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

Jack Fosten[†‡] Shaoni Nandi[§¶]

November 13, 2023

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. We find that nowcasts improve when incorporating timely data like electricity consumption, relative to a simple benchmark. 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 also notable well in advance of the data release, especially after the current year's energy consumption data are used in making the prediction.

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 information regarding 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 are much longer than is typical in 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 and monthly 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 (2023), 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 monthly electricity sales growth or quarterly real personal income growth. Since these economic series are at a higher frequency and have a much lower publication lag, this makes them highly appropriate for making regularly updated nowcasts and backcasts. 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 noteworthy findings. We find that the predictions of energy consumption growth improve on average across states when current economic data are used for nowcasting and backcasting, relative to a naïve benchmark. The use of monthly electricity sales data is particularly successful, more so than real personal income growth. We also find particularly sizeable gains in several individual states, which we assess by looking at the across-state distribution of the gain in predictive accuracy of our model relative to the benchmark. Given the increased timeliness of the predictor variables, we see gains in predictive accuracy occurring over a year ahead of the release of the energy consumption data. These results carry over to our bridge equation predictions of CO_2 emissions growth. We find that the energy consumption nowcasts using electricity sales data provide nowcast gains for CO_2 emissions relative to a simple benchmark model. The gains are, again, sizeable in some states, and the biggest gains occur when we wait until the backcast period and add in the current year's observed energy consumption data. This means we are able to provide accurate predictions 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 direct fuel use across various sectors: commercial, industrial, residential and transportation. We focus on the

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

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. One notable exception is Higgins (2014) who does propose a bottom-up approach for GDP nowcasting in the *GDPNow* methodology at the Federal Reserve Bank of Atlanta. Whether this type of approach can be useful in our context is something we leave for future study.

2.2 Energy Consumption

The data for state-level aggregate energy consumption ("EC" hereafter) 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 EC 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

[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]

published at the end of June 2021. Although the data are more timely, if we wish to use the current year's EC in predicting CO_2 emissions, this would constitute a backcast and not a nowcast. In order to obtain nowcasts of EC 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 EC nowcasts, it is natural to use state-level economic indicators. With limited available state-level indicators at our disposal, we will present results using two different predictors: electricity consumption (as measured by sales of electricity to ultimate customers) and real personal income (PI).[7] The former is available from the EIA while the latter is available from the Bureau of Economic Analysis (BEA).[8] The electricity consumption data are available for all states at the monthly frequency with data starting in 1990. The PI data are available at a quarterly frequency from 1950 onwards, which we deflate by the GDP deflator for the U.S. to obtain real figures.

There are two factors which make these series appropriate for nowcasting EC and therefore CO_2 emissions. Firstly, their higher frequency makes them much timelier than the annual data, especially in the case of electricity sales where monthly observations are available. Secondly, for both electricity sales and PI, the publication lag is relatively low which means that quite early in the year we already start to observe relevant data points for use in nowcasting. In the case of electricity sales we observe a month's data around two months after the end of the relevant month. For PI the publication lag is around three months after the end of the reference quarter. This implies that already in the middle of the nowcast year, we have data on the first four months of electricity sales and the first quarter of PI data available for making predictions of EC 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. In previous versions of the paper we also used state-level real GDP data alongside PI, but this had a much shorter history of data available. Other studies such as Bennedsen et al. (2021) note that the industrial production (IP) index is useful in nowcasting national CO_2, but unfortunately IP data are not available by state. We also experimented in with other economic indices such as the Federal Reserve Bank of Philadelphia's State Coincident Indexes.[9] However these indices, available at a monthly level and constructed using a dynamic factor model on four state-level employment type series, did not fare well in our analysis and were ultimately discarded. There are also weekly state-level economic conditions available through Baumeister et al. (2022). We do not consider these here as they are not available for as long a history as the PI data, as well as causing issues in the econometric modelling of an annual-to-weekly frequency mix.

[7]Other studies using state-level electricity consumption in the context of economic activity in various countries include Baumeister et al. (2022); Furukawa et al. (2022); Lehmann and Möhrle (2022).

[8]See: https://www.eia.gov/electricity/data/state/ [Last accessed: 16/08/23] and https://www.bea.gov/data/income-saving/personal-income-by-state [Last accessed: 12/02/2022]

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

3 Panel MIDAS and Bridge Equation Methodology

In this section we describe the models we use to predict the annual growth of EC and subsequently of CO_2 emissions growth.[10] As mentioned above, the CO_2 data are released in March over two 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 monthly electricity sales or quarterly PI growth.

3.1 Panel MIDAS Model for Energy Consumption

We now describe the panel model for nowcasting EC growth using economic data. We adopt the notation that there are T annual observations on the target variable and there are N states available in the panel. Since there are differences in the frequencies of the economic data (monthly for electricity sales and quarterly for PI), we will first write down the model for the annual-to-monthly frequency mix and then the annual-to-quarterly frequency mix.

Mixed Frequency Model with Monthly Data

We start out by writing down the model which predicts EC using the available autoregressive lags on day v of the nowcast period as well as the available monthly lags of the economic indicator:

$$c_{i,t} = \alpha_{vi}^{(m)} + \phi_v^{(m)} c_{i,t-d_v} + \beta_v^{(m)\prime} \boldsymbol{x}_{i,t-\frac{k_v}{12}}^{(m)} + u_{v,i,t}^{(m)} \tag{1}$$

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

[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 EC 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) 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 $EC_{i,t}$. Alternatively, we also explore the results where $c_{i,t}$ is the growth 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.

This model is a panel version of the unrestricted MIDAS (UMIDAS) model and can be estimated by panel least squares in order to obtain conditional mean predictions for each individual (see Foroni et al., 2015; Schumacher, 2016). We denote $\boldsymbol{x}_{i,t-\frac{k_v}{12}} = \begin{pmatrix} x_{i,t-k_v/12}, & x_{i,t-(k_v-1)/12}, & \cdots & x_{i,t-(k_v-11)/12} \end{pmatrix}'$ as the stacked skip-sampled electricity sales growth which is inserted into the model with a monthly lag of k_v at nowcast date v. Note that a lag of one month is denoted in annual terms as $t - \frac{1}{12}$. In equation (1), the slope coefficient $\beta_v^{(m)}$ is a vector of length twelve, corresponding to the stacked skip-sampled process $\boldsymbol{x}_{i,t-\frac{k_v}{12}}$. We note that the superscript m is used for the parameters and error term in the equation to distinguish this from the quarterly mixed frequency model below.

The lag structure of 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_v as the available lag of $c_{i,t}$ at the time of prediction, based on its publication lag. Similarly, k_v is used to denote the available monthly lag of $\boldsymbol{x}_{i,t}$ 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, whereas the superscript m denotes the monthly model to differentiate from the quarterly model below. To give an example, in nowcasting year t, if v is the start of year t (in other words January 1st of year t), based on the data flow described in the Data section above, the model would use $c_{i,t-3}$ and $\boldsymbol{x}_{i,t-3/12}$ (monthly data to October of the previous calendar year). At the end of January the model would change in light of the electricity sales data release to include $\boldsymbol{x}_{i,t-2/12}$ (data to November), and so on, with lags of c and $\boldsymbol{x}$ being sequentially updated as v changes. The full details of the 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) in the context of the quarterly frequency of the PI data. The modification is to clarify the notation and time indices:

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

where $\boldsymbol{x}_{i,t-\frac{q_v}{4}}^{(q)} = \begin{pmatrix} x_{i,t-q_v/4}, & x_{i,t-(q_v-1)/4}, & x_{i,t-(q_v-2)/4}, & x_{i,t-(q_v-3)/4} \end{pmatrix}'$ denotes the stacked skip-sampled PI growth which is inserted into the model with a quarterly lag of q_v at nowcast date v.[11] Here a lag of one quarter is denoted in annual terms as $t - \frac{1}{4}$. In equation (2), the slope coefficient $\beta_v^{(q)}$ is a vector of length four, and the superscripts are changed to q to distinguish them from the model above. The nowcast updating works in the same

[11] We also experimented with empirical results where we first aggregated the monthly electricity sales data to the quarterly frequency and used equation (2) instead of equation (1). However, the results were not substantially different so we left the monthly model to include the data at the original frequency.

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 and quarterly data for $\boldsymbol{x}$.

The main difference between equations (1) and (2) is that the dimension is higher in the monthly skip-sampled UMIDAS model. In principle one could guard against parameter proliferation by introducing a lag weighting function with fewer parameters as in standard MIDAS models, for instance using exponential Almon lags. However, our empirical results below show that the monthly model works very well so we are not concerned with this issue.

Equations (1) and (2) are panel versions of the ARX model (AR with an exogenous regressor) and we refer to these 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.[12] Next, we use the EC predictions from the panel ARX to predict CO_2 emissions growth.

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 $c_{i,t}$ for state i in year t on date v of the nowcast period. The main equation is a panel bridge equation model with a multi-factor error structure:

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

where we define emissions growth, $e_{i,t}$, which either represents the growth of $CO_{2,i,t}$, the CO_2 emissions in state i in year t, or the growth of per-capita emissions $E_{i,t} = \frac{CO_{2,i,t}}{pop_{i,t}}$. 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 g_v. 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:

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

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

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

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 (2023) 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:

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

where $\overline{z}_{v,i,t}$ 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 $z_{v,i,t} = [e_{i,t-g_v}, \hat{c}_{v,i,t}]'$. Chudik and Pesaran (2015) and Fosten and Nandi (2023) 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 including 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 use a calendar to 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.[13]. This allows us to see how the nowcasts and backcasts behave, on average, as we add more information as 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

[13] We note that, due to the lack of available past vintages of the EC and CO_2 data we are not able to perform a fully real time analysis as in Sinclair and Stekler (2013) and, more recently, Anesti et al. (2022)

of the reference year, at the end of January. Moving through the nowcast year, we update the predictions every month and we continue into the backcast period until all relevant data have been released. The full data flow can be seen in Table 1 which gives an example of the calendar for predicting the year 2021 and the data available at each prediction date. Throughout the nowcast year, a new observation of electricity sales is added every month, and a new quarterly observation of PI data is added every three months in March, June, September and December. The annual data for past years of EC and CO_2 (specifically 2019 and 2018) are added in June and March respectively. Moving into the backcast period, we continue to add monthly electricity data until the end of February when all data up to December of the target year are available, and we add PI data until the last quarter is released in March (in other words we do not use "future" data when making nowcasts). In June of the first backcast year, the previous year's EC data are released, so we stop making predictions of EC (indicated by the horizontal line in Table 1). This gives a total of 18 months in the prediction period for EC. 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. There will be a total of 30 months in the prediction period for CO_2, although there are many months towards the end of the backcast period with no new data updates.

We therefore have multiple nowcasts and backcasts made for each target year for a total of nine years from 2009 to 2018. 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.[14] 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:

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

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:

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

However, these results should only be treated as indicative since they are based on rather a small time series sample size and we will treat these with some caution.

[14] We also tried a weighted RMSFE with different weights by state, but the results were qualitatively very similar. We discuss this later in the results section.

		Calendar Date (v)	Latest Available Data EC	CO_2	PI	ELEC
Nowcast	1	2021:M1	2018	2017	2020:Q3	2020:M11
	2	2021:M2	2018	2017	2020:Q3	2020:M12
	3	2021:M3	2018	2018	2020:Q4	2021:M1
	4	2021:M4	2018	2018	2020:Q4	2021:M2
	5	2021:M5	2018	2018	2020:Q4	2021:M3
	6	2021:M6	2019	2018	2021:Q1	2021:M4
	7	2021:M7	2019	2018	2021:Q1	2021:M5
	8	2021:M8	2019	2018	2021:Q1	2021:M6
	9	2021:M9	2019	2018	2020:Q2	2021:M7
	10	2021:M10	2019	2018	2020:Q2	2021:M8
	11	2021:M11	2019	2018	2020:Q2	2021:M9
	12	2021:M12	2019	2018	2020:Q3	2021:M10
Backcast	13	2022:M1	2019	2018	2020:Q3	2021:M11
	14	2022:M2	2019	2018	2020:Q3	2021:M12
	15	2022:M3	2019	2019	2020:Q4	
	16	2022:M4	2019	2019		
	17	2022:M5	2019	2019		
	18	2022:M6	2020	2019		
	19	2022:M7	2020	2019		
	20	2022:M8	2020	2019		
	21	2022:M9	2020	2019		
	22	2022:M10	2020	2019		
	23	2022:M11	2020	2019		
	24	2022:M12	2020	2019		
	25	2023:M1	2020	2019		
	26	2023:M2	2020	2019		
	27	2023:M3	2020	2020		
	28	2023:M4	2020	2020		
	29	2023:M5	2020	2020		
	30	2023:M6	2021	2020		

Table 1: Nowcast Calendar and Latest Available Data, Example for 2021

Notes: For each calendar month (v) from 2021:M1 through 2023:M6, this table displays the last available year of data for EC and CO_2, the last available quarter of data for PI and the last available month for ELEC. The horizontal line after release 18 denotes the point at which we stop predicting EC.

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. In both cases we present accuracy in terms of a national average before then providing some state-level analysis. We present results only for the original EC and CO_2 growth series, with the per-capita growth being reported in the Appendix.[15] The findings are very similar between the main results and the per-capita results.

[15] In arriving at the per-capita figures for the quarterly PI 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

Figure 1 displays the average RMSFE across states at different nowcast release points according to the release schedule in Table 1. 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, the RMSFE of the ARX model falls when incorporating new information, both for electricity sales and PI. In the case of quarterly PI data ("ARX.PI"), we see a steady fall in RMSFE until the end of the nowcast year (M12:Y1) which is when the first three quarters of data have already been released. This fall corresponds to an improvement of 10% over the benchmark model. However, the results when using monthly electricity sales ("ARX.ELEC") are even more promising. Already by the middle of the nowcast year (M6:Y1) the model provides around a 15% improvement over the benchmark model. Once all of the nowcast year's data have been released at the beginning of the backcast period (M1:Y2), the gain is around 20%. This indicates that timely electricity sales information are, indeed, useful in nowcasting state-level EC. We also note that the RMSFE profiles look to be generally declining as we add information, which adds to the evidence for nowcast monotonicity observed in aggregate national-level studies (Giannone et al., 2008; Fosten and Gutknecht, 2020).

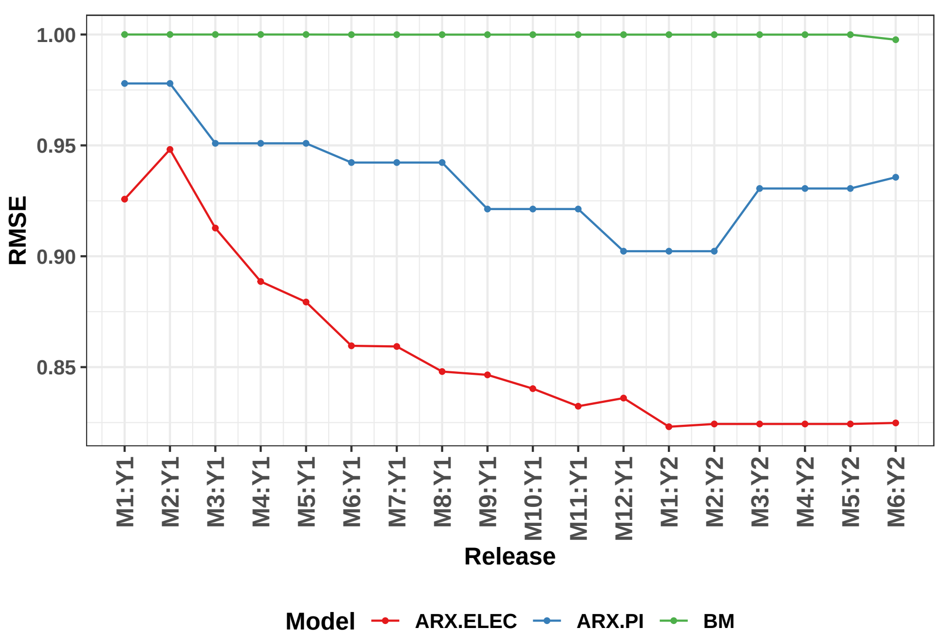


Figure 1: Average RMSFE Across States - Energy Consumption

Notes: ARX.ELEC/ARX.PI: ARX model with electricity sales/PI data respectively. BM: historic mean benchmark for EC. The RMSFE figures are normalised by the benchmark 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 already show a reasonable improvement over the benchmark after economic data have been released, when we dig into the state-level information we see much more substantial improvements of our method. To summarise the results across states, Table 2 presents the quantiles of the state-specific RMSFEs for the ARX model relative to the benchmark model, for the two predictors considered. In general, the table confirms what is seen in Figure 1 and we see that using economic data consistently improves over the benchmark in virtually all cases, and that the relative RMSFE is falling as new information is added. Moreover, focussing on the electricity sales results in Table 2b, by the end of the nowcast year we see that the ARX model has gains of almost 25% relative to the benchmark at the lower quartile, and gains of almost 40% at the 10th percentile. This shows that there are states in which the gains from our method are especially large.

To dig even further into the state-level results, Figure 3 in the Appendix presents the RMSFE results for every state (as in Figure 1), with the states ordered from largest to smallest in terms of their 2018 CO_2 emissions. From these plots we see that our method indeed provides big gains in some of the very large emitting states such as Florida, California and Pennsylvania which all see improvements of 20-30% relative to the benchmark. The gain is less notable in others, such as Texas. These results should, of course, be treated with an element of caution as the state-level RMSFEs are calculated on a small number of observations whereas the national average results have the benefit of pooling information across states. This means that in some cases large differences in relative RMSFEs across models can be driven by only a handful of observations.

As one further check regarding the state-level results, we also re-computed the equally-weighted average RMSFE results from Figure 1 to instead use a weighted average of the RMSFEs:

$$WRMSFE_v = \frac{1}{N}\sum_{i=1}^{N}\omega_i\sqrt{\frac{1}{P}\sum_{t=T-P+1}^{T}\widehat{\varepsilon}_{v,i,t}^2} \tag{8}$$

with the weights ω_i calculated according to the 2018 state CO_2 emissions levels. The results, displayed in Figure 5 in the Appendix, are incredibly similar to those of the unweighted average in Figure 1 which shows that we do not miss any “large state” effect when we use the unweighted average.

In summary of the EC results, we find that releases of current economic data, especially electricity sales, yield improvements in predicting growth rates of EC. The average improvement is of the order 15-20% relative to the benchmark model and can be even more sizeable when we dig into the individual state-level results. We already find good accuracy gains in the middle of the nowcast year and especially as we move towards the backcast period as all of the monthly data have been released. This occurs well over a year in advance of the release of the EC data, and so we are able to make huge timeliness gains using our nowcasting framework.

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). Figure 2 displays the results of the bridge equation method based on EC nowcasts from electricity sales or PI, relative to the historic mean benchmark. In predicting CO_2 emissions we notice that the more traditional economic indicator PI is barely able to beat the benchmark model in any of the nowcast periods. On the other hand, the use of electricity sales is capable of improving over the benchmark

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.9062	0.9414	0.9779	1.0107	1.0494
2	M2:Y1	0.9062	0.9414	0.9779	1.0107	1.0494
3	M3:Y1	0.8461	0.9140	0.9508	0.9946	1.0552
4	M4:Y1	0.8461	0.9140	0.9508	0.9946	1.0552
5	M5:Y1	0.8461	0.9140	0.9508	0.9946	1.0552
6	M6:Y1	0.8204	0.8893	0.9561	0.9909	1.0408
7	M7:Y1	0.8204	0.8893	0.9561	0.9909	1.0408
8	M8:Y1	0.8204	0.8893	0.9561	0.9909	1.0408
9	M9:Y1	0.7707	0.8767	0.9246	0.9834	1.0307
10	M10:Y1	0.7707	0.8767	0.9246	0.9834	1.0307
11	M11:Y1	0.7707	0.8767	0.9246	0.9834	1.0307
12	M12:Y1	0.7502	0.8631	0.9058	0.9575	1.0384
13	M1:Y2	0.7502	0.8631	0.9058	0.9575	1.0384
14	M2:Y2	0.7502	0.8631	0.9058	0.9575	1.0384
15	M3:Y2	0.7841	0.8837	0.9479	0.9859	1.0308
16	M4:Y2	0.7841	0.8837	0.9479	0.9859	1.0308
17	M5:Y2	0.7841	0.8837	0.9479	0.9859	1.0308
18	M6:Y2	0.7956	0.8820	0.9568	0.9918	1.0247

(a) Predictor - PI

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.7793	0.8690	0.9614	0.9990	1.0285
2	M2:Y1	0.8093	0.9100	0.9776	1.0236	1.0433
3	M3:Y1	0.7920	0.8672	0.9295	0.9892	1.0117
4	M4:Y1	0.7502	0.8398	0.9139	0.9671	0.9988
5	M5:Y1	0.7563	0.8258	0.9030	0.9638	0.9930
6	M6:Y1	0.7509	0.7941	0.8468	0.9350	0.9933
7	M7:Y1	0.7646	0.8003	0.8613	0.9265	0.9738
8	M8:Y1	0.6952	0.7849	0.8482	0.9216	0.9954
9	M9:Y1	0.6756	0.7651	0.8462	0.9477	1.0160
10	M10:Y1	0.6469	0.7663	0.8573	0.9299	1.0235
11	M11:Y1	0.6537	0.7494	0.8364	0.9426	1.0185
12	M12:Y1	0.6460	0.7620	0.8462	0.9504	0.9961
13	M1:Y2	0.6357	0.7393	0.8290	0.9235	1.0206
14	M2:Y2	0.6449	0.7489	0.8166	0.9342	1.0275
15	M3:Y2	0.6449	0.7489	0.8166	0.9342	1.0275
16	M4:Y2	0.6449	0.7489	0.8166	0.9342	1.0275
17	M5:Y2	0.6449	0.7489	0.8166	0.9342	1.0275
18	M6:Y2	0.6605	0.7327	0.8355	0.9426	0.9906

(b) Predictor - Electricity Sales

Table 2: Distribution of Relative RMSFE Across States - Energy Consumption

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 states below the relevant quantile.

by 10-15% once we reach the end of the nowcast period, which is over two years before the publication of the CO_2 data. We note that the addition of factors in the bridge equation model (displayed with dashed lines) can yield some minor improvements in the EC.ELEC model in the nowcast period but these are somewhat marginal.

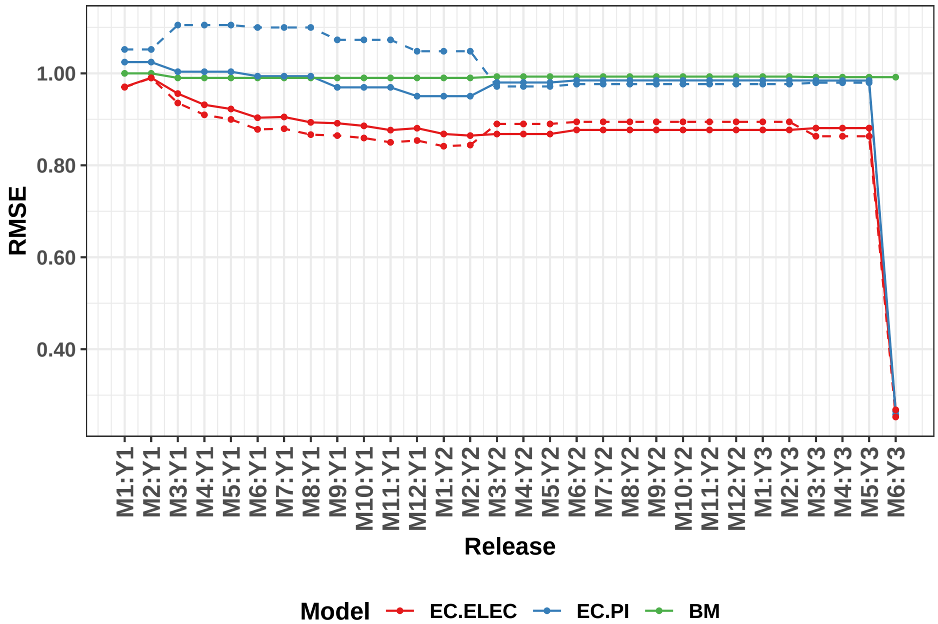


Figure 2: Average RMSFE Across States - CO_2 Emissions

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

In addition to these gains from using the timely economic data, another 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.

We also take a look into the state-level findings as we did before in the case of EC. Table 3 presents the relative RMSFE distributions across states, for the EC.ELEC and EC.PI models considered in Figure 2 without factors. Here we see that the gains from our method relative to the benchmark are as high as 25% at the 10th percentile in the the electricity sales bridge equation model (Table 3b) and around 20% for the lower quartile. We also see that, although PI does not provide much average gain in RMSFE, it can still give over 10% gains in selected states. The sudden drop in average MSFE on the release of the current year's EC data is also mirrored in these quantile results at nowcast point 30. The charts for individual states, displayed in Figure 4 in the Appendix, again show large gains in big CO_2 emitting states like Florida and Pennsylvania.

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.9635	0.9904	1.0206	1.0441	1.1002
2	M2:Y1	0.9635	0.9904	1.0206	1.0441	1.1002
3	M3:Y1	0.9459	0.9746	1.0134	1.0478	1.0990
4	M4:Y1	0.9459	0.9746	1.0134	1.0478	1.0990
5	M5:Y1	0.9459	0.9746	1.0134	1.0478	1.0990
6	M6:Y1	0.9234	0.9578	1.0077	1.0395	1.0761
7	M7:Y1	0.9234	0.9578	1.0077	1.0395	1.0761
8	M8:Y1	0.9234	0.9578	1.0077	1.0395	1.0761
9	M9:Y1	0.8990	0.9406	0.9835	1.0186	1.0542
10	M10:Y1	0.8990	0.9406	0.9835	1.0186	1.0542
11	M11:Y1	0.8990	0.9406	0.9835	1.0186	1.0542
12	M12:Y1	0.8807	0.9298	0.9622	1.0021	1.0249
13	M1:Y2	0.8807	0.9298	0.9622	1.0021	1.0249
14	M2:Y2	0.8807	0.9298	0.9622	1.0021	1.0249
15	M3:Y2	0.8871	0.9753	1.0021	1.0276	1.0435
16	M4:Y2	0.8871	0.9753	1.0021	1.0276	1.0435
17	M5:Y2	0.8871	0.9753	1.0021	1.0276	1.0435
18	M6:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
19	M7:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
20	M8:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
21	M9:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
22	M10:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
23	M11:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
24	M12:Y2	0.8804	0.9775	1.0062	1.0318	1.0397
25	M1:Y3	0.8804	0.9775	1.0062	1.0318	1.0397
26	M2:Y3	0.8804	0.9775	1.0062	1.0318	1.0397
27	M3:Y3	0.9005	0.9733	1.0049	1.0289	1.0521
28	M4:Y3	0.9005	0.9733	1.0049	1.0289	1.0521
29	M5:Y3	0.9005	0.9733	1.0049	1.0289	1.0521
30	M6:Y3	0.1741	0.2109	0.2755	0.3230	0.3955

(a) Predictor - PI

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.8993	0.9476	0.9756	1.0120	1.0390
2	M2:Y1	0.9283	0.9579	0.9999	1.0225	1.0578
3	M3:Y1	0.8966	0.9295	0.9659	1.0005	1.0269
4	M4:Y1	0.8705	0.9091	0.9429	0.9952	1.0228
5	M5:Y1	0.8529	0.9067	0.9335	0.9776	1.0056
6	M6:Y1	0.8364	0.8616	0.9048	0.9417	0.9814
7	M7:Y1	0.8278	0.8784	0.9052	0.9504	1.0345
8	M8:Y1	0.8038	0.8401	0.9109	0.9541	0.9925
9	M9:Y1	0.7960	0.8312	0.9135	0.9504	0.9972
10	M10:Y1	0.7588	0.8255	0.9033	0.9546	1.0178
11	M11:Y1	0.7378	0.8047	0.8975	0.9609	1.0083
12	M12:Y1	0.7538	0.8156	0.8991	0.9676	1.0035
13	M1:Y2	0.7501	0.8020	0.8780	0.9536	1.0191
14	M2:Y2	0.7403	0.8064	0.8754	0.9546	1.0237
15	M3:Y2	0.7439	0.8002	0.8816	0.9549	1.0111
16	M4:Y2	0.7439	0.8002	0.8816	0.9549	1.0111
17	M5:Y2	0.7439	0.8002	0.8816	0.9549	1.0111
18	M6:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
19	M7:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
20	M8:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
21	M9:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
22	M10:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
23	M11:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
24	M12:Y2	0.7518	0.8087	0.8770	0.9598	1.0521
25	M1:Y3	0.7518	0.8087	0.8770	0.9598	1.0521
26	M2:Y3	0.7518	0.8087	0.8770	0.9598	1.0521
27	M3:Y3	0.7583	0.8086	0.8772	0.9665	1.0717
28	M4:Y3	0.7583	0.8086	0.8772	0.9665	1.0717
29	M5:Y3	0.7583	0.8086	0.8772	0.9665	1.0717
30	M6:Y3	0.1836	0.2141	0.2769	0.3192	0.3963

(b) Predictor - Electricity Sales

Emissions

Table 3: Distribution of Relative RMSFE Across States - CO_2

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 states below the relevant quantile. Results are presented for different methods of computing the EC forecasts (PI and electricity sales) without factors.

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 EC and CO_2 data, motivated by the use of per capita figures in Bennedsen et al. (2021). The results in Appendix C 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 D, the evaluation sample 2001-2018 is used instead of that of 2009-2018 in Figure 2. The results are very similar, showing that the findings are indeed stable over time. We also explored the idea of nowcasting CO_2 directly instead of through the EC bridging variable. Figure 10 shows that the results are worse when using PI data and quite similar when using electricity sales data. However, 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. In previous versions of the paper we also tried various other versions of these models, including using combinations of multiple variables and switching to different variables like the Federal Reserve Bank of Philadelphia's state coincident index. However, none of these additional checks were able to outperform the best model using electricity sales data.

6 Conclusion

This paper has proposed methods for obtaining timely predictions of U.S. state-level EC 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 EC 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 over a naïve benchmark in predicting EC growth. Especially using electricity sales data, our methods deliver gains on average across states as well as large gains in certain individual states. These predictive gains can occur almost two years before the EC data are released. We also find that these predictions of EC are useful in bridge equation predictions of CO_2 emissions growth. Here we also see improvements in the CO_2 nowcasts as we add monthly electricity sales data. We also find that a very accurate prediction can be made by waiting until the release of the current year's EC data. This occurs 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. Another interesting avenue to explore is whether state-level nowcasts can be aggregated up to form accurate nowcasts of the national series. Conversely, it would also be interesting to explore in more detail whether national aggregates can further improve the state-level nowcasts.

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.

Baumeister, C., D. Leiva-León, and E. Sims (2022). Tracking weekly state-level economic conditions. *Review of Economics and Statistics*, 1–45.

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 41*(7), 675–696.

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 (2023). Nowcasting from Cross-Sectionally Dependent Panels. *Journal of Applied Econometrics (forthcoming).*

Furukawa, K., R. Hisano, Y. Minoura, and T. Yagi (2022). A Nowcasting Model of Industrial Production using Alternative Data and Machine Learning Approaches. Technical report, Bank of Japan.

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.

Higgins, P. C. (2014). GDPNow: A Model for GDP "Nowcasting". *FRB Atlanta Working Paper 2014-7*.

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.

Lehmann, R. and S. Möhrle (2022). Forecasting Regional Industrial Production with High-Frequency Electricity Consumption Data. *CESifo Working Paper No. 9917, Available at SSRN: https://ssrn.com/abstract=4215036*.

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.

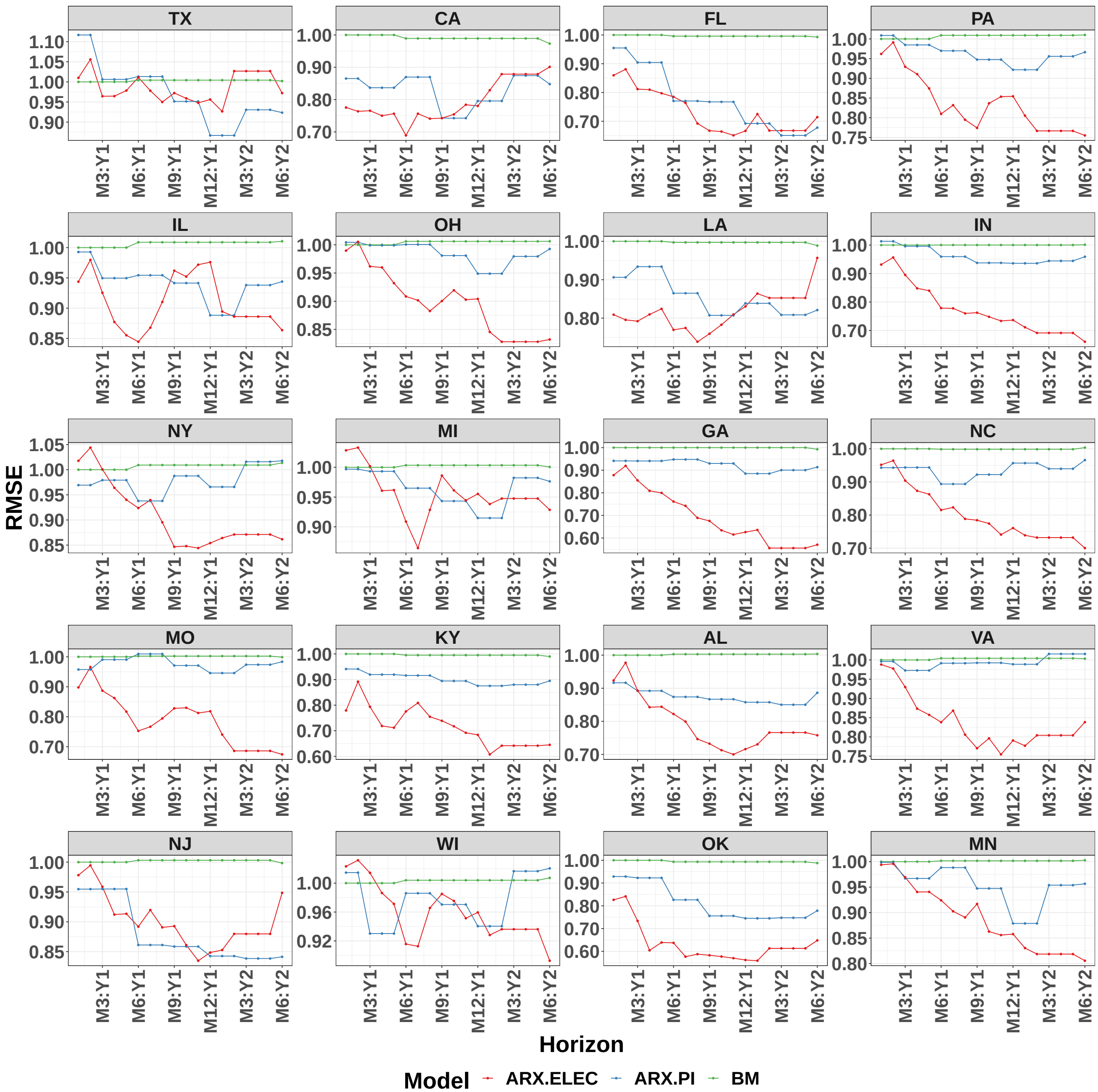
Sinclair, T. M. and H. O. Stekler (2013). Examining the quality of early GDP component estimates. *International Journal of Forecasting 29*(4), 736–750.

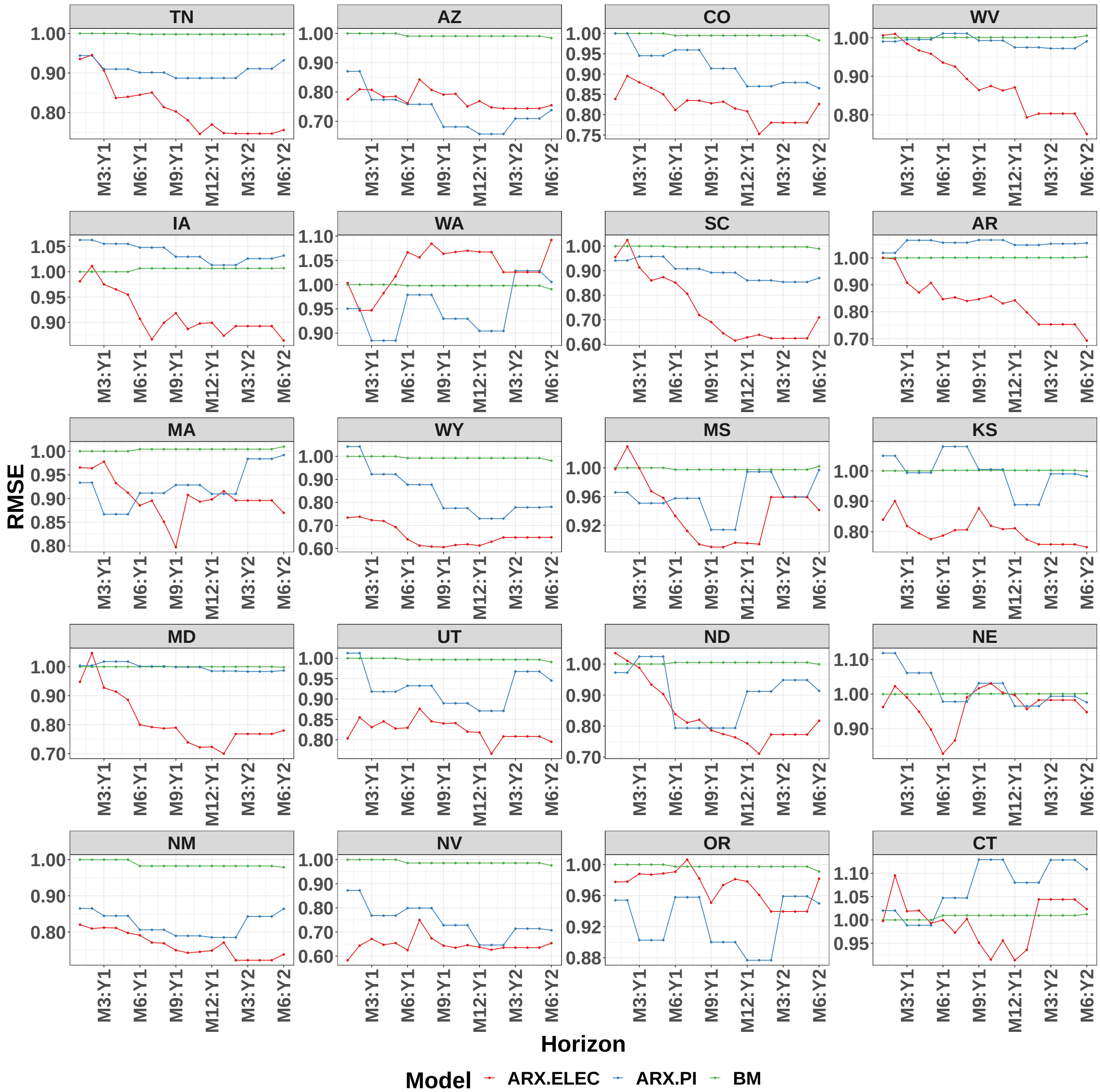
Wang, W., X. Zhang, and R. Paap (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics 34*(5), 724–745.

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

Appendix A State-Level RMSFE Results

A.1 Energy Consumption




TN
AZ
CO
WV
IA
WA
SC
AR
MA
WY
MS
KS
MD
UT
ND
NE
NM
NV
OR
CT
RMSE
Horizon
M3:Y1
M6:Y1
M9:Y1
M12:Y1
M3:Y2
M6:Y2
Model
ARX.ELEC
ARX.PI
BM

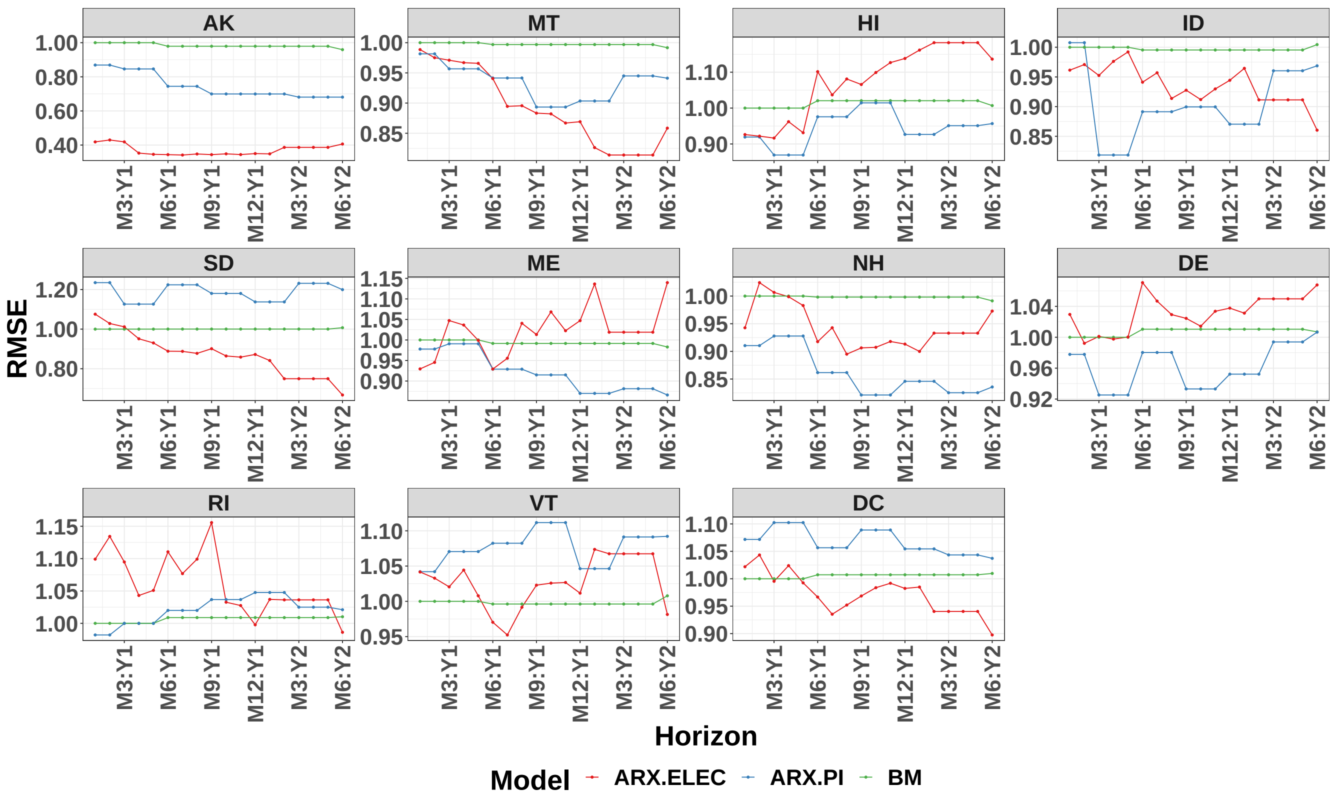
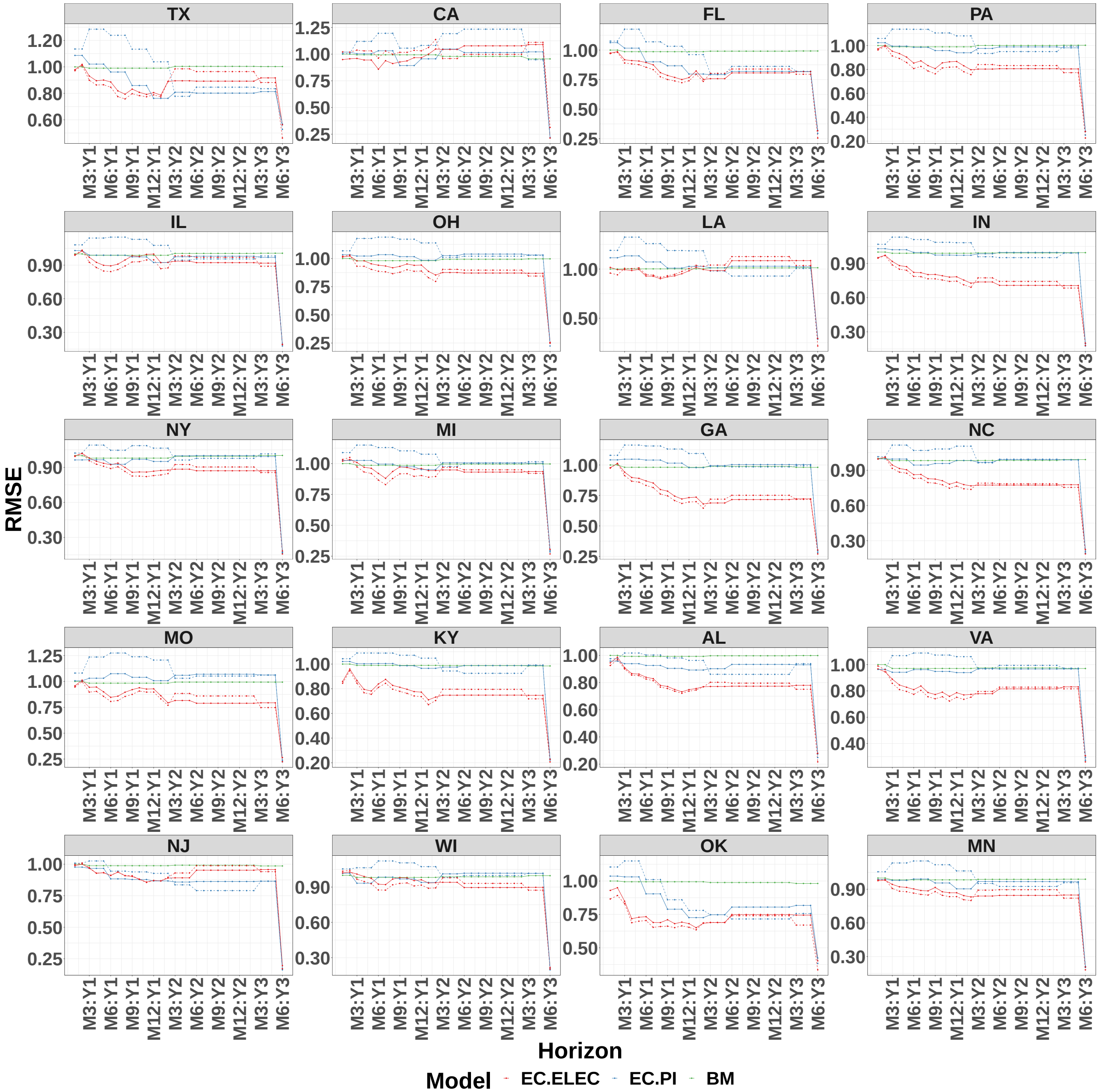


Figure 3: State-Level RMSFE - Energy Consumption

A.2 CO_2 Emissions




TN
AZ
CO
WV
IA
WA
SC
AR
MA
WY
MS
KS
MD
UT
ND
NE
NM
NV
OR
CT
RMSE
Horizon
M3:Y1
M6:Y1
M9:Y1
M12:Y1
M3:Y2
M6:Y2
M9:Y2
M12:Y2
M3:Y3
M6:Y3
Model
EC.ELEC
EC.PI
BM

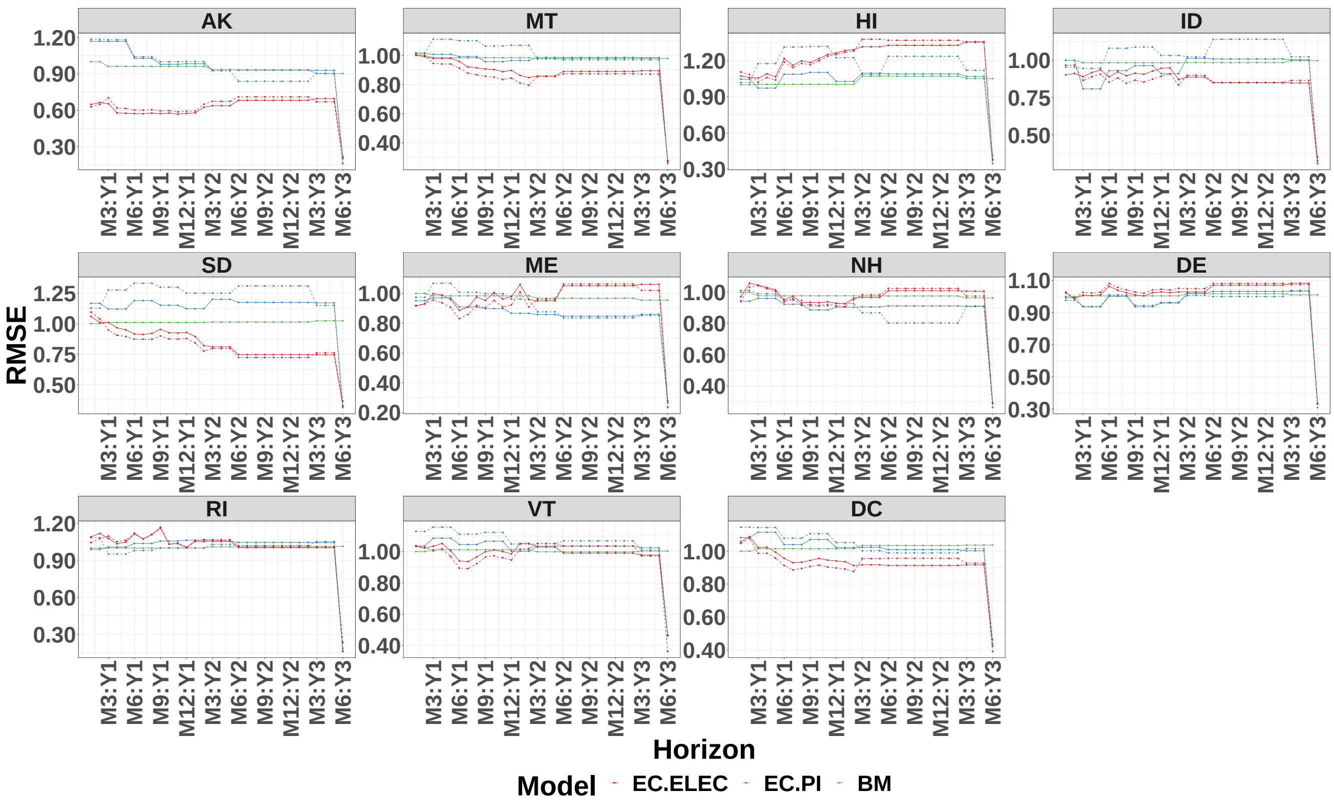


Figure 4: State-Level RMSFE - CO_2 Emissions

Appendix B   Weighted Average RMSFE Results

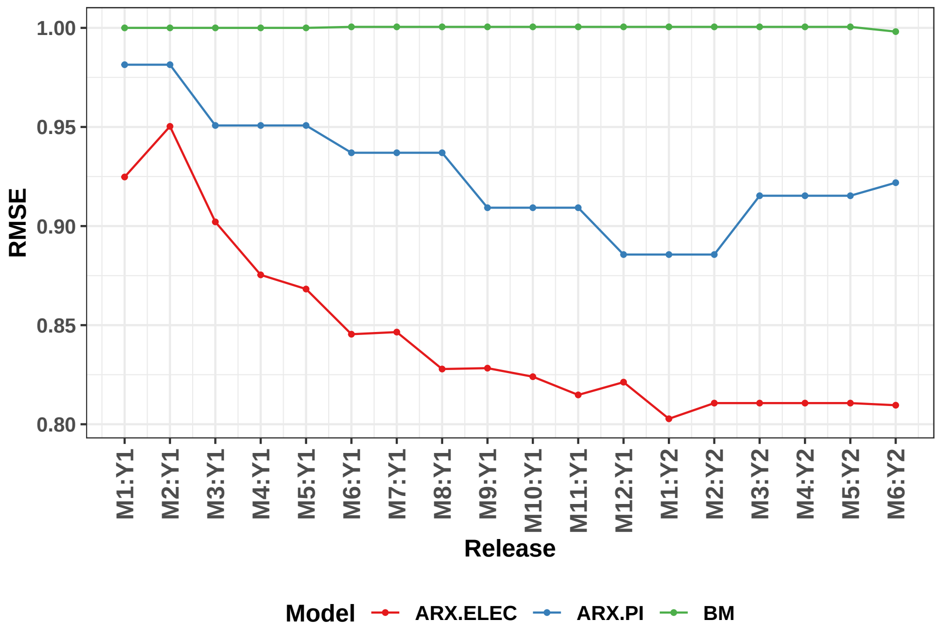


Figure 5: Weighted Average RMSFE Across States - Energy Consumption

Notes: Same as for Figure 1 whereas here the average RMSFE figure is calculated with weights corresponding to the level of state CO_2 emissions in 2018.

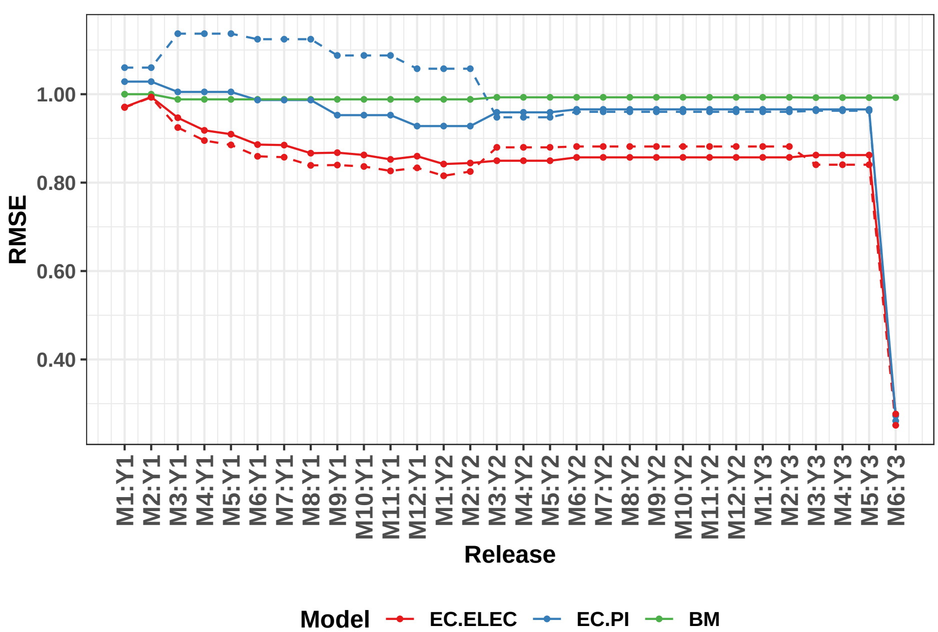


Figure 6: Weighted Average RMSFE Across States - CO_2 Emissions

Notes: Same as for Figure 2 whereas here the average RMSFE figure is calculated with weights corresponding to the level of state CO_2 emissions in 2018.

Appendix C Per Capita Results

C.1 Overall Results - Energy Consumption

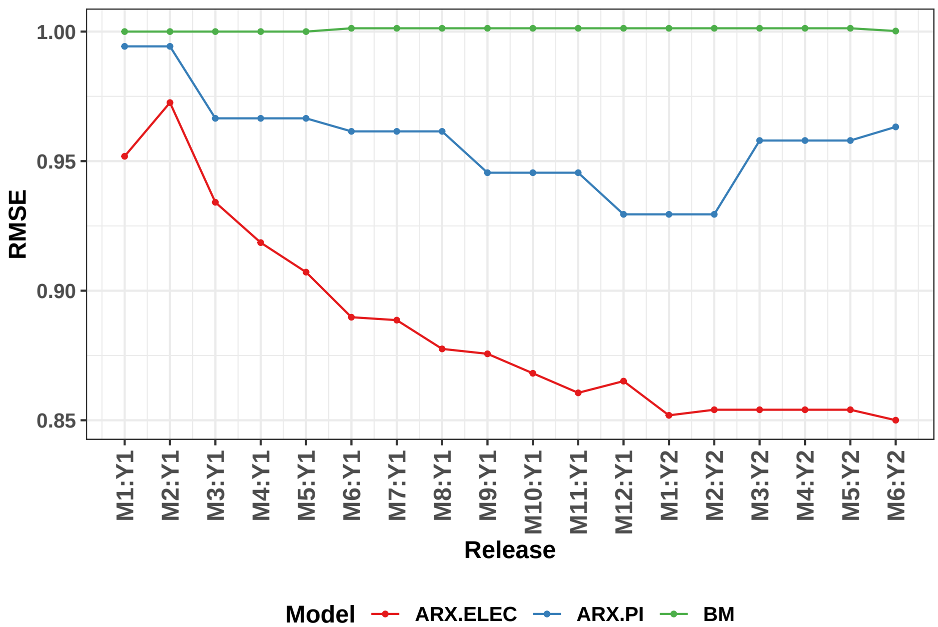


Figure 7: Average RMSFE Across States - Per Capita Energy Consumption

C.2 Overall Results - CO_2 Emissions

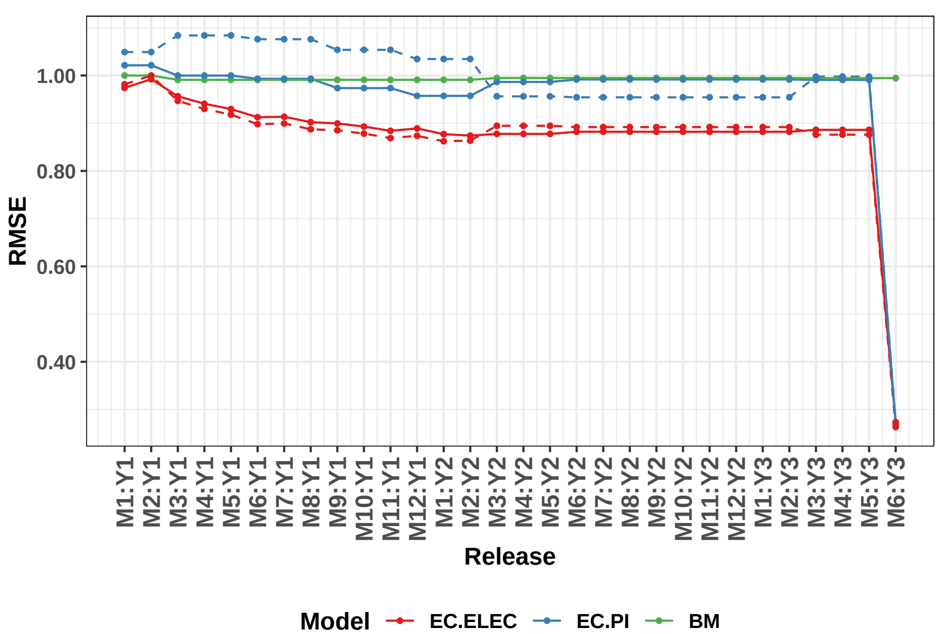


Figure 8: Average RMSFE Across States - Per Capita CO_2 Emissions

C.3 State-Level Results - Energy Consumption

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.9451	0.9611	0.9929	1.0171	1.0549
2	M2:Y1	0.9451	0.9611	0.9929	1.0171	1.0549
3	M3:Y1	0.9155	0.9353	0.9607	0.9958	1.0226
4	M4:Y1	0.9155	0.9353	0.9607	0.9958	1.0226
5	M5:Y1	0.9155	0.9353	0.9607	0.9958	1.0226
6	M6:Y1	0.8662	0.9111	0.9644	1.0064	1.0367
7	M7:Y1	0.8662	0.9111	0.9644	1.0064	1.0367
8	M8:Y1	0.8662	0.9111	0.9644	1.0064	1.0367
9	M9:Y1	0.8026	0.9022	0.9387	0.9898	1.0645
10	M10:Y1	0.8026	0.9022	0.9387	0.9898	1.0645
11	M11:Y1	0.8026	0.9022	0.9387	0.9898	1.0645
12	M12:Y1	0.8303	0.8911	0.9311	0.9668	1.0310
13	M1:Y2	0.8303	0.8911	0.9311	0.9668	1.0310
14	M2:Y2	0.8303	0.8911	0.9311	0.9668	1.0310
15	M3:Y2	0.8438	0.9154	0.9708	1.0005	1.0545
16	M4:Y2	0.8438	0.9154	0.9708	1.0005	1.0545
17	M5:Y2	0.8438	0.9154	0.9708	1.0005	1.0545
18	M6:Y2	0.8578	0.9262	0.9741	1.0072	1.0374

(a) Predictor - PI

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.8149	0.9194	0.9673	1.0100	1.0337
2	M2:Y1	0.8682	0.9463	0.9850	1.0197	1.0600
3	M3:Y1	0.8197	0.8860	0.9528	0.9951	1.0229
4	M4:Y1	0.8054	0.8630	0.9281	0.9882	1.0268
5	M5:Y1	0.8084	0.8515	0.9111	0.9898	1.0182
6	M6:Y1	0.7842	0.8235	0.8730	0.9408	1.0320
7	M7:Y1	0.7752	0.8321	0.8921	0.9357	1.0346
8	M8:Y1	0.7512	0.7994	0.8829	0.9537	1.0284
9	M9:Y1	0.7166	0.7867	0.8735	0.9657	1.0598
10	M10:Y1	0.7104	0.7962	0.8716	0.9591	1.0426
11	M11:Y1	0.7010	0.7552	0.8596	0.9532	1.0434
12	M12:Y1	0.6827	0.7892	0.8658	0.9620	1.0373
13	M1:Y2	0.6691	0.7686	0.8517	0.9422	1.0475
14	M2:Y2	0.6890	0.7591	0.8566	0.9714	1.0518
15	M3:Y2	0.6890	0.7591	0.8566	0.9714	1.0518
16	M4:Y2	0.6890	0.7591	0.8566	0.9714	1.0518
17	M5:Y2	0.6890	0.7591	0.8566	0.9714	1.0518
18	M6:Y2	0.6697	0.7455	0.8427	0.9615	1.0670

(b) Predictor - Electricity Sales

Table 4: Distribution of Relative RMSFE Across States - Per Capita Energy Consumption

C.4 State-Level Results - CO_2 Emissions

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.9676	0.9941	1.0184	1.0382	1.0808
2	M2:Y1	0.9676	0.9941	1.0184	1.0382	1.0808
3	M3:Y1	0.9468	0.9747	1.0104	1.0392	1.0753
4	M4:Y1	0.9468	0.9747	1.0104	1.0392	1.0753
5	M5:Y1	0.9468	0.9747	1.0104	1.0392	1.0753
6	M6:Y1	0.9261	0.9556	1.0059	1.0394	1.0684
7	M7:Y1	0.9261	0.9556	1.0059	1.0394	1.0684
8	M8:Y1	0.9261	0.9556	1.0059	1.0394	1.0684
9	M9:Y1	0.9076	0.9416	0.9834	1.0302	1.0619
10	M10:Y1	0.9076	0.9416	0.9834	1.0302	1.0619
11	M11:Y1	0.9076	0.9416	0.9834	1.0302	1.0619
12	M12:Y1	0.8897	0.9341	0.9702	1.0039	1.0366
13	M1:Y2	0.8897	0.9341	0.9702	1.0039	1.0366
14	M2:Y2	0.8897	0.9341	0.9702	1.0039	1.0366
15	M3:Y2	0.9172	0.9717	0.9989	1.0329	1.0734
16	M4:Y2	0.9172	0.9717	0.9989	1.0329	1.0734
17	M5:Y2	0.9172	0.9717	0.9989	1.0329	1.0734
18	M6:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
19	M7:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
20	M8:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
21	M9:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
22	M10:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
23	M11:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
24	M12:Y2	0.9100	0.9708	1.0019	1.0312	1.0674
25	M1:Y3	0.9100	0.9708	1.0019	1.0312	1.0674
26	M2:Y3	0.9100	0.9708	1.0019	1.0312	1.0674
27	M3:Y3	0.9321	0.9752	0.9974	1.0330	1.0546
28	M4:Y3	0.9321	0.9752	0.9974	1.0330	1.0546
29	M5:Y3	0.9321	0.9752	0.9974	1.0330	1.0546
30	M6:Y3	0.1762	0.2107	0.2713	0.3258	0.4044

(a) Predictor - PI

Release	Period	10%	25%	50%	75%	90%
1	M1:Y1	0.8836	0.9469	0.9813	1.0127	1.0388
2	M2:Y1	0.9283	0.9603	0.9972	1.0260	1.0530
3	M3:Y1	0.8833	0.9279	0.9659	1.0044	1.0350
4	M4:Y1	0.8754	0.9153	0.9500	0.9968	1.0500
5	M5:Y1	0.8547	0.9027	0.9347	0.9879	1.0272
6	M6:Y1	0.8436	0.8640	0.9094	0.9451	1.0487
7	M7:Y1	0.8348	0.8825	0.9010	0.9583	1.0381
8	M8:Y1	0.8081	0.8474	0.9004	0.9656	1.0306
9	M9:Y1	0.8004	0.8347	0.9015	0.9901	1.0177
10	M10:Y1	0.7784	0.8308	0.8979	0.9864	1.0327
11	M11:Y1	0.7533	0.8046	0.8898	0.9800	1.0237
12	M12:Y1	0.7677	0.8250	0.8904	0.9804	1.0265
13	M1:Y2	0.7589	0.8075	0.8810	0.9611	1.0689
14	M2:Y2	0.7487	0.8093	0.8711	0.9695	1.0665
15	M3:Y2	0.7651	0.8077	0.8773	0.9600	1.0552
16	M4:Y2	0.7651	0.8077	0.8773	0.9600	1.0552
17	M5:Y2	0.7651	0.8077	0.8773	0.9600	1.0552
18	M6:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
19	M7:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
20	M8:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
21	M9:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
22	M10:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
23	M11:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
24	M12:Y2	0.7584	0.8070	0.8702	0.9562	1.0764
25	M1:Y3	0.7584	0.8070	0.8702	0.9562	1.0764
26	M2:Y3	0.7584	0.8070	0.8702	0.9562	1.0764
27	M3:Y3	0.7582	0.8073	0.8642	0.9642	1.0828
28	M4:Y3	0.7582	0.8073	0.8642	0.9642	1.0828
29	M5:Y3	0.7582	0.8073	0.8642	0.9642	1.0828
30	M6:Y3	0.1871	0.2172	0.2737	0.3334	0.3962

(b) Predictor - Electricity Sales

Table 5: Distribution of Relative RMSFE Across States - Per Capita CO_2 Emissions

Appendix D Further Results

D.1 Robustness to Sample Split

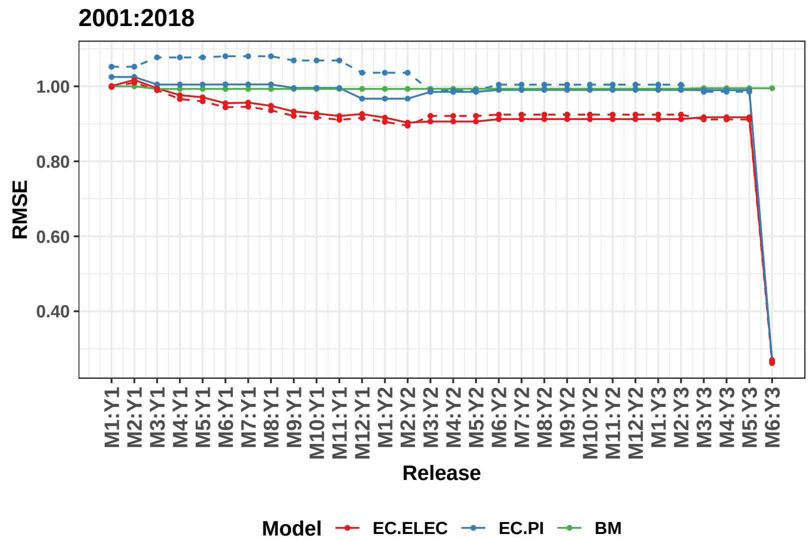


Figure 9: Sample Split - Average RMSFE Across States - CO_2 Emissions

D.2 Targeting CO_2 Emissions Directly Instead of Bridging

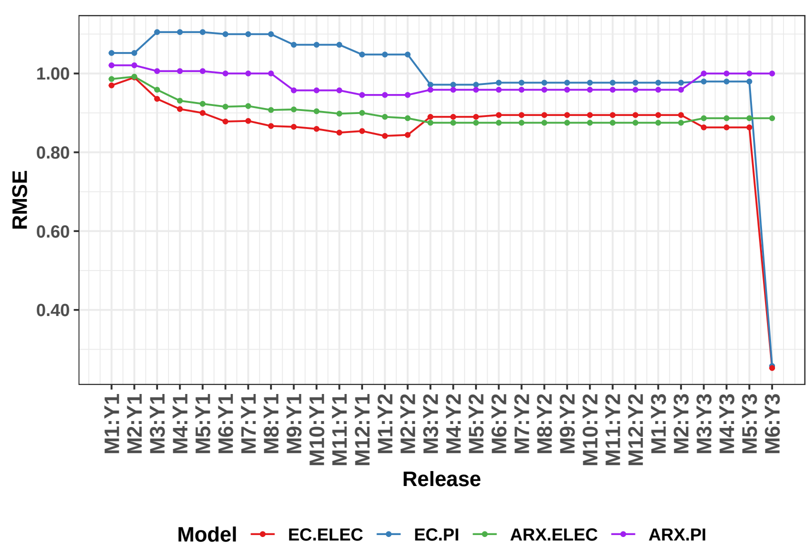


Figure 10: Targeting CO_2 Emissions Directly - Average RMSFE Across States

Notes: EC.ELEC/EC.PI denote the bridge equation model for CO_2 using predictions of EC produced by electricity sales/PI. ARX.ELEC/ARX.PI denote the direct mixed frequency model for CO_2 using electricity sales/PI.