ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

FRED-SD: A Real-Time Database for State-Level Data with Forecasting Applications

Authors	Kathryn O. Bokun, Laura E. Jackson, Kevin L. Kliesen, and Michael T. Owyang
Working Paper Number	2020-031E
Revision Date	April 2024
Citable Link	https://doi.org/10.20955/wp.2020.031
Suggested Citation	Bokun, K.O., Jackson, L.E., Kliesen, K.L., Owyang, M.T., 2024; FRED-SD: A Real-Time Database for State-Level Data with Forecasting Applications, Federal Reserve Bank of St. Louis Working Paper 2020-031. URL https://doi.org/10.20955/wp.2020.031

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

FRED-SD: A Real-Time Database for State-Level Data with Forecasting Applications*

Kathryn O. Bokun†

Laura E. Jackson‡§

Kevin L. Kliesen†

Michael T. Owyang†

keywords: factor model, Bayesian VAR, space-time autoregression

September 1, 2021

Abstract

We construct a real-time dataset (FRED-SD) with vintage data for the U.S. states that can be used to forecast both state-level and national-level variables. Our dataset includes approximately 28 variables per state, including labor market, production, and housing variables. We conduct two sets of real-time forecasting exercises. The first forecasts state-level labor-market variables using five different models and different levels of industrially-disaggregated data. The second forecasts a national-level variable exploiting the cross-section of state data. The state-forecasting experiments suggest that large models with industrially-disaggregated data tend to have higher predictive ability for industrially-diversified states. For national-level data, we find that forecasting and aggregating state-level data can outperform a random walk but not an autoregression. We compare these real-time data experiments with forecasting experiments using final-vintage data and find very different results. Because these final-vintage results are obtained with revised data that would not have been available at the time the forecasts would have been made, we conclude that the use of real-time data is essential for drawing proper conclusions about state-level forecasting models.

[JEL Codes: C33; R11]

*The authors benefitted from conversations with Mike McCracken and from comments and questions from participants at the 2018 Midwest Econometrics Group meetings. Julie K. Bennett, Joe McGillicuddy, and Hannah G. Shell provided research assistance. The views expressed here do not reflect the official opinions of the Federal Reserve Bank of St. Louis, the Federal Reserve Board of Governors or the Federal Reserve System. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: none.

†*Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63106 USA.*

‡*Department of Economics, Bentley University, 175 Forest Street, Waltham, MA 02452 USA.*

§Corresponding author. Email: ljackson@bentley.edu

1 Introduction

In the past, forecasting models were often restricted to small datasets to minimize parameter proliferation. Recent advances for estimating models using dimension reduction or shrinkage estimators have provided opportunities to forecast using large datasets. For example, factor models achieve dimension reduction by focusing on the common fluctuations across a set of variables.[1] More recently, advances in computational techniques and computing power—combined with Bayesian shrinkage—have demonstrated that vector autoregressions (VARs) with a large number of variables can be viable alternatives to factor models.[2]

Simultaneous with the eruption of large forecasting models has come rising interest in the use of geographically-disaggregate data for forecasting regional-level outcomes. Some studies have focused on forecasting particular states.[3] Others use a cross-section of state-level data.[4] There is also a smaller literature on forecasting national-level variables using disaggregate data.[5] A common theme across most of these papers is that they use a single vintage of data. Thus, the forecasts are made using ex post revisions, rather than using the data that would actually be available to the forecaster

Because in-sample experiments lead to overfitting, the forecast literature has generally adopted out-of-sample experiments that estimate models at each forecast origin using the data that would have been available at that time. The downside of out-of-sample forecasting comparisons is that they require vintage data that are not always easily obtained for regionally-disaggregated experiments. Recently, McCracken and Ng (2016) produced a dataset that includes real-time vintages of a large set of national data that had been used in factor models

[1]For applications using U.S. data, see Stock and Watson (2002a,b, 2006), Bernanke and Boivin (2003), Boivin and Ng (2005), and Forni, Hallin, Lippi and Reichlin (2005). For forecasting applications using European data, see Marcellino, Stock, Watson (2003).

[2]For example, Bańbura, Giannone, and Reichlin (2010) find that a 20-variable VAR can sufficiently capture relevant shocks that affect the macroeconomic outlook.

[3]For example, LeSage and Magura (1988, Ohio); Miller (1998, Idaho); and Rapach and Strauss (2005; Missouri). Lehmann and Wohlrabe (2014) survey the literature of forecasting regional economic variables.

[4]LeSage and Pan (1995) focus on the co-movement of activity between states or regions; Rapach and Strauss (2012) use U.S. state-level employment growth rates for forecasting; and LeSage and Hendrikz (2019) use the state-level coincident indices developed by Crone and Clayton-Matthews (2005).

[5]Owyang, Piger, and Wall (2015) include state-level employment growth series to improve the accuracy of nowcasts and short-horizon forecasts of U.S. business cycle phases, relative to models with only national-level data. Hernández-Murillo and Owyang (2006) forecast national-level employment growth with regional-level employment growth.

(e.g., Stock and Watson (1996, 2002a,b), among others). Their dataset contains over 100 series starting in January 1959 with vintages back through August 1999. The contribution of McCracken and Ng is to evaluate the dataset and maintain it over the course of time, making it accessible for researchers now and in the future.

For this paper, we construct a similar real-time dataset (FRED-SD) with vintage data for the U.S. states that can be used to forecast both state-level and national-level variables. While the number of macroeconomic indicators for each state number fewer than for the nation, there are obviously 50 times as many states. Our dataset includes approximately 28 variables per state, the number of series varying across states depending on their industrial composition. We include labor market, production, and housing variables that can be used, for example, to construct real-time coincident indicators for the states (e.g., Crone and Clayton-Matthews, 2005) or for forecasting experiments.

To demonstrate the utility of the dataset, we forecast state-level variables using a variety of models and levels of industrial disaggregation. We also forecast a national-level variable exploiting the cross-section of state data. We perform two types of experiments: (i) real-time forecasting experiments using only the data that a forecaster would have at the forecast origin and (ii) psuedo-out-of-sample experiments that use the most recent vintage of data to estimate and forecast at each forecast origin. In the latter case, the data have been revised and would not reflect the information that forecasters had.

For the real-time state-level forecasting experiments, we find that large models with industrially-disaggregated data tend to have higher predictive ability when the state of interest is industrially diversified. Moreover, a model with multiple states has greater predictive ability when the state of interest conducts large amounts of cross-state trade. In national forecasting experiments, we find that forecasting and aggregating state-level data can outperform a random walk but not an autoregression. Moreover, we find that these results would differ substantially if the experiments were pseudo-out-of-sample rather than real-time out-of-sample. For the state-level forecasting experiments, the most accurate model depends on whether the forecasts were constructed using real-time vintages verus final vintage data.

The balance of the paper follows: Section 2 describes the raw data, data transformations, and the sources. We outline the construction of the FRED-SD database that is intended to be publicly available and updated regularly. Section 3 describes some experiments for state-level data using the FRED-SD database and Section 4 describes some experiments for national-level data using the FRED-SD database. Section 5 summarizes and concludes.

2 The FRED-SD Database

In 2015, the Federal Reserve Bank of St. Louis unveiled FRED-MD, a database for economic research using large panels of vintage U.S. macroeconomic data. Described in McCracken and Ng (2016), that database contains 128 monthly U.S. economic and financial indicators (e.g., employment, prices, production). McCracken and Ng touted three benefits: (i) it is real-time with new vintages added once a month; (ii) it is publicly available; and (iii) researchers do not incur the costs of building, maintaining, and updating such a dataset themselves. Instead, this dataset is provided as a public service to researchers, with the cost borne by the Federal Reserve Bank of St. Louis.

Here, we construct a state-level analogue to FRED-MD, composed of state-level data, called FRED-SD. FRED-SD includes 28 state-level time series, listed in Table 1, encompassing three types of data: (i) labor market; (ii) income, product, imports and exports; and (iii) housing. Table 2 includes additional information on the FRED mnemonics used to access the data series. The data are mixed frequency, both monthly and quarterly, and organized in monthly vintages. A vintage reflects the time series of data that would be available that month, including revisions to past data that occurred before that month. Moreover, the vintage reflects potential variation in the publication lags across the many series. For every vintage, each series contains all previous observations that are available—i.e., the time series contain all of the data that would be available during that month.[6] Thus, the starting dates

[6] According to Croushore and Stark (2000), the Philadelphia Fed real-time data is updated quarterly on the 15th of each month in the relevant calendar quarter (Feb., May, Aug., and Nov.). Recently, however, they have changed to update the data monthly. By contrast, the FRED SD data is updated on the last business day of the month. Because state-level employment data is typically released two to three weeks after the national data, the January 2021 vintage will include data through December 2020.

vary across series, producing unbalanced panels, and the availability of series varies across vintages.

For convenience, we organize the data in two ways that will facilitate forecasting economic conditions within state and across states. In one set of spreadsheets, each vintage is organized by state on the tabs and the series in the columns. In this case, we list the series with the monthly data first and the quarterly data last. This type of organization should facilitate forecasting state-level outcomes using a cross-section of within-state data. In another set of spreadsheets with the same data, each vintage is organized by series on the tabs and the states in the columns. This organization facilitates forecasting aggregate outcomes using a cross-section of state-level series.

Table 1: FRED-SD Series

Variable	Description	Frequency	First Observation	First Vintage
NA	All employees, Total nonfarm	Monthly	1/1/1990	2007-06 *
UR	Unemployment rate	Monthly	1/1/1976	2007-06 *
MFGHRS	Average weekly hours, private employees	Monthly	1/1/2007	2014-01
STHPI	FHFA house price index	Monthly	1/1/1975	2008-05
OTOT	Nominal personal income	Monthly	1/1/1948 *	2010-09 *
GOVT	All employees, Total government	Monthly	1/1/1990	2007-06 *
CONS	All employees, Construction	Monthly	1/1/1990	2007-06 *
MFG	All employees, Manufacturing	Monthly	1/1/1990	2007-06 *
FIRE	All employees, Financial activities	Monthly	1/1/1990	2007-06 *
INFO	All employees, Information	Monthly	1/1/1990	2007-06 *
RENTS	All employees, Real estate, rental, leasing	Monthly	1/1/1990	2014-01 *
PSERV	All employees, Private services	Monthly	1/1/1990	2014-01
MINNG	All employees, Mining and logging	Monthly	1/1/1990 *	2007-06 *
ICLAIMS	Unemployment insurance, initial claims	Monthly	2/1/1986 *	2010-09
BPPRIVSA	Permits, new privately owned housing units	Monthly	1/1/1988	2011-12 *
LF	Civilian labor force	Monthly	1/1/1976	2007-06 *
PARTRATE	Participation rate	Monthly	1/1/1976	2019-04
EXPORTS	Total exports by state	Quarterly	8/1/1995	2019-03
IMPORTS	Total imports by state	Quarterly	1/1/2008	2019-04
NQGSP	Gross State Product, All Industries	Quarterly	1/1/2005	2016-08
CONSTNQGSP	Gross State Product, Construction	Quarterly	1/1/2005 *	2016-08
UTILNQGSP	Gross State Product, Utilities	Quarterly	1/1/2005	2016-08
INFONQGSP	Gross State Product, Information	Quarterly	1/1/2005	2016-08
GOVNQGSP	Gross State Product, Government	Quarterly	1/1/2005	2016-08
MANNQGSP	Gross State Product, Manufacturing	Quarterly	1/1/2005	2016-08
NATURNQGSP	Gross State Product, Agriculture and Mining	Quarterly	1/1/2005 *	2016-08
FIRENQGSP	Gross State Product, Finance and Real Estate	Quarterly	1/1/2005	2016-08 *
PSERVNQGSP	Gross State Product, Other Private Industries	Quarterly	1/1/2005	2016-08

Table 1: Information on state-level data series included in FRED-SD. * Observation and vintage starting dates may vary for some states. See Data Appendix, Section 1, Part e.

Table 2: FRED mnemonic series

Variable	Description	Units	FRED Mnemonic	Source
NA	All employees, Total nonfarm	Thousands of persons	*XX*NA	BLS
UR	Unemployment rate	Percent	*XX*UR	BLS
MFGHRS	Average weekly hours, private employees	Number of hours	SMU*XX*000000500000002SA *	BLS
STHPI	FHFA house price index	Index, Q1-1980 = 100	*XX*STHPI	HFA
OTOT	Nominal personal income	Thousands of dollars	*XX*OTOT *	BEA
GOVT	All employees, Total government	Thousands of persons	*XX*GOVT	BLS
CONS	All employees, Construction	Thousands of persons	*XX*CONS *	BLS
MFG	All employees, Manufacturing	Thousands of persons	*XX*MFG *	BLS
FIRE	All employees, Financial activities	Thousands of persons	*XX*FIRE *	BLS
INFO	All employees, Information	Thousands of persons	*XX*INFO *	BLS
RENTS	All employees, Real estate, rental, leasing	Thousands of persons	SMS*XX*000005553000001	BLS
PSERV	All employees, Private services	Thousands of persons	SMS*XX*000000800000001	BLS
MINNG	All employees, Mining and logging	Thousands of persons	*XX*NRMN *	BLS
ICLAIMS	Unemployment insurance, initial claims	Number of claims	*XX*ICLAIMS	ETA
BPPRIVSA	Permits, new privately owned housing units	Number of units authorized	*XX*BPPRIVSA *	Census
LF	Civilian labor force	Thousands of persons	*XX*LF	BLS
PARTRATE	Participation rate	Percent	LBSSA*XX*	BLS
EXPORTS	Total exports by state	Millions of dollars	EXPTOT*XX*	Census
IMPORTS	Total imports by state	Millions of dollars	IMPTOT*XX*	Census
NQGSP	Gross State Product, All Industries	Millions of dollars	*XX*NQGSP	BEA
CONSTNQGSP	Gross State Product, Construction	Millions of dollars	*XX*CONSTNQGSP	BEA
UTILNQGSP	Gross State Product, Utilities	Millions of dollars	*XX*UTILNQGSP	BEA
INFONQGSP	Gross State Product, Information	Millions of dollars	*XX*INFONQGSP	BEA
GOVNQGSP	Gross State Product, Government	Millions of dollars	*XX*GOVNQGSP	BEA
MANNQGSP	Gross State Product, Manufacturing	Millions of dollars	*XX*MANNQGSP	BEA
NATURNQGSP	Gross State Product, Agriculture and Mining	Millions of dollars	Aggregated series	BEA
FIRENQGSP	Gross State Product, Finance and Real Estate	Millions of dollars	Aggregated series	BEA
PSERVNQGSP	Gross State Product, Other Private Industries	Millions of dollars	Aggregated series	BEA

Table 2: Information on FRED mnemonic used to access data included in FRED-SD. Note that "XX" represents placeholder for desired state abbreviation or numeric identifier, depending on the category. * FRED mnemonic may vary for some states. See Data Appendix, Section 1, Part d.

2.1 Data Descriptions

Similar to FRED-MD, we break each state's data into subcategories: labor market, income and production, and housing.

2.1.1 Labor Market Data

The majority of the data available at the state level consists of labor market data. For each state and the District of Columbia, we obtain monthly total nonfarm payroll employment, government employment, and seven private-industry-level employment series at the two-digit NAICS level from the Current Employment Statistics (CES) survey conducted by the Bureau of Labor Statistics. For each of these series, the unit of account is thousands of persons.

Because states are industrially diverse, their industry employment classifications can vary. Thus, some industry employment series may be missing or included in other categories, especially for less populous states. For example, for Delaware, the District of Columbia, and Hawaii, construction employment includes employment in the mining and logging industry and these separate series are omitted. Similarly, employment for the real estate industry does not exist for New Mexico, Rhode Island, and South Dakota.

The first observations of total nonfarm payroll begin in January 1959, except for Alaska, which begins a year later. Data vintages begin in April 2000. Private-sector employment and state-level industry employment start dates are much later, beginning in January 1990. From the CES, we also include average weekly hours for all private employees. Average weekly hours begins in January 2007, with the first data vintage beginning in January 2014.

We also include four other state-level labor series. Three series are from the Current Population Survey (CPS): the unemployment rate, the civilian labor force, and the labor force participation rate. The first observations for the three CPS series begin in January 1976. Data vintages begin in June 2007 for the unemployment rate and the labor force; however, earlier vintages (from June 2005) are available for seven states (AR, IL, IN, KY, MO MS, and TN). Participation rate vintages begin in April 2019. The fourth labor series is weekly initial claims for state unemployment insurance benefits. The claims data is collected as part of the Department of Labor's Federal-State Unemployment Insurance Program. Start dates for unemployment insurance claims data generally begin in February 1986, except for seven states (AL, CO, IL, IN, MN, PA, and VT), with the first vintage starting in July 2013.[7]

2.1.2 Income and Product Data

The second group of data are income, product (gross state product), and import/export data. The data are quarterly except for import and export data, which is monthly.

The gross state product (GSP) series is published by the Regional Product Division of the U.S. Bureau of Economic Analysis. Although its name implies that it is a state-level equivalent of U.S. GDP, which measures the sum of final expenditures (i.e., household consumption of goods and services, business investment, government, and net exports), it is instead measured using the income approach. The income approach measures the current-dollar value of the sum of labor income, business taxes, and capital income. Real state GSP data is available in millions of 2012 chain-weighted dollars. However, because the BEA shifts the base-year weights about every five years, we have decided to instead report nominal GDP.

We report total state GSP data (all industries) and eight industries: construction, utilities, information, government, manufacturing, agriculture and mining, finance and real estate, and other private industries. State GSP data is generally available since the first quarter of 2005, although earlier data is available for some industries (construction and agriculture) and some states. Although data vintages for total GSP and the eight industries mostly begin in August 2016, there are a few exceptions. For example, government product begins in August 2014.

Like state-level GSP, state personal income (PI) is published by the BEA. Personal income is measured in current dollars by the BEA. PI is a broad measure of income received by all persons from all sources. This includes income from an individual's source of employment; self-proprietor income (ownership of a home or business); income from the ownership of financial assets (dividends and interest); and income transferred from the government and from businesses. State PI does not include realized or unrealized capital gains and losses. State

[7] While all of the labor series were seasonally adjusted, there may be some remaining seasonality. Researchers may consider deseasonalizing the labor data again at their discretion.

PI is available from the first quarter of 1948 to the present for all states. Data vintages for all states are available from September 2010, except for AR, IL, IN, KY, MO, MS, and TN, which begin in June 2005.

Total state-level imports and exports are published by the U.S. Bureau of the Census. The data measure exports and imports of manufactured (durable and nondurable) goods. Services exports and imports are not measured at the state level. State-level exports and imports are measured in current dollars and are reported on a not-seasonally-adjusted basis. Export data begin in August 1995 and import data begin in January 2008. Data vintages for exports begin in March 2019 and import vintages begin a month later, in April 2019.

2.1.3 Housing Data

The final group of state-level series measures housing markets. We include both house prices and permits for new, privately-owned buildings. The house price series is a quarterly series published by the Federal Housing Finance Agency (FHFA), which is an independent federal agency that oversees the Federal National Mortgage Association (Fannie Mae), the Federal Home Loan Mortgage Corporation (Freddie Mac), and the Federal Home Loan banks. The FHFA house price index (HPI) is based on sales-purchase transactions involving conforming, conventional mortgages that are purchased or securitized by Fannie Mae or Freddie Mac. The HPI only includes mortgage transactions on single-family properties. The HPI is available beginning in the first quarter of 1975. HPI data vintages are available from May 2008.

Housing permits measure the number of potential new construction starts authorized by local government entities. Not all permits result in new construction. Housing permits are published by the U.S. Bureau of the Census, and count the total number of building permits for all types of building structures. Structure types include 1-unit, 2-unit, 3-unit, 4-unit, and 5-unit or more. Permit data is seasonally adjusted and begins in January 1988. Data vintages begin in December 2011, except for the District of Columbia, which begins in November 2016.

2.2 Data Issues

Working with state-level data can be more challenging than working with macro-level data for several reasons. First, because there are 50 states and the District of Columbia, each new state-level variable adds many series to the dataset. For example, 10 labor market series becomes 510 data series. Assembling vintages of these data for real-time analysis also adds complexity.

Second, the state data are typically more volatile than national data. Figure 1 plots the annualized monthly growth in nonfarm payroll employment from January 1990 through December 2019 for the United States and for a few interesting cases: Ohio, Texas, Alaska, and West Virginia. Over this period, the standard deviation of average annualized monthly employment growth for the United States was 1.9 percent. States such as Ohio and Texas are more correlated with the aggregate labor market, correlations of 0.74 and 0.68, respectively. These states also exhibit comparable volatility at 2.6 percent (Ohio) and 2.4 percent (Texas). Alternatively, for states with much smaller populations and more idiosyncratic labor markets when compared with the nation, employment statistics are much more volatile. The standard deviation of average annualized monthly employment growth in Alaska is 4.1 percent, and the state-level data shows almost no correlation with the aggregate data (0.08). The diversion is even more stark for a state like West Virginia which exhibits volatility of 7.0 percent, almost four times that of the nation, and shares a correlation of only 0.28 with U.S. employment.

Third, each state's industry composition can differ, making some series censored for some time periods or completey unavailable for the whole sample. When small amounts of data are missing, we linearly interpolate the missing values but note where these occur so future users can make alternative assumptions. The Appendix provides more details.

3 Forecasting State-Level Indicators

To demonstrate how the FRED-SD dataset can be used, we first conduct a set of real-time state-level forecasting exercises. At each forecast origin t, we use the information that would be available that month to forecast the value of the variables of interest at time $t + h$. In

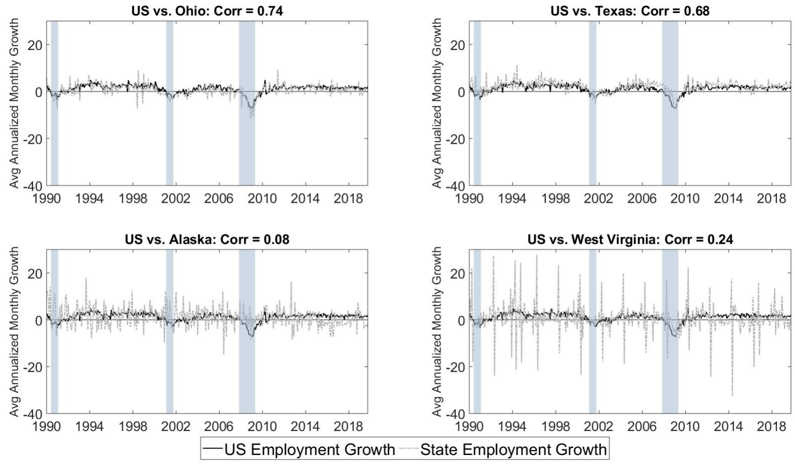

Figure 1: Comparing data on employment growth in the US versus four interesting states: Ohio, Texas, Alaska, and West Virginia

separate out-of-sample experiments, we forecast monthly state-level employment growth and the monthly state-level unemployment rate at horizons of $h = \{1, 3, 6, 12\}$.[8] We then compare these to pseudo-out-of-sample experiments that one would conduct by using only the latest vintage of data.[9]

In all cases, we generate the multi-step-ahead forecasts by iterating forward one-step-ahead forecasts.

3.1 The State-Level Forecasting Models

We consider both smaller models that minimize the number of parameters and larger models that exploit the cross-section of state-level variables to forecast employment growth and the unemployment rate. We compare five models: (i) a random walk benchmark; (ii) an $AR(P)$ in each variable; (iii) a state-level bivariate VAR that includes both variables; (iv) a factor-augmented VAR (FAVAR) that utilizes state-industry-level data; and (v) a panel VAR that uses the state cross-section of employment and unemployment data.[10]

[8] In a previous draft, we also considered quarterly forecasts of Gross State Product. These results are now included in the online appendix.

[9] Our approach parallels exercises from Stock and Watson (2002a,b) and McCracken and Ng (2016). For state-level employment, we use the annualized growth rate, $Y_{n,t+1} = 1200 \ln (EMP_{n,t+1}/EMP_{n,t})$; for the unemployment rate, we use the average annualized monthly changes, $Y_{n,t+1} = 12 (UR_{n,t+1} - UR_{n,t})$.

[10] The VAR emphasizes the dependence of the within-state variables. The FAVAR considers the predictive ability of state-industry level data. The panel VAR exploits cross-state relationships. While we do not view

The first three models are straightforward and not reviewed here. Here, we outline the FAVAR and the panel VAR. Let $\mathbf{y}_{nt}$ reflect the time$-t$, state$-n$ vector of variables of interest, which include the employment growth rate and the change in the unemployment rate. We forecast each state's outcomes independently, where state n's FAVAR is written in state-space form. Let $\mathbf{F}_{nt}$ reflect a time$-t$ vector of K latent factors for state n that are dynamically related via a VAR:

$$\begin{bmatrix} \mathbf{F}_{nt} \\ \mathbf{y}_{nt} \end{bmatrix} = \Phi_n(L) \begin{bmatrix} \mathbf{F}_{n,t-1} \\ \mathbf{y}_{n,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{nt}^F \\ \mathbf{e}_{nt}^y \end{bmatrix},$$

where $\mathbf{e}_{nt} = \left[\mathbf{e}_{nt}^F, \mathbf{e}_{nt}^y\right] \sim N(0, \Omega_n)$ and Ω_n is assumed to be diagonal and uncorrelated with other states. The factors are obtained from the measurement equation:

$$\mathbf{X}_{nt} = \begin{bmatrix} \Lambda_n^F & \Lambda_n^y \end{bmatrix} \begin{bmatrix} \mathbf{F}_{nt} \\ \mathbf{y}_{nt} \end{bmatrix} + \boldsymbol{\varepsilon}_{nt},$$

where $\mathbf{X}_{nt}$ is a vector containing the period$-t$ values of state-sector–level employment growth and $\boldsymbol{\varepsilon}_{nt}$ has a diagonal covariance matrix. We restrict $\Lambda_n^Y = 0$ so that the latent factors capture any comovement among the disaggregate data.

Let $\mathbf{y}_t = [\mathbf{y}'_{1t}, ..., \mathbf{y}'_{Nt}]'$ reflect the stacked vector of the N states' data on the variables of interest: the employment growth rate and the change in the unemployment rate in the model for the labor markets. The panel VAR has the form:

$$\mathbf{y}_t = A(L)\,\mathbf{y}_{t-1} + \mathbf{u}_t,$$

where we leave the dynamic and static cross-correlations unrestricted. The intention here is to exploit the cross-state information.

---

these relationships as structural, they may be enlightening nevertheless.

3.2 Estimation and Evaluation

For each vintage, we estimate the models using all of the data available at that time to construct real-time, iterated multi-step forecasts.[11] To estimate the VARs, we use the procedure outlined in Giannone, Lenza and Primiceri (2015; henceforth GLP) that imposes a prior to shrink the dimensionality of the VARs.[12] We construct point forecasts using the parameter values at the posterior mode, iterating forward to the desired forecast horizon. To estimate the factor model, we construct the factors using principal components and then estimate the VAR using the method described above. The factor loadings are obtained simultaneously with the principal components.

We start the monthly recursive-window estimation with the first vintage that contains state-level employment data (2007:06). Our first sample includes observations from 1990:01-2007:05. Taking the parameter estimates for the five models, we forecast these variables for 2007:06, 2007:08, 2007:11, and 2008:05. Then, we roll forward to the 2007:07 vintage with observations from 1990:01-2007:06 and construct our forecasts for 2007:07, 2007:09, 2007:12, and 2008:06. We repeat this real-time forecasting exercise for every vintage through the final, 2019:12.[13]

To highlight the contribution of the real-time vintage dataset, we compare the real-time vintage results to pseudo-out-of-sample results that use final-vintage data. Our intention is to demonstrate how using revised data can affect forecasting results. For the pseudo-out-of-sample experiments, we forecast $y_{t+h|t}$ using final vintage (period T) data up through t to estimate the model and forecast using the same final vintage data to forecast y_{t+h}. Thus, the pseudo-out-of-sample experiments use revised data, while the real-time out-of-sample experiments use the actual data that was available to the forecaster. We seek to determine whether using revised data produces qualitatively different results.

To illustrate how revisions affect the data, Figures 2 and 3 plot the value of each date's

[11] To ensure convergence, we utilize multiple chains of the sampler. The reported results are the average of ten chains.

[12] GLP chooses the optimal shrinkage by maximizing the marginal data density. We use all default settings except for adjusting the application for use with stationary data.

[13] We use the BIC to select the number of lags or lags at each vintage for each model. For the panel VARs, the lag length is set to 1 due to the large dimension of the dataset.

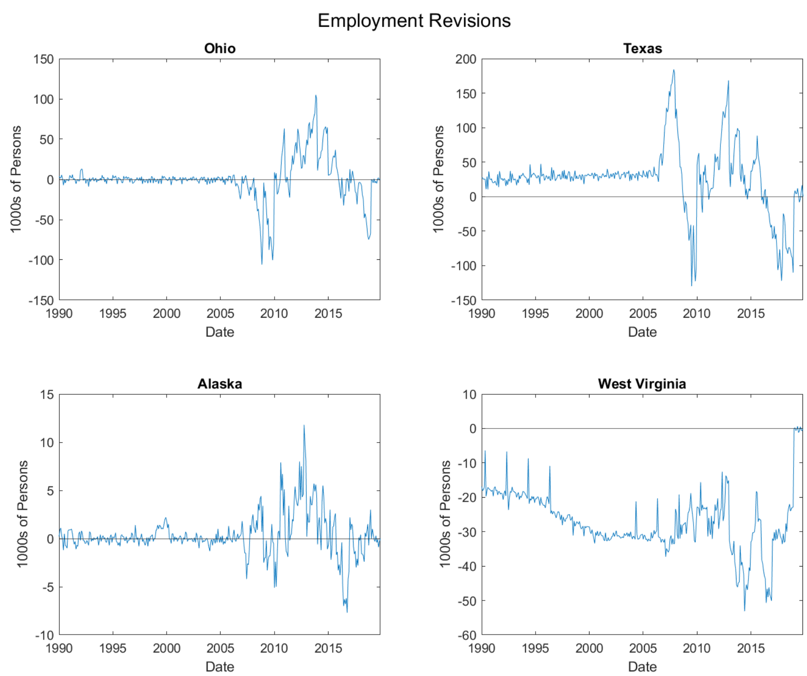

Figure 2: Revisions in state-level employment data between the initial release and the December 2019 vintage. The y-axis gives the level of the overall revision, in 1000's of persons. The x-axis shows the date corresponding to the data.

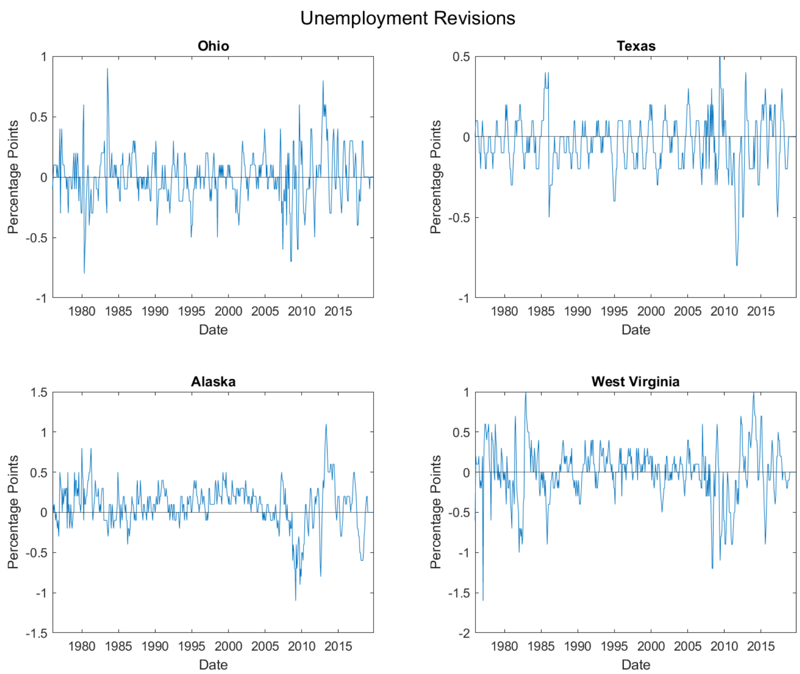

Figure 3: Revisions in state-level unemployment data between the initial release and the December 2019 vintage. The y-axis gives the level of the overall revision, in percentage points. The x-axis shows the date corresponding to the data.

revision from initial release until our end-of-sample vintage (December 2019) for state-level employment and unemployment data, respectively, for the four states from Figure 1. The y-axis shows the total revision—in 1000's of persons for employment and percentage points for unemployment. The x-axis shows the date corresponding to the data. Figure 2 shows an increase in the magnitude and variability of the revisions of employment since 2010. Figure 3 shows that revisions to unemployment show less of a systematic bias in these states. These figures suggest that data revisions are not white noise and may affect the results of pseudo-out-of-sample forecasting experiments that use final vintage data.

3.3 Results

Each of the columns labeled (i) through (v) of Table 3 corresponds to one of the five models outlined above; each row depicts a different forecast horizon. The values in each cell show the number of states that the corresponding model produces the minimum MSE computed from out-of-sample experiments. The seventh column shows the number of states where the best model for revised data differs from the best model for real-time data. The last column reports the number of data releases for which we compare forecasted values with the truth in each of these exercises. We evaluate forecasting performance over 128-139 observations, depending on the forecast horizon.[14] The top panel shows results for state-level employment and the bottom panel shows results for the unemployment rate. The mean squared error of the forecasts for the five models is computed using the initial release of the data as the truth for the real-time experiments.[15] For the final-vintage experiments, the truth is taken from that vintage.

Figures 4 and 5 show the geographical distribution of the minimum MSE forecasting model for each state's employment and unemployment data, respectively. The left column of each figure shows the minimum MSE model when using real-time data and the right column shows the results with revised, final-vintage data. The rows show different horizons. In many cases,

[14] Because of the large number of U.S. states, we do not report the MSE for each model/state combination; these are available in the online appendix.

[15] We also used the one-year-later revision and the latest vintage release as the truth. Results were qualitatively similar and are available in the online appendix.

Table 3: Most Accurate Forecasting Method Across States							
State-Level Employment Growth (Monthly)							
Forecast Horizon	(i) **RW**	(ii) **AR**	(iii) **State VAR**	(iv) **FAVAR**	(v) **Panel VAR**	**# Incorrect w/ Revised Data**	**# Releases**
$h = 1$	0	36	8	0	6	31	139
$h = 3$	0	30	6	2	12	22	137
$h = 6$	0	32	3	1	14	24	134
$h = 12$	0	30	7	4	9	29	128
State-Level Unemployment Rate (Monthly)							
Forecast Horizon	(i) **RW**	(ii) **AR**	(iii) **State VAR**	(iv) **FAVAR**	(v) **Panel VAR**	**# Incorrect w/ Revised Data**	**# Releases**
$h = 1$	3	7	18	19	3	37	139
$h = 3$	0	3	10	10	27	40	137
$h = 6$	0	9	7	14	20	25	134
$h = 12$	0	14	12	11	13	32	128

Table 3: Number of states for which each forecasting method produces the minimum mean squared error for state-level employment growth and the unemployment rate at forecasting horizons of 1, 3, 6, and 12 months. We include the number of observations for which the forecasts are evaluated. We compare the accuracy of the various forecasting models against the initial release of the data. The State VAR for the labor market variables includes data on employment and the unemployment rate in a given state. The factor-augmented VAR (FAVAR) uses disaggregate employment data for each state. The panel VAR uses the full state cross-section of employment and unemployment data. The final column indicates the number of states, out of 50, for which the revised data would suggest an incorrect identification of the most accurate forecasting method.

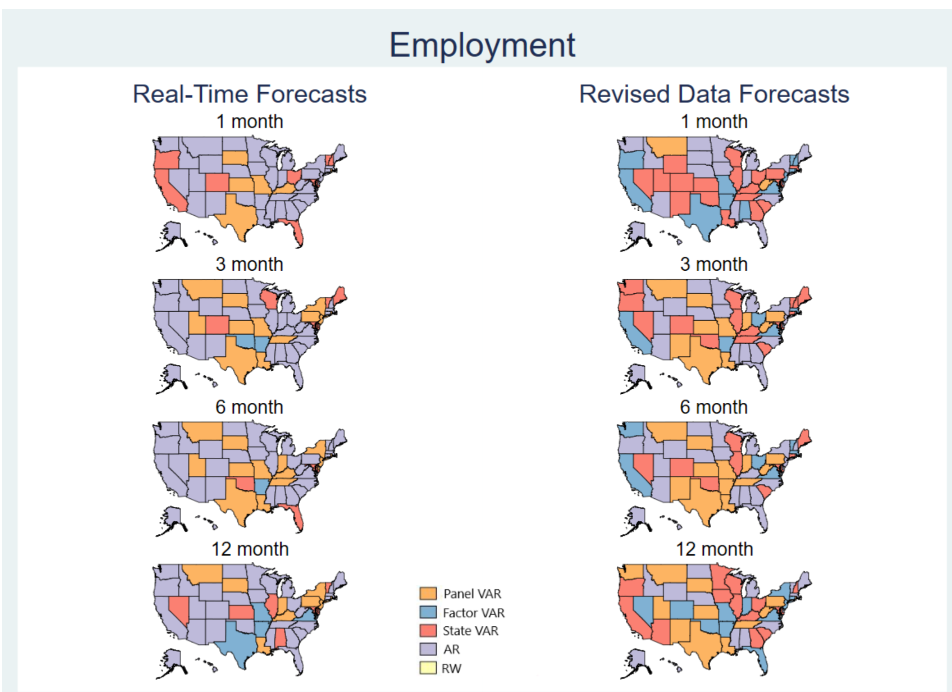

Figure 4: Best Forecasting Model for Each US State: Employment

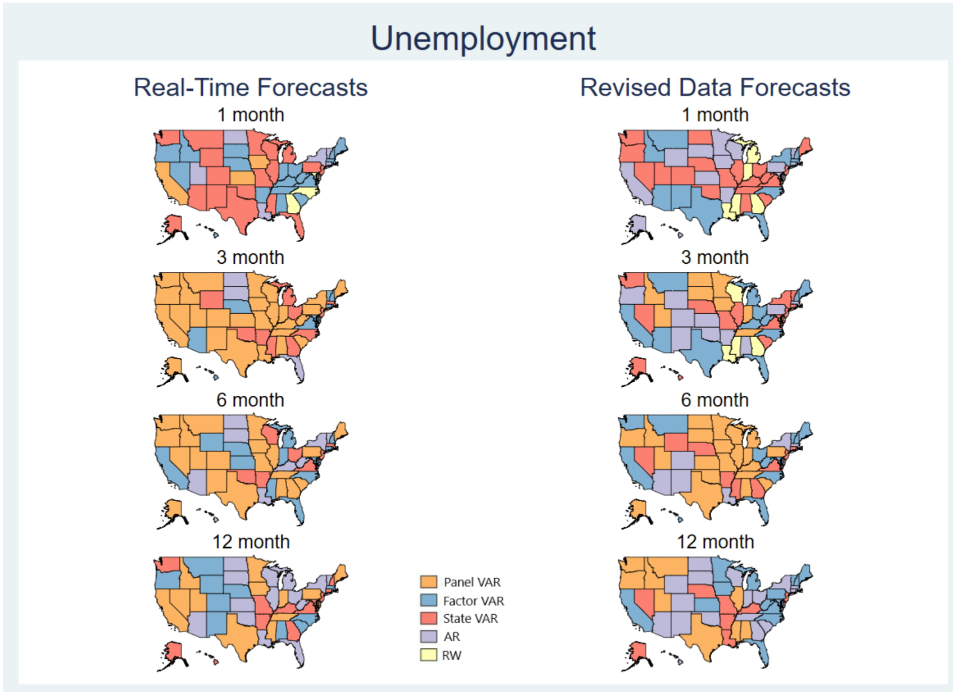

Figure 5: Best Forecasting Model for Each US State: Unemployment

the best forecasting model appears geographically clustered. We first discuss the real-time results and then will contrast this with conclusions one would make with the revised dataset.

For employment growth, the univariate AR outperforms other models for most states at most horizons; the panel VAR outperforms all other models for about 20% of the states. In particular, for Montana, South Dakota, Kansas, Kentucky, Tennessee, Missouri, New York, and Texas, the panel VAR outperforms all other models for at least three out of the four reported horizons. Montana and South Dakota are both relatively small states in population. Kansas, Missouri, Tennessee, Kentucky, and New York all have large MSAs that overlap with other states. For these reasons, these states may be sensitive to the economic conditions in neighboring states, resulting in better forecasting performance of the model that accounts for the co-movement between states. For unemployment, models incorporating more information—in particular, the panel VAR—perform better on average.

Figures 6 and 7 contain spider plots for the same states that we compared to the U.S. national data above for employment growth and the unemployment rate, respectively. Each

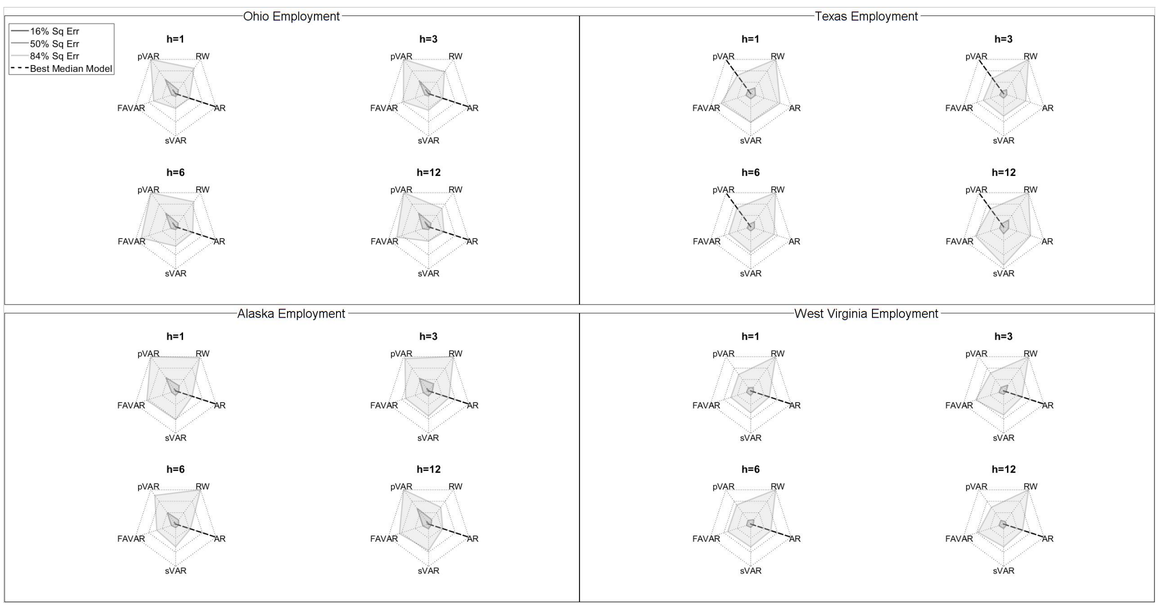

Figure 6: State-Level Forecasting Performance for Employment: 16-, 50-, and 84-percentiles of the distribution of squared errors for a few select states (Ohio, Texas, Alaska, West Virginia). The node closest to the center corresponds to the model with the lowest median squared error indicated in Table 2, and the corresponding model is highlighted with the dashed line. The State VAR (sVAR) includes data on employment and the unemployment rate in a given state. The factor-augmented VAR (FAVAR) uses disaggregate employment data for each state. The panel VAR (pVAR) uses the full state cross-section of employment and unemployment data.

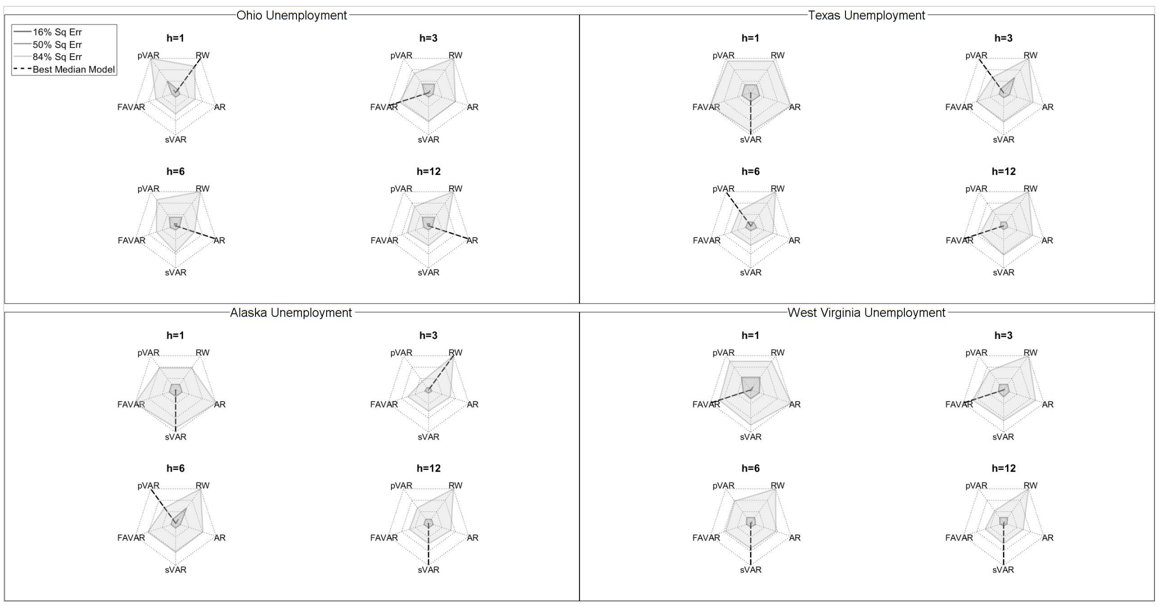

Figure 7: State-Level Forecasting Performance for Unemployment: 16-, 50-, and 84-percentiles of the distribution of squared errors for a few select states (Ohio, Texas, Alaska, West Virginia). The node closest to the center corresponds to the model with the lowest median squared error indicated in Table 2, and the corresponding model is highlighted with the dashed line. The State VAR (sVAR) includes data on employment and the unemployment rate in a given state. The factor-augmented VAR (FAVAR) uses disaggregate employment data for each state. The panel VAR (pVAR) uses the full state cross-section of employment and unemployment data.

node plots the 16th, 50th, and 84th percentiles of the squared errors for a single model for the state of interest. Dark dashed lines highlight the best-performing model according to the median-squared error, corresponding to the node closest to the origin; the node furthest from the center corresponds to the worst-performing model.

For two of the four states, the models perform similarly for both of the labor market indicators; Alaska (all horizons) and West Virginia (one-period-ahead) are exceptions. For Alaska, employment is better predicted by smaller models (e.g., the univariate AR) but unemployment is better predicted by larger models like the panel VAR. Overall, the plots show that, in many cases, the predictive differences across some models can be small. In particular, for employment growth in Ohio (top left), the AR and state VAR models forecast comparably at the $h = \{1, 12\}$ month horizons, each reducing the MSE about 40 percent compared with the random walk benchmark. For Texas employment growth (top right), the panel VAR is most accurate for most horizons (reducing the MSE from 32 to 49 percent versus the random walk) but the FAVAR performs similarly.

The sixth column of Table 3 suggests that conclusions regarding the best-performing forecasting model would be different—and, more importantly, inaccurate—for the majority of states. For $h = 1$, using real-time or final vintage data would identify different models as the most accurate forecast of employment (unemployment) for 31 (37) states. Comparing the left and right columns of Figures 4 and 5 illustrates the extent to which using revised, final-vintage data might affect one's conclusions. For nine (ten) states, the best model is misidentified using final-vintage data for employment (unemployment) at all horizons.

3.4 What Might Explain Model Predictability Across States?

Are there state-level differences that might help determine which models have superior predictability across states? The models that we use can highlight intrastate dependence or interstate dependence; univariate models might have higher predictability in states that are subject to more idiosyncratic fluctuations.

We then consider whether two state-level labor market features affect the probability that a model will forecast best for that state: employment diversity and cross-state trade. Our hypothesis is that greater industrial diversity within a state increases the likelihood that accounting for fluctuations in the state's industrial employment will have higher predictive ability. Thus, greater employment diversity should favor the factor model. Similarly, more cross-state trade should increase the correlation with other states, favoring the panel VAR.

To explore these hypotheses, we take the lowest MSE model as given and estimate a multinomial logit using the 5 models as the LHS variable. The RHS variables consist of a measure of state industry employment diversity and a measure of state-to-state trade. We consider two measures of Employment Diversity: the Hachman Index (Benway, 2019) and an Entropy Index (Smith and Gibson, 1988). The Hachman Index measures state employment diversity relative to the U.S., taking values from 0 to 1, where 1 indicates industrial similarity the U.S. (i.e., diverse).[16] The Entropy Index compares the existing employment across industries in the state to an equiproportional distribution.[17] Higher entropy indicates greater relative diversification and lower entropy indicates more specialization. Each state-level index uses employment data from 2000. The measure of state-to-state trade is taken from the Commodity Flow Survey conducted by the Census.

Table 4 shows the marginal effects from multinomial logit regressions of the winning model (excluding the random walk) for different horizon employment forecasts on employment diversity and state-to-state trade. The left panel shows the results using the Hachman index; the right panel shows the results for the Entropy index. While results for the Hachman index are not conclusive, results for the entropy index are more promising. While not necessarily consistent at all horizons, the salient result is that information about economic conditions in other states are more predictive for states with industrial diversity and large cross-state trade flows. States that are industrially diverse may look more like other states (or the nation) and

[16] The Hachman index is calculated from $H = 1/\sum_{i=1}^{N}[(S_i^{state}/S_i^{US})S_i^{state}]$, where $S_i^{state}$is the share of total employment of industry i in a state, S_i^{US} is the share of total employment of industry i in the U.S., and N is the number of industries.

[17] The Entropy Index is calculated from $E = \sum_{i=1}^{N} S_i \ln(\frac{1}{S_i})$, where S_i is the share of total employment of industry i in a state and N is the number of industries.

may exhibit similar dynamics. States that trade more with other states have economies that may depend on the activity in those states. In both cases, economic activity in other states may be informative, consistent with these results.

Table 4: Marginal Effects							
	Employment	**Hachman**	**Trade**		**Employment**	**Entropy**	**Trade**
h=1	AR	−0.11	0.04	**h=1**	AR	**−0.11**	**0.11**
	State VAR	0.07	−0.06		State VAR	0.05	**−0.14**
	Factor	–	–		Factor	–	–
	Panel VAR	0.03	0.02		Panel VAR	0.06	0.03
h=3	AR	−0.11	−0.05	**h=3**	AR	**−0.16**	−0.08
	State VAR	**0.05**	−0.01		State VAR	−0.01	−0.01
	Factor	−0.01	−0.02		Factor	0.01	−0.01
	Panel VAR	0.08	0.08		Panel VAR	**0.16**	**0.10**
h=6	AR	−0.07	−0.07	**h=6**	AR	**−0.13**	−0.09
	State VAR	0.00	−0.03		State VAR	0.00	−0.03
	Factor	0.00	0.00		Factor	0.01	0.00
	Panel VAR	0.07	0.10		Panel VAR	**0.14**	**0.12**
h=12	AR	−0.05	0.00	**h=12**	AR	−0.02	0.00
	State VAR	−0.01	−0.08		State VAR	−0.06	**−0.09**
	Factor	0.06	0.02		Factor	0.04	0.03
	Panel VAR	0.00	0.06		Panel VAR	0.03	0.06

Table 4: Marginal effects from multinomial logit regressions of the winning model (excluding the random walk) for forecasts of employment growth at the 1, 3, 6, and 12-month horizons. Regressors include employment diversity and state-to-state trade. The State VAR includes data on employment and the unemployment rate in a given state. The factor-augmented VAR (FAVAR) uses disaggregate employment data for each state. The panel VAR uses the full cross-section of employment and unemployment data across states.

4 Forecasting National-Level Macroeconomic Indicators

Hernández-Murillo and Owyang (2006) and Owyang, Piger and Wall (2015), among others, showed that regionally-disaggregated data can predict national-level data. Here, we consider whether state-level data predicts national-level payroll employment.[18] Because national-level data are (typically) not exactly equal to the sum of the state-level data, we level the playing field by replacing the BEA's reported national series with the sum of the state-level data—in this case, state-level payroll employment.[19]

[18] In a previous draft, we also considered quarterly forecasts of Gross Domestic Product. These results are now included in the online appendix.

[19] The BEA explains that the national series and the sum of the state-level series are not equal because, among other things, foreign civilian and military output is included in the national series but not any state

We construct forecasts of the growth rates of the disaggregated data series and convert these into level forecasts using the values at the forecast origin. Let Y_{nt} and y_{nt} represent the level and growth rate of the state-level variable of interest, respectively; $Y_t = \sum Y_{nt}$ and y_t represent the aggregated level and growth rate of the variable of interest, respectively; and $\mathbf{y}_t = [y_{1t}, ..., y_{Nt}]'$ represent the vector collecting the individual state-level growth rates.

As benchmarks, we consider a random walk and an $AR(P)$ model of each national variable. We also consider two models intended to take advantage of the large cross-section of state-level data: a block factor model that includes a national-level factor and a set of state- and/or regional-level factors and a space-time autoregression, first proposed by Giacomini and Granger (2004) and applied to state-level data in Hernández-Murillo and Owyang (2006). We also consider a model-average forecast that combines forecasts from the other models.

4.1 The National-Level Forecasting Models

The factor model is similar to the FAVAR outlined in the previous section but applied to cross-state data. Because the vector of measurement data is larger, we impose restrictions on the factor loadings to interpret them as state- or regional-level factors. Define G_t, a vector of K^G latent national factors, and R_t, a vector of K^R latent regional factors, and their loadings, Λ^G and Λ^R, respectively. The latent factors are determined from the common movements in a large vector of state-sector data, X_t:

$$X_t = \Lambda^G G_t + \Lambda^R R_t + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \Sigma)$.[20] We assume Σ is diagonal, so that the factors capture all of the cross-sectional comovements in the elements of X_t.

Once we obtain the latent factors, we can consider the dynamic relationship between the

series.

[20] Restrictions on the loading matrices allow us to interpret the factors. The national factors Λ^G are unrestricted. We set $K^R = 8$ and construct regions based on those defined by the BEA. Thus, the factor for region n does not load on variables in other BEA regions. In this case, the national factors account for all of the correlation across the entire vector of variables, while the regional factors account for cross-state correlations occurring within a particular BEA region.

variables of interest $\mathbf{y}_t$ and the unobserved factors, G_t and R_t, in a factor-augmented VAR:

$$\begin{bmatrix} G_t \\ R_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \Phi_{GG}(L) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_{RR}(L) & \mathbf{0} \\ \Phi_{YG}(L) & \Phi_{YR}(L) & \Phi_{YY}(L) \end{bmatrix} \begin{bmatrix} G_{t-1} \\ R_{t-1} \\ \mathbf{y}_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^G \\ e_t^R \\ e_t^y \end{bmatrix},$$

where the covariance matrix $E_t e_t e_t' = \Omega$ is assumed to be diagonal.[21] The VAR coefficients are restricted such that the factors are only functions of their own lags; the elements of $\mathbf{y}_t$ are functions of their own lags, lags of the national factor, and lags of their respective regional factor. Thus, $\Phi_{GG}(L)$ and $\Phi_{RR}(L)$ are restricted to be diagonal; $\Phi_{YG}(L)$ is left unrestricted; the (m,n)-element of $\Phi_{YR}(L)$ is restricted to 0 if state m is not in region n; and $\Phi_{YY}(L)$ is restricted to be diagonal. As indicated above, we construct forecasts of the individual state data $\mathbf{y}_t$ and then sum these to obtain a national forecast.

The fourth model is the space-time autoregression proposed by Giacomini and Granger (2004):

$$\mathbf{y}_t = \phi(L)\,\mathbf{y}_{t-1} + \psi(L)\,W\mathbf{y}_{t-1} + \mathbf{e}_t,$$

where $\phi(L)$ and $\psi(L)$ are *scalar* polynomials in the lag operator, W is a spatial weighting matrix with elements:

$$w_{nm} = \begin{cases} 0 & if \quad n = m \\ 1/C_n & \quad otherwise \end{cases},$$

where C_n is the number of states contiguous with state n. In the cases of Alaska and Hawai'i, we set all weights equal to zero. Thus, the forecast for each state is composed of two components: (i) an AR process that is the same for all states and (ii) an AR process that reflects the spillover effect from contiguous states.

[21] Given that the Ω matrix is of dimension $(50 + K^G + K^R) \times (50 + K^G + K^R)$, allowing for correlation across all factors and state-level variables would require estimation of $\frac{1}{2}(50 + K^G + K^R)(50 + K^G + K^R - 1)$ elements. This is possible but is computationally intensive and we opt to model innovations to each element of the factor-augmented VAR separately. We leave this extension for future work.

The model-average forecast is computed as a weighted sum of the forecasts of a subset of the other models. The weight on model i is determined by $l\,(.|M_i)$, the log likelihood of model i,

$$\omega_i = \frac{l\,(Y|M_i)}{\sum_j l\,(Y|M_j)}.$$

Note that because, by construction, the aggregate level is the sum of the disaggregate levels, the aggregate log likelihood is the weighted sum of the disaggregate log likelihoods.

4.2 Estimation

For the factor-augmented VAR, the factors are first constructed using principal components; then, the VAR is estimated using Bayesian methods, using a Normal-Inverse Gamma prior. Based on the restrictions on the factor loadings, the factors are estimated in a hierarchy: The national factor(s) are estimated from the full vector of data, Y_t; the effects of the national factor(s) are removed; and, finally, the regional factors are estimated from subsets of the state-level data associated with each BEA region.

The space-time AR is also estimated using a Normal-Inverse Gamma prior, conditional on a fixed weighting matrix. Conditional on the fixed weighting matrix, the posteriors are conjugate.

For the national-level forecasting application, evaluation is conducted as described in Section 3 for the state-level forecasting exercises. We use the BIC to select the number of lags in the AR model, $1 \leq P \leq 6$ at each vintage. For the factor-augmented VAR and the space-time AR, we set the lag length equal to 1 due to the dimensionality of the dataset. We consider forecast horizons of $h = \{1, 3, 6, 12\}$ months and construct point estimates for $E_t Z_{t+h|t}$ via iterated, multi-step forecasts using the posterior mean parameter values.

As with the state-level forecasting exercises above, we redo the national experiments using revised, final-vintage data to demonstrate how the results would differ from real-time versus revised datasets.

4.3 Results

Table 5 reports the MSEs of the national-level forecasting models relative to the random walk benchmark for aggregate payroll employment. For payroll employment, the simple AR produces the most accurate forecasts at all horizons. The two disaggregate forecasting models improve upon the random walk, reducing the MSE by between 14 and 40-percent, depending on the horizon. This suggests that state-level employment data has some predictive value for national employment; however, because we chose iterative rather than direct forecasting methods, disaggregation generally cannot beat the simplicity of the AR.[22]

The last two columns of the top panel show the relative MSEs for the model-averaging approach. The model average in the second to last column combines all other models. Because we compute the likelihood based on one-period-ahead model fit, the random walk seems to be overweighted. Thus, the last column includes a model average forecast that excludes the random walk. For payroll employment, the forecasting performance of the model average excluding the random walk is similar to the other disaggregated model but still not comparable to the simpler AR. We conjecture that an alternative weighting scheme that places more weight on the simpler models would likely perform better.

Table 5: MSE Comparison for National-Level Forecasting							
Payroll Employment (Monthly, Real-Time Data)							
Forecast Horizon	**# Obs**	**RW**	**AR**	**Factor Model**	**Space-Time**	**Model Avg**	**Model Avg ex. RW**
$h = 1$	139	1.00	0.70	0.87*	0.86	1.00	0.85*
$h = 3$	137	1.00	0.51	0.71*	0.74*	0.99*	0.72*
$h = 6$	134	1.00	0.47	0.64*	0.66*	0.99	0.64*
$h = 12$	128	1.00	0.40	0.61	0.58	0.98	0.58

Table 5: Relative MSE, compared with RW, for all national-level forecasting models. We compare the accuracy of the various models against the initial release of the data. The factor model is hierarchical and includes PCA factors extracted from data on all state-level employment data. The factors are extracted from states grouped by BEA region. The space-time autoregression uses spatial weighting for employment data across states to forecast the national variables. We forecast the growth rate for each individual state and use these to construct forecasts of the level of employment in each state. Then we aggregate by summing across states and treat this as the national-level measure which we are forecasting. An asterisk indicates a model for which the conclusion regarding the performance relative to the random walk benchmark would be incorrect with revised data, relative to the real-time results.

[22] These results are consistent with Miller's (1998) finding that aggregating regions for a state-level "bottom-up" forecast is less accurate than a direct univariate forecast, while using state-level data for regional "top-down" forecasts is more accurate than a direct univariate forecast of each region.

To highlight the benefit of using real-time vintages, we rerun the experiments using final-vintage data. Recall that numbers in Table 5 less than 1 reflect models that have lower MSE than the random walk. An asterisk (*) in Table 5 identifies a model that would have higher MSE than the random walk had the experiment been conducted using revised data. The factor model and model-average improve forecast accuracy over the random walk for all forecasting horizons with real-time data. However, with final-vintage data, these models appear to have less predictive ability than the random walk at forecasting horizons of $h = \{1, 3, 6\}$ months. Similarly, with real-time data, the space-time AR improves upon the MSE of the random walk over all horizons, but appears to perform worse for $h = \{3, 6\}$ using final-vintage data.

As a final example of the difference between real-time vintages and revised, final vintages, Figure 8 plots the correlations between expanding-window estimates of the factors from the national forecasting experiments. The factors are estimated using three datasets: (i) real-time vintages; (ii) final-vintage data up through the forecast origin (pseudo-real-time); and (iii) the full final-vintage time series (in-sample). The correlation between real-time versus pseudo-out-of-sample data in each window is larger than that between real-time versus in-sample data. While some of the pseudo-out-of-sample factors are highly correlated with the out-of-sample factors, those corresponding to the Plains, Southeast, and Southwest regions show reduced correlation, suggesting a substantial effect of data revisions.

5 Conclusion

The recent surge in the use of spatially-disaggregated data in macroeconomics has given rise to the need for a collection of real-time state-level data suitable for forecasting experiments. This paper demonstrates the use of such a dataset, constructed, housed, and, most importantly, consistently updated at the Federal Reserve Bank of St. Louis. While the data are not per se new, the dataset eases the burden on researchers, collecting and documenting approximately 28 series per state in real-time vintages.

We show that these data can be useful for both state-level and aggregate-level forecasting. For state-level forecasting, the benefits of such a dataset are transparent. One requires

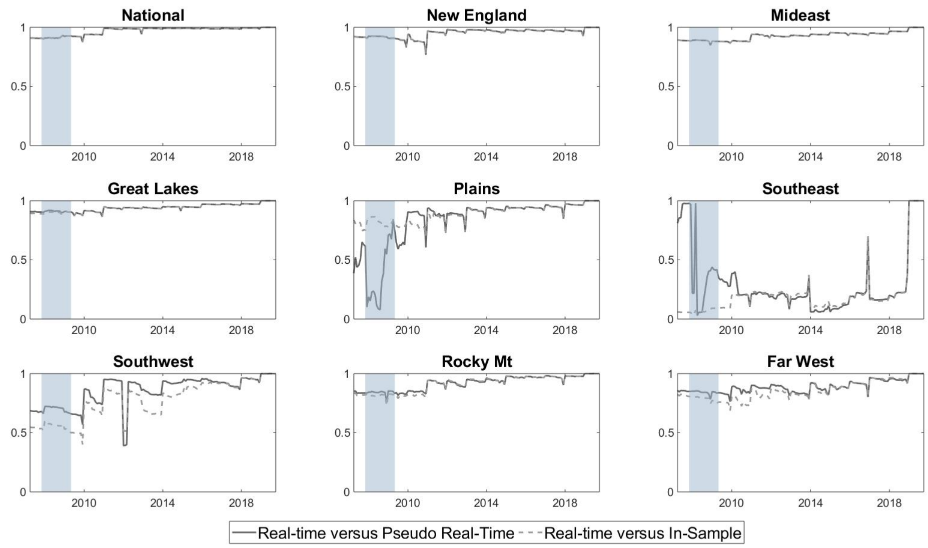

Figure 8: Correlation between factor estimates using real-time vintage data or revised, end-of-sample data. “Real-time versus Pseudo Real-Time” compares factor estimates using data from each vintage or the final-vintage data through the forecast origin - datasets (i) versus (ii). “Real-time versus In-Sample” compares factor estimates using data from each vintage or the full final-vintage sample - datasets (i) versus (iii).

vintage data to construct out-of-sample forecast experiments. At the state-level, these experiments demonstrate the heterogeneity across states based, perhaps in part, on the industry concentration or industrial composition of the states.

At the aggregate level, the forecasting results for the simple experiments that we performed here are mixed. However, unlike some of the past literature, we aggregated the state-forecasts using their respective employment or output weights—that is, a state's contribution to the national employment forecast is determined by its share of national employment. Other papers (e.g., Owyang, Piger, and Wall, 2015) use estimated weights that might perform better. Rapach and Strauss (2005), forecasting Missouri employment growth at several horizons, find that combination methods work well at shorter horizons but that shrinkage methods work better for longer horizons. Combining these techniques with cross-sectional information may result in better predicitive models.

More importantly, we highlight how using final-vintage data differs from using a dataset with real-time vintages that replicates the data available to the forecaster at each forecast origin. In particular, we show that the conclusions one draws from pseudo-out-of-sample experiments can differ substantially from those from true out-of-sample experiments. In state-level forecasting experiments, pseudo-out-of-sample experiments identify the wrong "best" forecasting model for the majority of states.

References

[1] Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. "Large Bayesian Vector Auto Regressions," *Journal of Applied Econometrics*, 25(1), 2010, pp. 71-92. https://doi.org/10.1002/jae.1137

[2] Benway, DJ. "Measuring Economic Diversity: The Hachman Index, 2017," Kem C. Gardner Policy Institute Research Brief, February 2019. https://gardner.utah.edu/wp-content/uploads/HachmanBriefFinal.pdf?x71849

[3] Bernanke, Ben S. and Jean Boivin. "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics*, 50(3), 2003, pp. 525-546. https://doi.org/10.1016/S0304-3932(03)00024-2

[4] Boivin, Jean and Serena Ng. "Understanding and Comparing Factor-Based Forecasts," *International Journal of Central Banking*, 1(3), 2005, pp. 117-151. https://www.ijcb.org/journal/ijcb05q4a4.htm

[5] Crone, Theodore M. and Alan Clayton-Matthews. "Consistent Economic Indexes for the 50 States," *The Review of Economics and Statistics*, 87(4), 2005, pp. 593-603. https://doi.org/10.1162/003465305775098242

[6] Croushore, Dean and Tom Stark. "A Real-time Data Set for Macroeconomists," *Journal of Econometrics*, 105(1), 2001, pp. 111-130. https://doi.org/10.1016/S0304-4076(01)00072-0

[7] Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100(471), 2005, pp. 830-840. https://doi.org/10.1198/016214504000002050

[8] Giacomini, Raffaella, and Clive W.J. Granger. "Aggregation of Space-Time Processes," *Journal of Econometrics*, 118, 2004, pp. 7-26. https://doi.org/10.1016/s0304-4076(03)00132-5

[9] Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97(2), 2015, pp. 412-435. https://doi.org/10.1162/REST_a_00483

[10] Hernández-Murillo, Rubén and Michael T. Owyang. "The Information Content of Regional Employment Data for Forecasting Aggregate Conditions," *Economics Letters*, 90(3), 2006, pp. 335-339. https://doi.org/10.1016/j.econlet.2005.08.023

[11] Lehman, Robert and Klaus Wohlrabe. "Regional Economic Forecasting: State of the Art Methodology and Future Challenges," *Economics and Business Letters*, 3(4), 2014. https://doi.org/10.17811/ebl.3.4.2014.218-231

[12] LeSage, James P. and Daniel Hendrikz. "Large Bayesian vector autoregressive forecasting for regions: A comparison of methods based on alternative disturbance structures," *The Annals of Regional Science*, 62(3), 2019, pp. 563-599. https://doi.org/10.1007/s00168-019-00908-z

[13] LeSage, James P. and Michael Magura. "A Regional Payroll Forecasting Model That Uses Bayesian Shrinkage Techniques for Data Pooling," *Journal of Regional Analysis and Policy*, 18(1), 1988, pp. 1-17. https://doi.org/10.22004/ag.econ.128961

[14] LeSage, James P. and Zheng Pan. "Using Spatial Contiguity as Bayesian Prior Information in Regional Forecasting Models," *International Regional Science Review*, 18(1), 1995, pp. 33-53. https://doi.org/10.1177/016001769501800102

[15] Marcellino, Massimiliano, James H. Stock and Mark W. Watson. "Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-Wide Information," *European Economic Review*, 47(1), 2003, pp. 1-18. https://doi.org/10.1016/S0014-2921(02)00206-4

[16] McCracken, Michael W. and Serena Ng. "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34(4), 2016, pp. 574-589. https://doi.org/10.1080/07350015.2015.1086655

[17] Miller, Jon R. "Spatial aggregation and regional economic forecasting," *The Annals of Regional Science*, 32(2), 1998, pp. 253-266. https://doi.org/10.1007/s001680050073

[18] Owyang, Michael T., Jeremy Piger, and Howard J. Wall. "Forecasting National Recessions Using State-Level Data," *Journal of Money, Credit and Banking*, 47(5), 2015, pp. 847-866. https://doi.org/10.1111/jmcb.12228

[19] Rapach, David E. and Jack K. Strauss. "Forecasting Employment Growth in Missouri with Many Potentially Relevant Predictors: An Analysis of Forecast Combining Methods," *Federal Reserve Bank of St. Louis Regional Economic Development*, 1(1), 2005, pp. 97-112.

[20] Smith, Stephen M., and Cosette M. Gibson. "Industrial diversification in nonmetropolitan counties and its effect on economic stability," *Western Journal of Agricultural Economics*, 13(2), 1988, pp.193-201.

[21] Stock, James H and Mark W. Watson. "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14(1), 1996, pp. 11-30. https://doi.org/10.1080/07350015.1996.10524626

[22] Stock, James H. and Mark W. Watson. "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20(2), 2002a, pp. 147-162. https://doi.org/10.1198/073500102317351921

[23] Stock, James H. and Mark W. Watson. "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 2002b, pp. 1167-1179. https://doi.org/10.1198/016214502388618960

[24] Stock, James H., and Mark W. Watson. "Forecasting with Many Predictors." Handbook of Economic Forecasting, 2006, pp. 515–554. doi:10.1016/s1574-0706(05)01010-4.

A Data Appendix

This version of the appendix summarizes the availability of data and general cleaning procedures used to assemble the FRED-SD database. Sections proceed as follows:

I. Availability of Raw Data

II. Cleaning Raw Data

III. Assigning Vintage Dates

IV. Missing Values

V. Accessing Data

The FRED-SD database includes 28 variables for the 50 U.S. states and Washington D.C. All data is sourced from ALFRED (Archival FRED). Throughout this appendix, note that the date format mm-YYYY (ex. Aug-2019) refers to observation dates and YYYY-MM (ex. 2019-08) refers to vintage dates.

A.1 I. Availability of Raw Data

a. **Substituted Series:** The following series are substituted in the FRED-SD database because the desired series does not exist (in ALFRED).

i. MFGHRS: Manufacturing Hours is substituted with Goods Producing Hours (which aggregates Manufacturing and Mining Hours) in the following series:

1. DCMFGHRS: Manufacturing Hours for Washington D.C.
2. MTMFGHRS: Manufacturing Hours for Montana
3. NVMFGHRS: Manufacturing Hours for Nevada
4. NMMFGHRS: Manufacturing Hours for New Mexico
5. WYMFGHRS: Manufacturing Hours for Wyoming

ii. CONS: Construction Employment is substituted with Mining, Logging and Construction Employment in the following series:

1. DECONS: All employees, Construction for Delaware
2. DCCONS: All employees, Construction for Washington D.C.
3. HICONS: All employees, Construction for Hawaii

b. **Unavailable Series:** The following series are omitted from the FRED-SD database because they do not exist (in ALFRED) and lack adequate substitutes.

i. RENTS: Real Estate, Rental and Leasing Employment

1. NMRENTS: All employees, Real Estate, Rental, Leasing for New Mexico

2. RIRENTS: All employees, Real Estate, Rental, Leasing for Rhode Island

3. SDRENTS: All employees, Real Estate, Rental, Leasing for South Dakota

ii. MINNG: Mining and Logging Employment

1. DEMINNG: All employees, Mining and Logging for Delaware

2. DCMINNG: All employees, Mining and Logging for Washington D.C.

3. HIMINNG: All employees, Mining and Logging for Hawaii

c. **Omitted Series:** The following series are omitted from the FRED-SD database because they are currently discontinued and contain only a small number of observations and vintages:

i. DCCONSTNQGSP: Construction GSP for District of Columbia

ii. DCMANNQGSP: Manufacturing GSP for District of Columbia

iii. DENATURNQGSP: Agriculture and Mining GSP for Delaware

iv. RINATURNQGSP: Agriculture and Mining GSP for Rhode Island

v. RICONSTNQGSP: Construction GSP for Rhode Island

vi. WYAGRNQGSP: Agriculture, Forestry, Fishing and Hunting GSP for Wyoming[i]

vii. WYMANNQGSP: Manufacturing GSP for Wyoming

d. **Variation in FRED Mnemonic:** The FRED mnemonic pattern deviates for the following series.

i. MFGHRS: Average weekly hours, private employees

SMU*XX*000000500000002SA deviates for DC, MT, NV, NM, and WY.

ii. OTOT: Nominal personal income

*XX*OTOT deviates for NC.

iii. CONS: All employees, Construction

*XX*CONS deviates for DE, DC, HI, MD, NE, SD, and TN.

iv. MFG: All employees, Manufacturing

*XX*MFG deviates for AL and DC.

v. FIRE: All employees, Finance, Insurance, Real Estate

*XX*FIRE deviates for DC, HI, MS, and WY.

vi. INFO: All employees, Information

*XX*INFO deviates for CA, HI, NV, NM, OR, RI, VT, WV, and WY.

vii. MINNG: All employees, Mining and Logging

*XX*NRMN deviates for AL, CT, FL, GA, IL, IN, LA, MD, MS, MO NE, NH, NJ NC, RI, SC, SD, TN, and VT.

viii. BPPRIVSA: Permits, New Privately Owned Housing Units

*XX*BPPRIVSA deviates for DC.

e. Variation in Start Dates:

i. **First Observation:** The following series have a first observation date that varies from the date listed in Table 1.

1. OTOT: Nominal personal income

First observation is 1/1/1950 for AK and HI.

2. MINNG: All employees, Mining and Logging

First observation is 1/1/2002 for FL.

First observation is 1/1/2009 for RI.

3. ICLAIMS: Unemployment insurance, initial claims

First observation is 9/1/1985 for CO, SD, UT, and WY.

First observation is 10/1/1985 for MT and ND.

First observation is 1/1/1986 for DC .

4. CONSTNQGSP: Gross Domestic Product by state, Construction

First observation is 1/1/2006 for DE.

5. NATURNQGSP: Gross Domestic Product by state, Agriculture and Mining

First observation is 1/1/2006 for DE.

ii. **First Vintage:** The following series have a first vintage date that varies from the date listed above.

1. NA: All employees, Total nonfarm

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

2. UR: Unemployment rate

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

3. OTOT: Nominal personal income

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

4. GOVT: All employees, Total government

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

5. CONS: All employees, Construction

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

First vintage is 2014-01 for DE, DC, and HI.

First vintage is 2016-03 for MD and NE.

6. MFG: All employees, Manufacturing

First vintage is 2005-06 for AR, IL, IN, KY, MO, and MS.

First vintage is 2007-07 for AK.

First vintage is 2013-08 for DE, OK, and WY.

First vintage is 2015-05 for GA, HI, and TN.

First vintage is 2015-08 for AL and DC.

7. FIRE: All employees, Finance, Insurance, Real Estate

First vintage is 2005-06 for AR, IL, IN, KY, MO, and TN.

First vintage is 2014-01 for DC, HI, MS, and WY.

8. INFO: All employees, Information

First vintage is 2005-06 for IL, IN, KY, and TN.

First vintage is 2012-05 for ME, MS, OK, and SD.

First vintage is 2015-05 for CA and WY

First vintage is 2015-08 for HI, NV, NM, OR, RI, and VT.

9. RENTS:

First vintage is 2017-03 for IA.

10. MINNG: All employees, Mining and Logging

First vintage is 2005-06 for AR and TN.

First vintage is 2008-03 for KS.

First vintage is 2010-11 for KY.

First vintage is 2014-01 for AL, CT, FL, GA, IL, IN, LA, MS, MO, NH, NJ, NC, SC and VT.

First vintage is 2014-03 for RI.

First vintage is 2016-03 for MD and NE.

11. BPPRIVSA: Permits, New Privately Owned Housing Units

First vintage is 2007-11 for DC.

12. LF: Civilian Labor Force

First vintage is 2005-06 for AR, IL, IN, KY, MO, MS, and TN.

13. NATURNQGSP: Gross Domestic Product by state, Agriculture and Mining

First vintage is 2016-11 for DC.

A.2 II. Cleaning Raw Data

a. Adjusting Units

i. OTOT: Nominal personal income

Vintages from 2013-09 to 2018-09 are published in thousands of dollars for all fifty states and DC. These vintages are adjusted to reflect millions of dollars for comparison to previous and future vintages in ALFRED.

ii. LF: Civilian Labor Force

Vintages from 2007-06 to 2016-01 are published in thousands of persons for all fifty states and DC. These vintages are adjusted to reflect number of persons for comparison to current and future vintages in ALFRED.

b. Aggregating Series: The following series aggregate related GSP categories from ALFRED and create new GSP categories to reduce the number of variables in the FRED-SD database.

i. NATURNQGSP: Agriculture and Mining GSP

1. AGRNQGSP: Agriculture, Forestry, Fishing, and Hunting GSP
2. MINNQGSP: Mining GSP

ii. FIRENQGSP: Finance and Real Estate GSP

1. FININSNQGSP: Finance and Insurance GSP
2. RERENTLEANQGSP: Real Estate, Rental and Leasing GSP

iii. PSERVNQGSP: Other Private Industries GSP

1. WHOLENQGSP: Wholesale Trade GSP

2. RETAILNQGSP: Retail Trade GSP

3. TRANSWARENQGSP: Transportation and Warehousing GSP

4. PROBUSNQGSP: Professional, Scientific, and Technical Services GSP

5. MNGCOENTPRNQGSP: Management of Companies and Enterprises GSP

6. ADMINWASTNQGSP: Administrative and Waste Management Services GSP

7. EDCATNQGSP: Educational Services GSP

8. HLTHSOCASSNQGSP: Health Care and Social Assistance GSP

9. ARTENTRECNQGSP: Arts, Entertainment, and Recreation GSP

10. ACCOMDNQGSP: Accommodation and Food Services GSP

11. OTHSERVENQGSP: Other Services, Except Government GSP

For each aggregated series, all the subcomponent series were available for the same observation period with the same number of vintages, with the exception of one aggregated series:

i. DCNATURNQGSP: Agriculture and Mining GSP for District of Columbia

Mining and Agriculture data had slightly different observation and vintage dates. The combined series here represents only common observations. This resulted in dropping 4 observations from DCAGRNQGSP (Jan-2005 to Oct-2005) and one vintage from DCMINNQGSP (2016-08).[ii]

c. Aggregating Frequency

i. ICLAIMS: Initial Claims

Vintages are aggregated from weekly to monthly frequency. The monthly vintage reflects the average value of weeks observed in a month using the last weekly vintage in a given month.

d. Seasonal Adjustments: The following series were not available with seasonal adjustment (in ALFRED) but have been manually adjusted using the X12-ARIMA program from the IRIS Matlab toolbox for the FRED-SD database:

i. Applied to all 50 states and DC:

1. MFGHRS: Manufacturing Hours
2. STHPI: Home Price Index
3. ICLAIMS: Initial Claims

ii. Applied to select states:

1. SDCONS: All employees, Construction for South Dakota
2. TNCONS: All employees, Construction for Tennessee
3. WVINFO: All employees, Information for West Virginia
4. SDMINNG: All employees, Mining and Logging for South Dakota
5. TNMINNG: All employees, Mining and Logging for Tennessee
6. DCBPPRIVSA: New Private Housing Units for District of Columbia

A.3 III. Assigning Vintage Dates

a. Monthly Data

Generally, a new observation is added each month for a series, and the vintage date is forced to the last date of the respective month.

i. Delays in Data due to Release Schedule

The State Employment and Unemployment news release presents data from the Local Area Unemployment Statistics and State and Metro Area Employment, Hours, & Earnings programs.

The State Employment and Unemployment Report typically publish data on a one month lag; however, data for January is not released in February. As a result, January and February observations are both released in March.

The FRED-SD database preserves the real-time availability of these employment series: no new observation in February and two new observations in March. In some cases, the second data release was not captured in ALFRED, so a pseudo vintage was created to reflect the real-time availability of these series.

Link explaining releases: https://www.bls.gov/bls/news-release/laus.htm#2014

ii. Delays in Data due to Government Shutdown

The partial government shutdown from December 22, 2018 to January 25, 2019 delayed the release of some data series. FRED-SD reflects the real-time availability of these series. As a result, some vintages will not add a new observation or will increase by more than one observation. For example:

1. BPPRIVSA: New Private Housing Units
   a. 2018-12: zero new observations
   b. 2019-02: zero new observations
   c. 2019-03: three new observations

Census Release Schedule can be found here: https://www.census.gov/construction/bps/schedule.html

b. Quarterly Data

FRED-SD assigns quarterly series vintage dates to the middle month of the quarter, with the exception of the Personal Income series which is assigned to the last month of the quarter. By using this method, quarterly data in FRED-SD has consistent timing between observations but may reflect data released later than the real-time availability in some instances.

iii. For instance, most quarterly series in FRED-SD are released at the beginning of the second month in the quarter. If the 1st falls on a weekend, the data is sometimes released at end of the first month. However, FRED-SD would not recognize the observation until the end of the second month in the quarter.

A.4 IV. Missing Values

The following series have at least one missing observation in FRED/ALFRED

a. Filling in Missing Values

i. Using Data from Another Vintage

1. When data is missing, the first procedure is to fill in the observation using the previous available vintage.

2. When a previous vintage is not available (i.e. a missing observation occurs in the first vintage), the second procedure is to backfill the observation from the next available vintage.

ii. Using Data from Original Source

1. In some rare cases, an observation was not recorded in FRED for any vintage, in which case, the third procedure is to fill in the observation using the most recent value from the original data source. For example:

a. DCBPPRIVSA: New Private Housing Units for District of Columbia

Observation for Sep-2006 was not available via FRED. Most recent Census data used to fill in missing value across all available vintages.

Data sourced from: https://www.census.gov/construction/bps/txt/tb2u200609.txt.

b. Missing Values Due to Series Discontinuation

The following series are intentionally left as missing values from FRED-SD because the series was discontinued at some point in time:

i. MENATURNQGSP: Agriculture and Mining GSP for Maine

This series was temporarily discontinued in 2008, 2009, and 2012.

c. Omitting Values Due to Inconsistent Report

In some cases, either vintages or early observations of a series were omitted from the FRED-SD database because of inconsistent reporting between ALRED vintages or the original source. For example:

i. IMPTOT: Import of Goods

1. Although the first vintage in FRED is 2019-03, vintages begin at 2019-04 in FRED-SD.

ii. CTICLAIMS: Initial Claims for Connecticut

1. Although the first observation in FRED is 10/1/1985, the series begins in 2/1/1986 in FRED-SD due to missing weekly observations.

A.5 V. Accessing Data

a. Downloading FRED-SD Database

Data can be downloaded from the following link: https://research.stlouisfed.org/data/owyang/fred-sd/.

b. How to Read the FRED-SD Database

The database is structured after FRED-MD, where

i. Each file name corresponds to the vintage date.

ii. Each tab in the file corresponds to a variable.

iii. Each column corresponds to a U.S. state or Washington D.C.

iv. Each row corresponds to an observation date

Endnotes:

i. The AGRNQGSP series does not appear directly in the FRED-SD database but is aggregated with MINNQGSP to create the NATURNQGSP series. In this case, WYNATURNQGSP represents only WYMINNQGSP.

ii. The AGRNQGSP series does not appear directly in the FRED-SD database but is aggregated with MINNQGSP to create the NATURNQGSP series. In this case, only the DCAGRNQGSP series was adjusted and later merged with DCMINNQGSP.