frequency

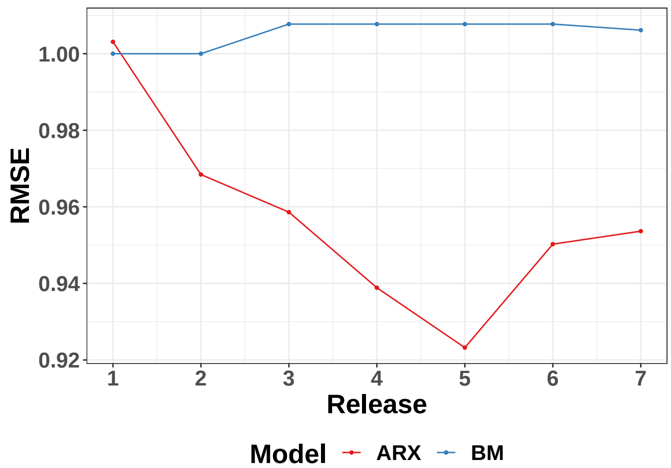


Figure 2: Average RMSFE Across States - Energy Consumption - Quarterly Data Flow

Notes: Same as Figure 1 but we note that only PI data are used in the quarterly data flow set-up due to the short time span of quarterly state-level GDP data.

analysis.

Figure 2 shows the average RMSFE across states from the mixed frequency panel ARX model, in contrast to the naïve benchmark. As in the annual case above, we find a noteworthy drop in RMSFE once the PI data for successive quarters of the target year starts to get released and is included in the model. The RMSFE gains relative to the benchmark are as much as 10% on average across states, which is somewhat larger than that when using annual data. Furthermore, in the annual model the drop in RMSFE could be observed only after all four quarter's data have been released. In the mixed frequency case we notice falling RMSFE right from the release of the first quarter of data (release date 3 in Table 2). By the end of the prediction period, while the benchmark does not improve at all, the mixed frequency ARX model has shown improvements using economic data.

As with the annual results, we also display the distribution of the relative RMSFE across quantiles, which can be seen in Table 4. Here we see that there are nowcast accuracy gains of up to 25% in the best 10th percentile of states, which is even larger that in the annual case, with the added benefit that the quarterly predictions can be derived in a more timely fashion. Even at the 25th percentile, there are gains of around 15% from using the ARX model relative to the benchmark, once sufficient data have been added into the model.

Overall, these quarterly results show an improvement over the annual results both in terms of the relative gain of the ARX compared to the benchmark, but especially due to their additional timeliness. Since we start to get the quarterly information on the nowcast year in around June of the same year, we can see improvements in RMSFE around two years before the EC data are published.

Release	10%	25%	50%	75%	90%
1	0.9494	0.9792	0.9991	1.0250	1.0566
2	0.8984	0.9336	0.9728	1.0034	1.0254
3	0.8644	0.9009	0.9598	0.9918	1.0408
4	0.7891	0.8726	0.9344	0.9888	1.0399
5	0.7961	0.8745	0.9112	0.9608	1.0340
6	0.8038	0.9013	0.9607	0.9950	1.0266
7	0.8114	0.9054	0.9667	1.0010	1.0270

Table 4: Distribution of Relative RMSFE Across States - Energy Consumption - Quarterly Data Flow

Notes: Same as for Table 3.

5.2 CO_2 Emissions Results

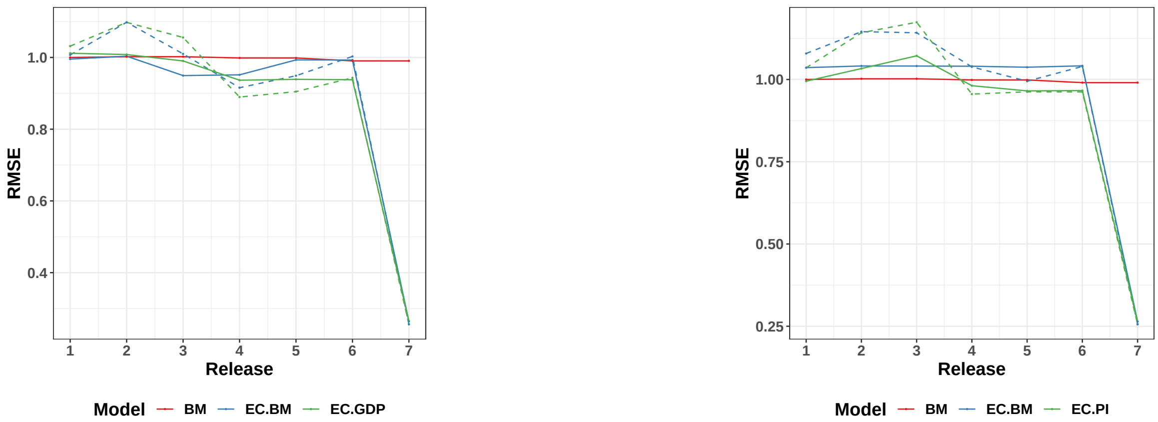
Now having the predictions of the EC for the target year, we can proceed to predict the CO_2 emissions growth rate using the bridge equation model in equation (3). We do this in two parts as earlier, first evaluating the nowcasts when only the annual economic data are incorporated into the EC nowcasts, and then allowing for quarterly economic data to be used. As before, both PI and GDP are used in the annual frequency results whereas only PI is used in the quarterly results.

5.2.1 CO_2 Emissions Predictions with the Annual Data Flow

Figure 3 displays the results of the bridge equation method based on energy consumption nowcasts from either the ARX or benchmark method (as in Figure 1 above). In a similar way to above, there is improvement in predictive accuracy from the fourth data release onwards, in other words when the economic data for the target year is released. However, these gains are less obvious than in the case of energy consumption. Some gains of up to 10% on average can be seen when bridging using the predictions of EC including GDP data ("EC.GDP") which improves more than when predicting EC with the benchmark method ("EC.BM") and no economic data. We note that the addition of factors in the bridge equation model (displayed with dashed lines) does yield some minor improvements but these are somewhat marginal.

The most striking finding is the very sharp drop of almost 75% at the final release date when we incorporate the actual observed EC data into the bridge equation model. This clearly makes sense as the CO_2 data are derived from energy consumption, however it is noteworthy that we are able to generate good predictions many months before the CO_2 data are released, even when using a simple panel data regression model which is far simpler than the methodology used to construct the actual CO_2 data.

Tables 5 and 6 present the relative RMSFE distributions across states, for all of the models considered in Figure 3. As above, these tables reveal more information about the performance of the bridge equation method than looking at the average across all states. For instance, in the case where the EC nowcasts used in the bridge equation have been derived from GDP data (tables 5a and 5b), we see gains of over 20% relative to the benchmark



(a) Predictor: State-Level GDP

(b) Predictor: State-Level PI

Figure 3: Average RMSFE Across States - CO_2 Emissions - Annual Data Flow

Notes: Dashed lines indicate that factors were used in the CO_2 model. EC.BM: bridge equation predictions for CO_2, benchmark model for EC; EC.GDP/EC.PI: bridge equation predictions for CO_2, GDP/PI model for EC. The RMSFE is normalised on the benchmark in the first nowcast period as in previous figures.

in the top 10th percentile at nowcast point four when recent economic data are available. We also notice a similar pattern to the average results above when looking across all percentiles, with a sudden drop in the RMSFE of the bridge equation method (relative to the benchmark) at the end of the prediction period when the current year's energy consumption data are released.

5.2.2 CO_2 Emissions Predictions with the Quarterly Data Flow

Finally, we perform the evaluation of the CO_2 predictions where the quarterly data were used in the EC predictions. These are, again, only performed with PI as the economic indicator as in Figure 2. The results of the pseudo-out-of-sample experiment (Figure 4) are comparable to those of the annual results discussed earlier. As with the annual CO_2 results using the PI model for the EC nowcasts, the gains on average are not very large until the release of the current year's EC data which improves the predictive accuracy remarkably.

Table 7 shows, in a similar way to above, that if we dig down into the quantiles of the relative RMSFE across states, then there are some sizeable RMSFE gains relative to the benchmark even when quite early on in the nowcast period. These gains are as large as 15% in the case where the EC.PI bridge model is used with factors (7b). However, in general the findings tend to show that is fairly difficult to improve much over the benchmark in predicting CO_2 emissions until the point at which energy consumption data become available.[13] As mentioned above, this still presents an opportunity to obtain reliable CO_2 nowcasts several months before the statistical authority releases the actual data.

[13] We note that formal statistical testing of the relative predictive accuracy is not available in our context with only 10 out-of-sample observations, where the power of Diebold-Mariano type tests will be very low.

Release	10%	25%	50%	75%	90%
1	0.8875	0.9383	1.0207	1.0747	1.1628
2	0.9350	0.9680	1.0116	1.0476	1.0872
3	0.8872	0.9250	0.9818	1.0395	1.1152
4	0.8650	0.9028	0.9410	0.9799	1.0202
5	0.8240	0.8752	0.9327	1.0075	1.0943
6	0.8431	0.8865	0.9410	0.9974	1.0940
7	0.1764	0.2134	0.2650	0.3249	0.4000

(a) Model: EC.GDP

Release	10%	25%	50%	75%	90%
1	0.8594	0.9449	1.0414	1.1007	1.2016
2	1.0124	1.0465	1.1048	1.1718	1.2197
3	0.9363	1.0083	1.0575	1.1064	1.1672
4	0.7953	0.8277	0.8916	0.9518	1.0288
5	0.7715	0.8127	0.8933	1.0009	1.1013
6	0.8439	0.8846	0.9381	0.9985	1.1116
7	0.1757	0.2003	0.2470	0.3114	0.3955

(b) Model: EC.GDP with factors

Release	10%	25%	50%	75%	90%
1	0.9072	0.9542	0.9991	1.0485	1.0924
2	0.9145	0.9561	1.0042	1.0595	1.0908
3	0.8889	0.9183	0.9410	0.9837	1.0178
4	0.8945	0.9149	0.9522	0.9957	1.0555
5	0.9402	0.9651	0.9964	1.0252	1.0691
6	0.9481	0.9722	1.0025	1.0361	1.0870
7	0.1764	0.2134	0.2650	0.3249	0.4000

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9124	0.9704	1.0093	1.0510	1.1082
2	0.9792	1.0487	1.0965	1.1786	1.2236
3	0.9351	0.9739	1.0154	1.0484	1.0942
4	0.8389	0.8684	0.9005	0.9744	1.0285
5	0.8465	0.9049	0.9436	1.0028	1.0699
6	0.9437	0.9776	1.0133	1.0567	1.0904
7	0.1757	0.2003	0.2470	0.3114	0.3955

(d) Model: EC.BM with factors

Table 5: Distribution of Relative RMSFE Across States - CO2 Emissions - Annual Data Flow (Predictor: GDP)

Notes: The numbers represent the quantiles of the distribution of relative RMSFE across states, where we take the RMSFE of the bridge equation model relative to the benchmark. Figures lower than 1 indicate that the RMSFE of the bridge equation model was lower than that of the benchmark for all of the countries below the relevant quantile. Results are presented for different methods of computing the EC forecasts (EC.GDP and EC.BM) as well as with and without factors.

Release	10%	25%	50%	75%	90%
1	0.8818	0.9232	0.9549	1.0255	1.1692
2	0.9287	0.9750	1.0280	1.0661	1.1906
3	0.9236	1.0140	1.0717	1.1233	1.2499
4	0.8935	0.9490	0.9855	1.0227	1.0466
5	0.8714	0.9251	0.9743	1.0209	1.0411
6	0.8935	0.9477	0.9852	1.0191	1.0401
7	0.1764	0.2134	0.2650	0.3249	0.4000

(a) Model: EC.PI

Release	10%	25%	50%	75%	90%
1	0.9147	0.9573	0.9972	1.0672	1.2872
2	0.9831	1.0769	1.1451	1.1990	1.3388
3	0.9879	1.0815	1.1846	1.2547	1.4261
4	0.8608	0.8951	0.9470	1.0099	1.0981
5	0.8430	0.8827	0.9751	1.0322	1.1459
6	0.8901	0.9449	0.9821	1.0124	1.0389
7	0.1757	0.2003	0.2470	0.3114	0.3955

(b) Model: EC.PI with Factors

Release	10%	25%	50%	75%	90%
1	0.9419	0.9613	0.9908	1.0348	1.0838
2	0.9425	0.9710	0.9948	1.0398	1.0889
3	0.9767	1.0000	1.0132	1.0523	1.1074
4	0.9902	1.0053	1.0232	1.0480	1.1119
5	0.9812	0.9929	1.0147	1.0471	1.0924
6	0.9948	1.0027	1.0214	1.0545	1.0971
7	0.1764	0.2134	0.2650	0.3249	0.4000

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9715	0.9969	1.0197	1.0711	1.1963
2	1.0217	1.0502	1.0856	1.1521	1.2017
3	1.0300	1.0635	1.1194	1.1656	1.2676
4	0.9565	0.9720	0.9994	1.0602	1.1214
5	0.9112	0.9460	0.9779	1.0251	1.0966
6	0.9891	1.0022	1.0340	1.0520	1.0817
7	0.1757	0.2003	0.2470	0.3114	0.3955

(d) Model: EC.BM with factors

Table 6: Distribution of Relative RMSFE Across States - CO2 Emissions - Annual Data Flow (Predictor: PI)

Notes: Same as Table 5 but with EC.PI instead of EC.GDP. The estimation sample is larger for the PI results compared with the GDP results due to the data availability. See text for further details.

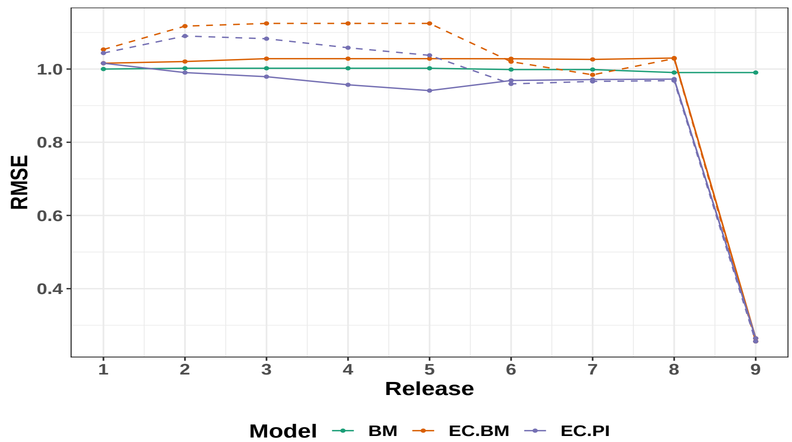


Figure 4: Average RMSFE Across States - CO_2 Emissions - Quarterly Data Flow

Notes: Same as for Figure 3.

Release	10%	25%	50%	75%	90%
1	0.9523	0.9724	1.0058	1.0368	1.0913
2	0.9042	0.9386	0.9839	1.0359	1.0704
3	0.9022	0.9262	0.9644	1.0122	1.0685
4	0.8531	0.9056	0.9528	0.9966	1.0384
5	0.8414	0.8910	0.9409	0.9771	1.0136
6	0.8739	0.9355	0.9743	1.0111	1.0487
7	0.8821	0.9347	0.9912	1.0165	1.0512
8	0.9104	0.9625	0.9931	1.0226	1.0394
9	0.1764	0.2134	0.2650	0.3249	0.4000

(a) Model: EC.PI

Release	10%	25%	50%	75%	90%
1	0.9587	0.9964	1.0347	1.0773	1.1311
2	0.9757	1.0380	1.0755	1.1467	1.2402
3	0.9772	1.0215	1.0706	1.1434	1.2528
4	0.9372	0.9955	1.0743	1.1154	1.1601
5	0.9266	0.9866	1.0409	1.0933	1.1239
6	0.8460	0.9205	0.9641	1.0083	1.0841
7	0.8498	0.9052	0.9812	1.0184	1.1325
8	0.8956	0.9587	0.9840	1.0159	1.0364
9	0.1757	0.2003	0.2470	0.3114	0.3955

(b) Model: EC.PI with factors

Release	10%	25%	50%	75%	90%
1	0.9627	0.9736	0.9990	1.0203	1.0933
2	0.9625	0.9821	0.9997	1.0286	1.1000
3	0.9698	0.9896	1.0066	1.0366	1.1060
4	0.9698	0.9896	1.0066	1.0366	1.1060
5	0.9698	0.9896	1.0066	1.0366	1.1060
6	0.9802	0.9930	1.0145	1.0423	1.1302
7	0.9763	0.9950	1.0095	1.0476	1.1043
8	0.9912	1.0017	1.0242	1.0630	1.1099
9	0.1764	0.2134	0.2650	0.3249	0.4000

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9823	1.0039	1.0269	1.0891	1.1372
2	1.0190	1.0628	1.1133	1.1554	1.2464
3	1.0246	1.0684	1.1173	1.1615	1.2413
4	1.0246	1.0684	1.1173	1.1615	1.2413
5	1.0246	1.0684	1.1173	1.1615	1.2413
6	0.9444	0.9606	0.9993	1.0706	1.1483
7	0.9041	0.9404	0.9709	1.0214	1.0878
8	0.9890	1.0053	1.0268	1.0591	1.1028
9	0.1757	0.2003	0.2470	0.3114	0.3955

(d) Model: EC.BM with factors

Table 7: Distribution of Relative RMSFE Across States - CO_2 Emissions - Quarterly Data Flow

Notes: Same as Table 5.

5.3 Further Results

We also explored the robustness of these empirical results to a number of additional checks, the results of which we display in the Appendix. Firstly, we re-ran all results of the paper using the per capita energy consumption and CO_2 data. The results in Appendix A and B demonstrate very little difference to the results reported in the main text which indicates that the same results hold if we use the per capita or level figures when computing the growth rates. Secondly, we performed an additional set of results to explore the robustness to the sample split used in generating the out-of-sample predictions. In Figure 9 in Appendix C, the evaluation sample 2000-2018 is compared to that of 2009-2018, showing that the results are indeed stable over time. In terms of the models we run and the variables used, we attempted several additional checks. Figure 10 displays results using the Philly Fed's state coincident index which is like a principal component from a set of state-level employment series. The results are no better than the main results where we use GDP or PI as predictors. Figure 11 shows that the CO_2 nowcasts are worse when we use the economic variables directly instead of through the energy consumption bridging variable, and this direct model cannot pick up the large drop in RMSFE we see at the end of the sample on the release of the EC data. Finally, Figure 12 shows that the results are not improved by combining both GDP and PI in the same model instead of using them individually.

6 Conclusion

This paper has proposed methods for obtaining timely predictions of U.S. state-level energy consumption and CO_2 emissions growth. Motivated by the very long publication lags for these variables, we use the flow of more timely economic data to make nowcasts and backcasts. Our contribution is a first step in the direction of making real time predictions of sub-national variables related to environmental degradation. We have moved the focus of existing panel nowcasting studies away from the classic GDP and macroeconomic nowcasting setting.

Our empirical study produces historic out-of-sample nowcasts of state-level energy consumption growth and CO_2 emissions growth, from which we draw the following conclusions. Firstly, we conclude that the use of timely economic data can give important improvements in predicting energy consumption growth on average across all states, and can deliver especially large gains in a smaller group of states including larger ones such as Florida. These predictive gains can occur almost two years before the energy consumption data are released. On the other hand, we conclude that the CO_2 predictions are less affected by the release of economic data and that it is better to wait until the release of the current year's energy consumption data, at which point a very accurate prediction can be made. This is, nevertheless, able to produce a reliable CO_2 prediction many months before the statistical authority releases the data and using a method which is far simpler.

There is still much more work to be done on state-level energy and CO_2 nowcasting. With the 'big data' revolution increasing the granularity of available data, it would be useful to see our method perform with a more complete dataset. An interesting example would be to assess whether firm-level emissions data can be aggregated in a timely fashion for the purpose of predicting state-level emissions.

References

Anesti, N., A. B. Galvao, and S. Miranda-Agrippino (2022). Uncertain Kingdom: nowcasting GDP and its revisions. *Journal of Applied Econometrics 37*(1), 42–62.

Babii, A., R. T. Ball, E. Ghysels, and J. Striaukas (2020). Machine learning panel data regressions with an application to nowcasting price earnings ratios. *arXiv:2008.03600v1*.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2A, pp. 195–236. Amsterdam: North-Holland.

Bennedsen, M., E. Hillebrand, and S. J. Koopman (2021). Modeling, forecasting, and nowcasting US CO2 emissions using many macroeconomic predictors. *Energy Economics 96*, 105118.

Bok, B., D. Caratelli, D. Giannone, A. M. Sbordone, and A. Tambalotti (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review of Economics 10*(1), 615–643.

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

Choi, I. (2001). Unit root tests for panel data. *Journal of international money and Finance 20*(2), 249–272.

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

Dahlhaus, T., J.-D. Guénette, and G. Vasishtha (2017). Nowcasting BRIC+M in Real Time. *International Journal of Forecasting 33*(4), 915–935.

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

Foroni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting 30*(3), 554–568.

Foroni, C., M. Marcellino, and C. Schumacher (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. and R. Greenaway-McGrevy (2022). Panel Data Nowcasting. *Econometric Reviews (forthcoming)*.

Fosten, J. and D. Gutknecht (2020). Testing Nowcast Monotonicity with Estimated Factors. *Journal of Business & Economic Statistics 38*(1), 107–123.

Fosten, J. and S. Nandi (2021). Nowcasting from Cross-Sectionally Dependent Panels. *DAFM Working paper No. 2021/9*.

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

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

Hultman, N. E., L. Clarke, C. Frisch, K. Kennedy, H. McJeon, T. Cyrs, P. Hansel, P. Bodnar, M. Manion, M. R. Edwards, and Others (2020). Fusing subnational with national climate action is central to decarbonization: the case of the United States. *Nature Communications 11*(1), 1–10.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*(1), 53–74.

Knotek, E. S. and S. Zaman (2017). Nowcasting US Headline and Core Inflation. *Journal of Money, Credit and Banking 49*(5), 931–968.

Koop, G., S. McIntyre, and J. Mitchell (2020). UK regional nowcasting using a mixed frequency vector autoregressive model with entropic tilting. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 183*(1), 91–119.

Larson, W. D. and T. M. Sinclair (2022). Nowcasting unemployment insurance claims in the time of COVID-19. *International Journal of Forecasting 38*(2), 635–647.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics 108*(1), 1–24.

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

Nordhaus, W. D. (1992). An optimal transition path for controlling greenhouse gases. *Science 258*(5086), 1315–1319.

Pesaran, M. H. (2016). *Time series and panel data econometrics* (First ed.). Oxford: Oxford University Press.

Rennert, K., B. C. Prest, W. A. Pizer, R. G. Newell, D. Anthoff, C. Kingdon, L. Rennels, R. Cooke, A. E. Raftery, and H. Ševčiková (2021). The social cost of carbon: Advances in long-term probabilistic projections of population, GDP, emissions, and discount rates. *Brookings Papers on Economic Activity Fall*, 223–275.

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

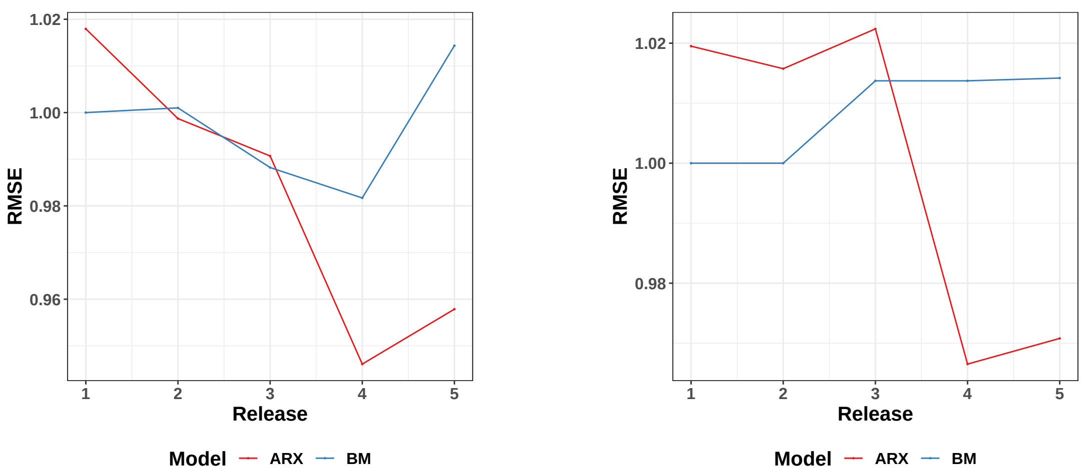
Wang, W., X. Zhang, and R. Paap (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics 34*(5), 724–745.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067–1084.

Appendix A Per Capita Energy Consumption Results

A.1 Per Capita Energy Consumption Predictions with the Annual Data Flow

A.1.1 Overall Results



(a) Predictor: Per Capita GDP

(b) Predictor: Per Capita PI

Figure 5: Average RMSFE Across States - Per Capita Energy Consumption - Annual Data Flow

A.1.2 State-Level Results

Release	10%	25%	50%	75%	90%
1	0.9668	0.9929	1.0182	1.0379	1.0585
2	0.9604	0.9717	1.0055	1.0180	1.0231
3	0.9568	0.9780	1.0020	1.0243	1.0575
4	0.8845	0.9092	0.9439	1.0122	1.0539
5	0.8514	0.8847	0.9324	0.9977	1.0568

(a) Predictor - Per Capita GDP

Release	10%	25%	50%	75%	90%
1	0.9994	1.0080	1.0176	1.0281	1.0478
2	0.9836	1.0108	1.0213	1.0304	1.0405
3	0.9819	0.9965	1.0124	1.0220	1.0291
4	0.8103	0.9134	0.9628	1.0020	1.0499
5	0.8456	0.9120	0.9669	1.0053	1.0507

(b) Predictor - Per Capita PI

Table 8: Distribution of Relative RMSFE Across States - Per Capita Energy Consumption - Annual Data Flow

A.2 Per Capita Energy Consumption Predictions with the Quarterly Data Flow

A.2.1 Overall Results

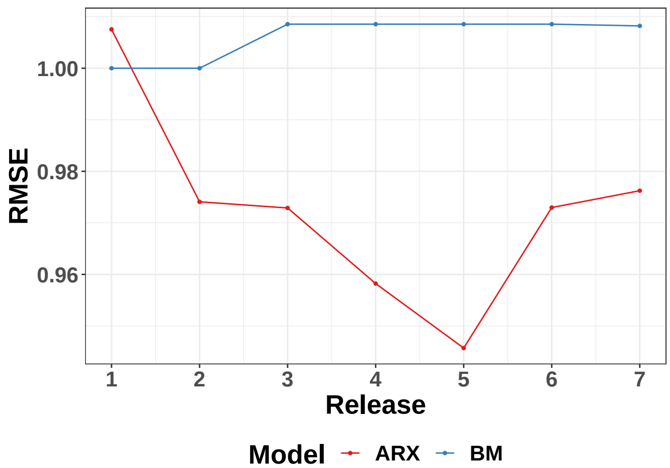


Figure 6: Average RMSFE Across States - Per Capita Energy Consumption - Quarterly Data Flow

A.2.2 State-Level Results

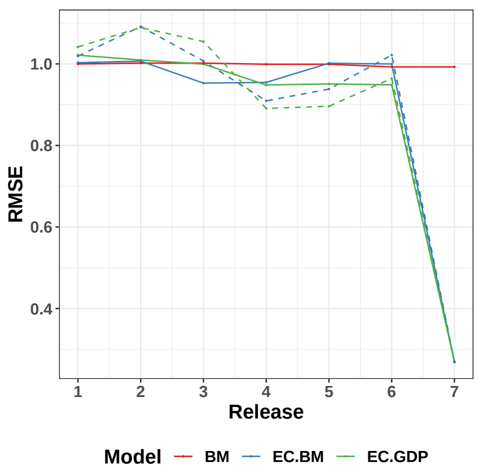
Release	10%	25%	50%	75%	90%
1	0.9583	0.9894	1.0075	1.0318	1.0692
2	0.9236	0.9459	0.9731	1.0035	1.0188
3	0.8804	0.9144	0.9676	1.0069	1.0399
4	0.8300	0.9027	0.9490	0.9953	1.0666
5	0.8560	0.8965	0.9432	0.9840	1.0283
6	0.8461	0.9217	0.9751	1.0096	1.0559
7	0.8602	0.9349	0.9728	1.0103	1.0458

Table 9: Distribution of Relative RMSFE Across States - Per Capita Energy Consumption - Quarterly Data Flow

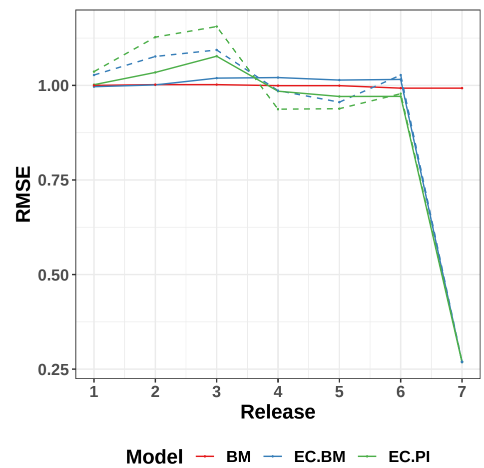
Appendix B Per Capita CO_2 Emissions Results

B.1 Per Capita CO_2 Predictions with the Annual Data Flow

B.1.1 Overall Results



(a) Predictor: Per Capita GDP



(b) Predictor: Per Capita PI

Figure 7: Average RMSFE Across States - Per Capita CO_2 Emissions - Annual Data Flow

B.1.2 State-Level Results

Release	**10%**	**25%**	**50%**	**75%**	**90%**
1	0.8676	0.9594	1.0243	1.0882	1.1671
2	0.9232	0.9747	1.0069	1.0558	1.1061
3	0.8989	0.9278	0.9894	1.0755	1.1335
4	0.8831	0.9129	0.9410	0.9916	1.0487
5	0.8359	0.8853	0.9509	1.0191	1.0867
6	0.8583	0.8928	0.9611	0.9971	1.0843
7	0.1792	0.2128	0.2653	0.3309	0.3962

(a) Model: EC.GDP

Release	10%	25%	50%	75%	90%
1	0.9151	0.9726	1.0326	1.1002	1.1986
2	0.9751	1.0265	1.0910	1.1601	1.2149
3	0.9608	1.0038	1.0541	1.0991	1.1855
4	0.7959	0.8185	0.8745	0.9498	1.0335
5	0.7486	0.8095	0.8936	0.9941	1.0850
6	0.8589	0.9130	0.9679	1.0283	1.1120
7	0.1820	0.2144	0.2641	0.3268	0.3976

(b) Model: EC.GDP with factors

Release	10%	25%	50%	75%	90%
1	0.8987	0.9573	1.0075	1.0505	1.1261
2	0.9152	0.9579	1.0103	1.0565	1.1201
3	0.8734	0.9083	0.9416	1.0052	1.0686
4	0.8834	0.9160	0.9530	1.0026	1.0596
5	0.9204	0.9706	1.0047	1.0356	1.0834
6	0.9292	0.9675	1.0068	1.0486	1.0985
7	0.1792	0.2128	0.2653	0.3309	0.3962

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9082	0.9631	1.0328	1.0740	1.1489
2	0.9570	1.0177	1.0961	1.1655	1.2432
3	0.8974	0.9619	1.0043	1.0446	1.1096
4	0.8135	0.8453	0.8978	0.9647	1.0529
5	0.8556	0.8879	0.9172	0.9742	1.1293
6	0.9445	0.9808	1.0264	1.0752	1.1358
7	0.1820	0.2144	0.2641	0.3268	0.3976

(d) Model: EC.BM with factors

Table 10: Distribution of Relative RMSFE Across States - Per Capita CO_2 emissions - Annual Data Flow (Predictor: GDP)

Release	10%	25%	50%	75%	90%
1	0.9011	0.9313	0.9728	1.0437	1.1550
2	0.9528	0.9926	1.0329	1.0593	1.1229
3	0.9434	1.0404	1.0832	1.1343	1.1925
4	0.8880	0.9556	0.9908	1.0198	1.0527
5	0.8773	0.9247	0.9708	1.0250	1.0720
6	0.9076	0.9437	0.9822	1.0159	1.0625
7	0.1792	0.2128	0.2653	0.3309	0.3962

(a) Model: EC.PI

Release	10%	25%	50%	75%	90%
1	0.9120	0.9684	0.9983	1.0694	1.2493
2	0.9980	1.0772	1.1284	1.1986	1.2498
3	0.9879	1.0891	1.1577	1.2335	1.3150
4	0.8492	0.8879	0.9291	0.9860	1.0616
5	0.8070	0.8782	0.9481	1.0020	1.1088
6	0.9101	0.9464	0.9915	1.0300	1.0540
7	0.1820	0.2144	0.2641	0.3268	0.3976

(b) Model: EC.PI with factors

Release	10%	25%	50%	75%	90%
1	0.9464	0.9575	0.9937	1.0162	1.0625
2	0.9449	0.9698	0.9952	1.0212	1.0512
3	0.9641	0.9954	1.0137	1.0344	1.0622
4	0.9885	0.9959	1.0173	1.0335	1.0719
5	0.9776	0.9888	1.0093	1.0358	1.0658
6	0.9879	0.9976	1.0099	1.0428	1.0645
7	0.1792	0.2128	0.2653	0.3309	0.3962

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9616	0.9882	1.0109	1.0483	1.1421
2	1.0069	1.0447	1.0762	1.1288	1.1521
3	1.0047	1.0517	1.0996	1.1349	1.1932
4	0.9301	0.9489	0.9756	1.0163	1.0450
5	0.8851	0.9203	0.9531	0.9960	1.0145
6	0.9934	1.0067	1.0390	1.0558	1.0739
7	0.1820	0.2144	0.2641	0.3268	0.3976

(d) Model: EC.BM with factors

Table 11: Distribution of Relative RMSFE Across States - Per Capita CO_2 emissions - Annual Data Flow (Predictor: PI)

B.2 Per Capita CO_2 Predictions with the Quarterly Data Flow

B.2.1 Overall Results

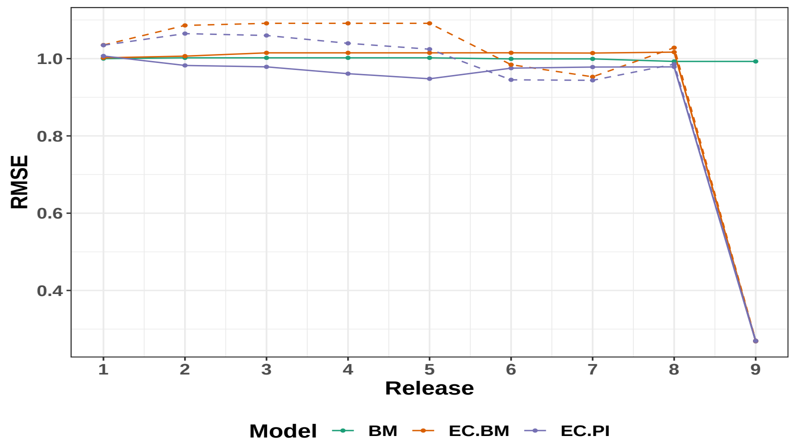


Figure 8: Average RMSFE Across States - Per Capita CO_2 Emissions - Quarterly Data Flow

B.2.2 State-Level Results

Release	**10%**	**25%**	**50%**	**75%**	**90%**
1	0.9569	0.9770	1.0012	1.0283	1.0670
2	0.9206	0.9404	0.9758	1.0166	1.0490
3	0.9000	0.9309	0.9656	1.0121	1.0627
4	0.8688	0.9058	0.9578	1.0087	1.0592
5	0.8530	0.8933	0.9494	0.9977	1.0309
6	0.8692	0.9384	0.9836	1.0257	1.0673
7	0.8934	0.9411	0.9837	1.0260	1.0560
8	0.9193	0.9633	0.9928	1.0269	1.0434
9	0.1792	0.2128	0.2653	0.3309	0.3962

(a) Model: EC.PI

Release	10%	25%	50%	75%	90%
1	0.9760	0.9971	1.0322	1.0586	1.1139
2	0.9845	1.0307	1.0658	1.1135	1.1707
3	0.9479	1.0127	1.0563	1.1220	1.1727
4	0.9524	0.9889	1.0358	1.0918	1.1586
5	0.9492	0.9832	1.0169	1.0708	1.1270
6	0.8492	0.8910	0.9513	0.9916	1.0642
7	0.8125	0.8831	0.9413	1.0105	1.1149
8	0.9361	0.9716	0.9995	1.0322	1.0455
9	0.1820	0.2144	0.2641	0.3268	0.3976

(b) Model: EC.PI with factors

Release	10%	25%	50%	75%	90%
1	0.9556	0.9745	0.9944	1.0099	1.0601
2	0.9506	0.9829	0.9993	1.0128	1.0631
3	0.9645	0.9868	1.0056	1.0328	1.0576
4	0.9645	0.9868	1.0056	1.0328	1.0576
5	0.9645	0.9868	1.0056	1.0328	1.0576
6	0.9752	0.9906	1.0062	1.0347	1.0722
7	0.9798	0.9920	1.0103	1.0270	1.0620
8	0.9903	0.9976	1.0110	1.0435	1.0613
9	0.1792	0.2128	0.2653	0.3309	0.3962

(c) Model: EC.BM

Release	10%	25%	50%	75%	90%
1	0.9692	1.0007	1.0213	1.0561	1.1356
2	1.0104	1.0472	1.0921	1.1319	1.1696
3	1.0109	1.0548	1.0970	1.1439	1.1765
4	1.0109	1.0548	1.0970	1.1439	1.1765
5	1.0109	1.0548	1.0970	1.1439	1.1765
6	0.9237	0.9384	0.9691	1.0145	1.0685
7	0.8827	0.9190	0.9498	0.9921	1.0364
8	0.9928	1.0071	1.0268	1.0600	1.0898
9	0.1820	0.2144	0.2641	0.3268	0.3976

(d) Model: EC.BM with factors

Table 12: Distribution of Relative RMSFE Across States - Per Capita CO_2 Emissions - Quarterly Data Flow

Appendix C Further Results

C.1 Robustness to Sample Split

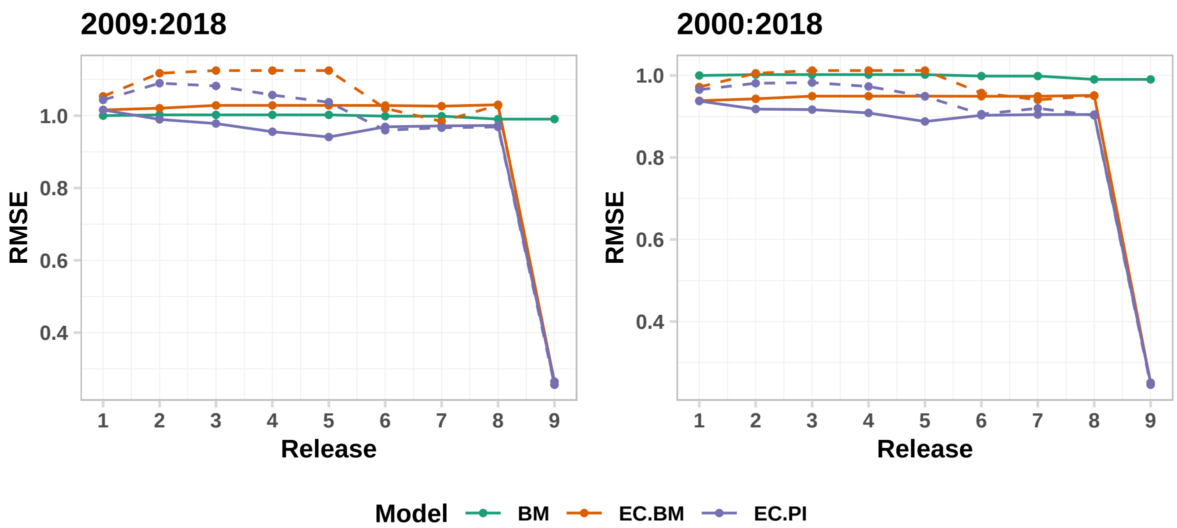


Figure 9: Sample Split - Average RMSFE Across States - CO_2 Emissions - Quarterly Data Flow

C.2 Using the Philly Fed's State Coincident Indexes (Quarterly)

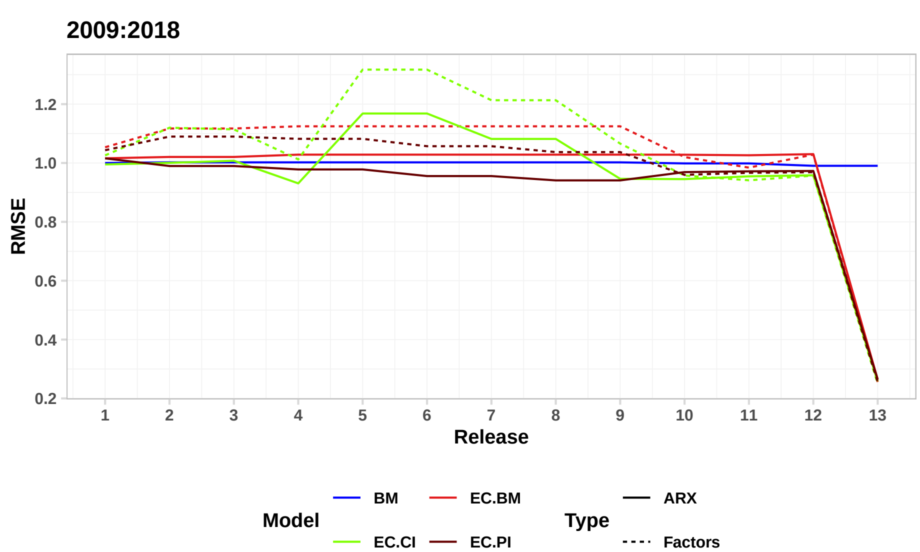


Figure 10: Average RMSFE Across States - CO_2 Emissions - Quarterly Data Flow

Notes: Same as Figure 4 with the addition of the model EC.CI which uses the Philly Fed's State Coincident index as a predictor.

C.3 Targeting CO_2 Emissions Directly Instead of Bridging

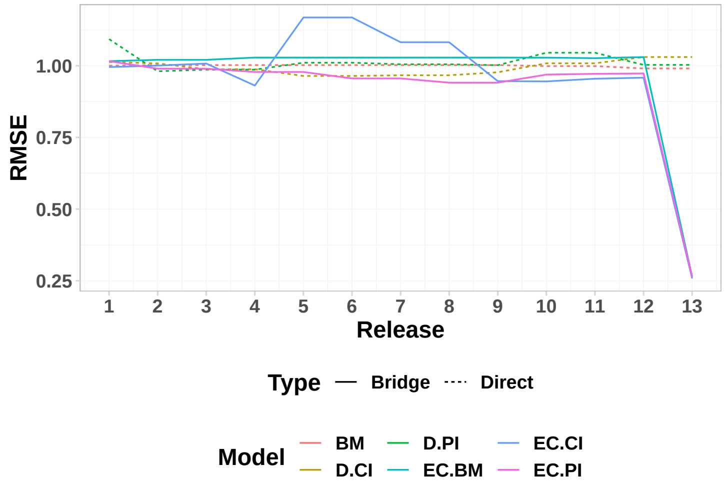


Figure 11: Targeting CO_2 Emissions Directly - Average RMSFE Across States - Quarterly Data Flow

Notes: Same as Figure 4 with the addition of the models D.PI and D.CI which directly predict CO_2 using PI or CI instead of through the bridging method.

C.4 Using Both GDP and PI in the Model (Annual Only)

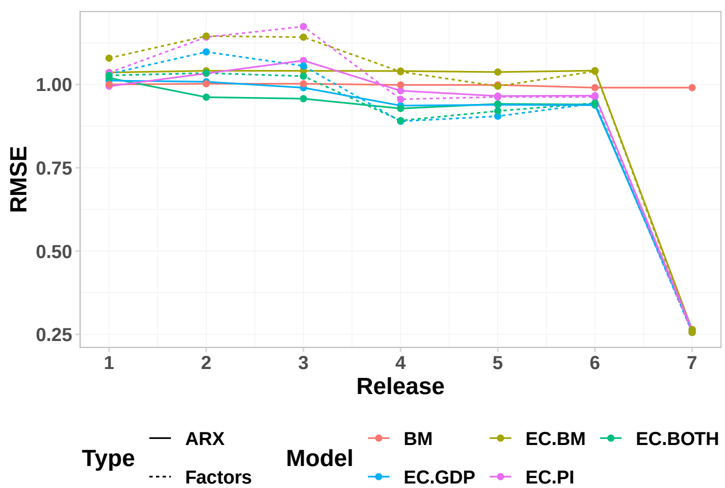


Figure 12: Average RMSFE Across States - CO_2 Emissions - Annual Data Flow

Notes: Same as Figure 3 with the addition of the model EC.BOTH which uses both PI and GDP in making the EC predictions for the bridge equation.

The Data Analytics for Finance and Macro (DAFM) Research Centre
Forecasting trends is more important now than it has ever been. Our quantitative financial and macroeconomic research helps central banks and statistical agencies better understand their markets. Having our base in central London means we can focus on issues that really matter to the City. Our emphasis on digitalisation and the use of big data analytics can help people understand economic behaviour and adapt to changes in the economy.

Find out more at kcl.ac.uk/dafm

KING'S BUSINESS SCHOOL

Data Analytics for Finance & Macro Research Centre