Bayesian Dependent Functional Mixture Estimation for Area and Time-Indexed Data: An Application for the Prediction of Monthly County Employment

Terrance D. Savitsky [1]   Matthew R. Williams [2]

[1] U.S. Bureau of Labor Statistics (Office of Survey Methods Research)

[2]RTI International (Division for Statistical and Data Sciences)

Seasonal Adjustment Practitioners Workshop
June 7-8, 2022

Outline

Motivation: LAUS Forecasts

Model: Four Major Components

Forecast Performance: Comparing Alternatives

Background

- Local Area Un/employment "survey" (LAUS) publishes by county
  - Employment and Unemployment totals
  - Monthly
  - For *every* county and Municipal Civil Division (MCD) in the U.S.
  - ... there is no survey.

Background (2)

- LAUS project forward census instrument
  - Quarterly Census of Employment and Wages (QCEW)
  - by 7 months
  - for each county time series, separately
  - Includes seasonality
- Simultaneously model collection of county time-series
  - To produce more accurate predictions.

LAUS Employment Estimation

- LAUS (Local Area Unemployment Survey) partners with States for county-level monthly employment
- CES (Current Employment Statistics) is unavailable for 1751 out of 3108 counties
- Partnering with QCEW (Quarterly Census of Employment and Wages) program to use lagged data and project forward 7 months
- Data set is $N = 3108 \times T = 180$,
- $i = 1, \ldots, (N = 3108)$ counties
- $j = 1, \ldots, (T = 180)$ months
  - Observe Jan 2002 - May 2016
  - Predict 7 months, June - December 2016
- Project monthly values, by county, for remainder of 2016.

Outline

County-indexed Time Series

- $y_{ij} \sim \mathcal{N}\left(f_{ij} = \text{pred}_{ij} + \text{tr}_{ij} + \text{seas}_{ij}, \tau_y^{-1}\right)$
- $\text{pred}_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta}_i$; $\boldsymbol{\beta}_i \sim \mathcal{N}_P\left(\boldsymbol{\mu}_i, \Lambda_i^{-1}\right)$
- $T \times 1$, $\text{tr}_i \sim f_{\nu_i}$, autoregressive, bw 1 $(\text{tr}_{i,j-1}, \text{tr}_{i,j+1})$.

$$\text{tr}_i \overset{\text{ind}}{\sim} {\nu_i}^{\frac{T-1}{2}} \exp\left(-\frac{\nu_i}{2}\sum_{j=1}^{T-1}\left(\text{tr}_{i(j+1)} - \text{tr}_{ij}\right)^2\right) \quad (1)$$

$$= {\nu_i}^{\frac{T-1}{2}} \exp\left(-\frac{\nu_i}{2}{\text{tr}_i}^T Q \text{tr}_i\right) \quad (2)$$

- Precision matrix, $Q = (D - \Omega)$
- $\text{tr}_{ij} \perp \text{tr}_{ik} \mid \text{tr}_{i,-jk} \leftrightarrow \Omega_{ij} = 0$
- Rank-deficient since mean level not identified
- Probabilistic local smoother

County-indexed Time Series

- $y_{ij} \sim \mathcal{N}\left(f_{ij} = \text{pred}_{ij} + \text{tr}_{ij} + \text{seas}_{ij}, \tau_y^{-1}\right)$
- $\text{pred}_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta}_i$; $\boldsymbol{\beta}_i \sim \mathcal{N}_P\left(\boldsymbol{\mu}_i, \Lambda_i^{-1}\right)$
- $T \times 1$, $\text{tr}_i \sim f_{\nu_i}$, autoregressive, bw 1 $\left(\text{tr}_{i,j-1}, \text{tr}_{i,j+1}\right)$.
- 2 options for $T \times 1$, seas_i:
  - $\text{seas}_i \sim g_{\phi_i}$, autoregressive, bw $(O = 12) - 1$ $\left(\text{seas}_{ij}, \ldots, \text{seas}_{i(j+(O-1))}\right)$
    - Improper, local, $\text{seas}_i = \mathcal{N}_T\left(\mathbf{0}, Q_i^{-1} = \left[\tau_i\left(D - \Omega\right)\right]^{-1}\right)$
    - Proper, global $\text{seas}_i = \mathcal{N}_T\left(\mathbf{0}, Q_i^{-1} = \left[\tau_i\left(D - \rho_i\Omega\right)\right]^{-1}\right)$
  - seas_{ij} = fourier basis = $\left[\overset{O-1\times 1}{\mathbf{z}_{ij}} = \left\{\cos\left(\frac{2\pi k_1 j}{O}\right), \sin\left(\frac{2\pi k_2 j}{O}\right)\right\}^{'}_{k_1=1,\ldots,O/2,\ k_2=1,\ldots,(O/2-1)}\right] \times \boldsymbol{\kappa}_i$
    - $\mathbf{x}_{ij} \leftarrow \left(\mathbf{x}_{ij}, \mathbf{z}_{ij}\right)$ and $\boldsymbol{\beta}_i \leftarrow \left(\boldsymbol{\beta}_i, \boldsymbol{\kappa}_i\right)$.

County-indexed Time Series

- $y_{ij} \sim \mathcal{N}\left(f_{ij} = \text{pred}_{ij} + \text{tr}_{ij} + \text{seas}_{ij}, \tau_y^{-1}\right)$
- $\text{pred}_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta}_i$; $\boldsymbol{\beta}_i \sim \mathcal{N}_P\left(\boldsymbol{\mu}_i, \Lambda_i^{-1}\right)$
- $T \times 1,\ \text{tr}_i \sim f_{\nu_i}$, autoregressive, bw 1 $(\text{tr}_{i,j-1}, \text{tr}_{i,j+1})$.
- 2 options for $T \times 1,\ \text{seas}_i$:
  - $\text{seas}_i \sim g_{\phi_i}$, autoregressive, bw $(O = 12) - 1$ $\left(\text{seas}_{ij}, \ldots, \text{seas}_{i(j+(O-1))}\right)$
  - seas_{ij} = fourier basis = $\overset{O-1\times 1}{\mathbf{z}_{ij}} \times \boldsymbol{\kappa}_i$
    - $\mathbf{x}_{ij} \leftarrow (\mathbf{x}_{ij}, \mathbf{z}_{ij})$ and $\boldsymbol{\beta}_i \leftarrow (\boldsymbol{\beta}_i, \boldsymbol{\kappa}_i)$.
- Probabilistic Clustering:
  - Collect, $\boldsymbol{\theta}_i = (\nu_i, \phi_i, \boldsymbol{\mu}_i, \Lambda_i)$
  - Unique cluster parameter values, $\boldsymbol{\theta}_k^*,\ k = 1, \ldots, K \leq n$
  - If counties $i, \ell \in$ cluster $k \rightarrow \boldsymbol{\theta}_i = \boldsymbol{\theta}_\ell = \boldsymbol{\theta}_k^*$

Predictors Used for Clustering

- location quotient $\in [0, 1]$,employment concentration of economic sector in county compared to national average.
- Sectors constructed from the first $2-$ digits of detailed NAICS industry code
- Sectors: Construction, Transportation, Services, Leisure, Public, Mining, Manufacturing, Information, Education.
- Assertion: location quotient more useful than spatial contiguity.
  - e.g., Rural county adjacent to urban county
  - Distinct economic drivers / bases
- Other predictors:
  - Unemployment insurance (UI) claims in each month for each county to measure economic health.
  - Latitude and Longitude, computed based on population (rather than geographic) centroids

Outline

Compare Seasonality Methods: Less Expressed

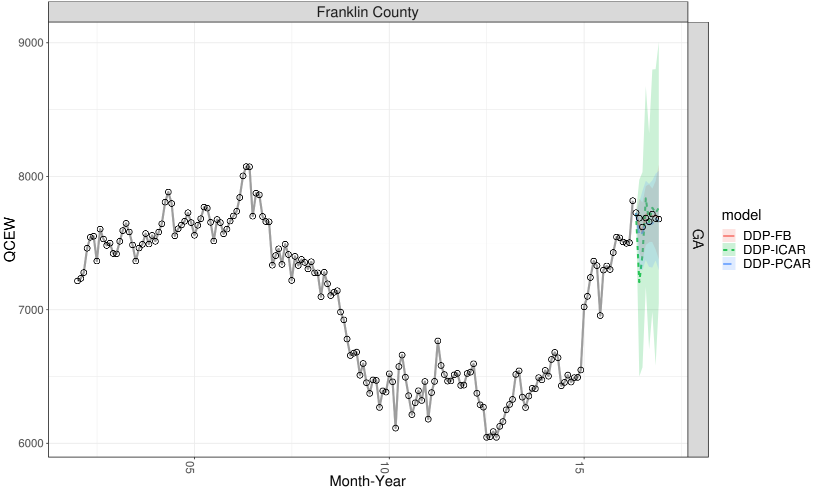

Figure: Fourier Basis (pink). Proper AR (blue), Local AR (green).

Compare Seasonality Methods: More Expressed

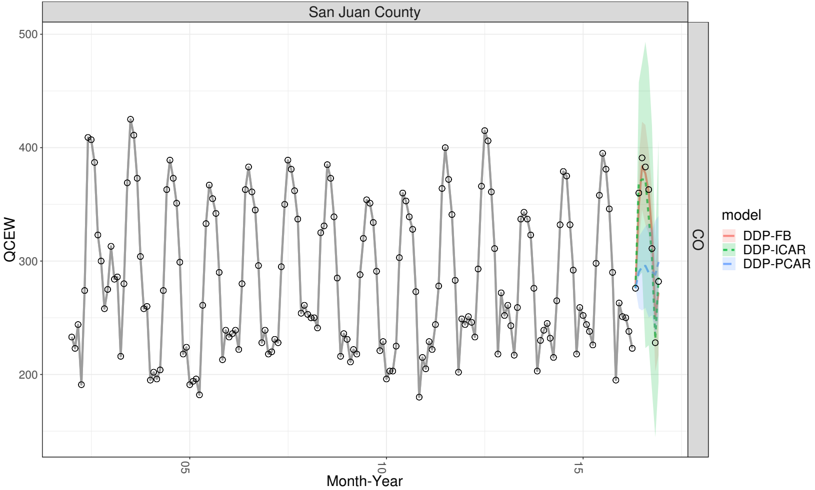

Figure: Fourier Basis (pink). Proper AR (blue), Local AR (green).

Smaller County

- Little seasonality expressed

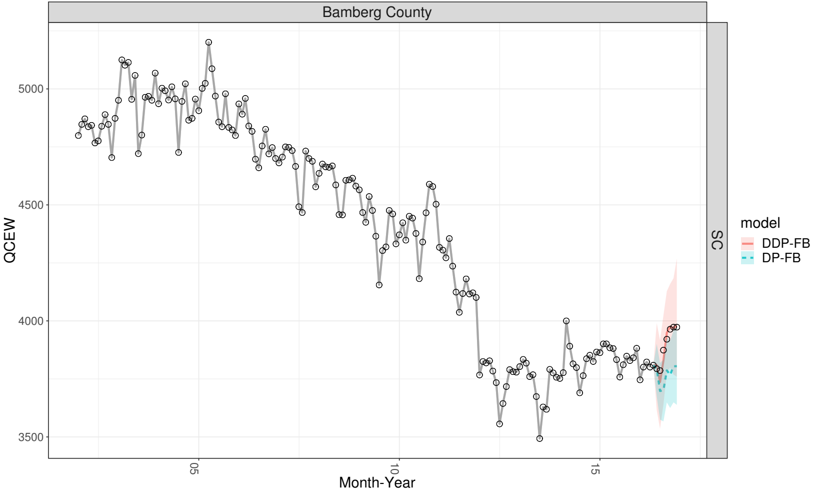

Figure: Predictor Assist (pink). Unsupervised (turquoise).

Medium-sized County

- Higher, but irregular seasonality expressed

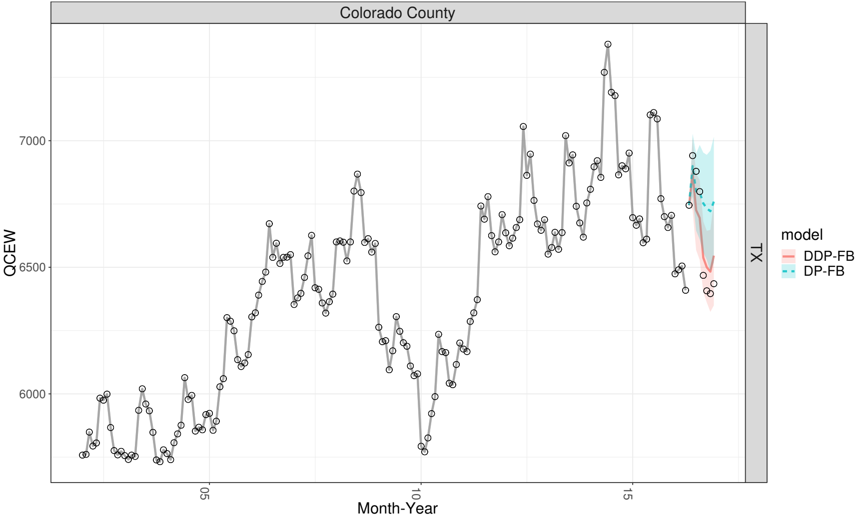

Figure: Predictor Assist (pink). Unsupervised (turquoise).

Tiny County

- Fibrilation

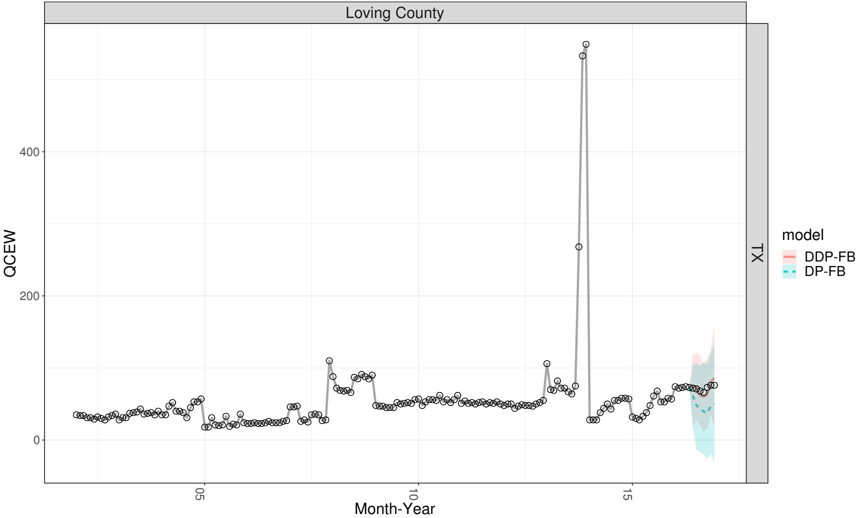

Figure: Predictor Assist (pink). Unsupervised (turquoise).

Spatial Process vs. Time-series

- Higher, but irregular seasonality expressed

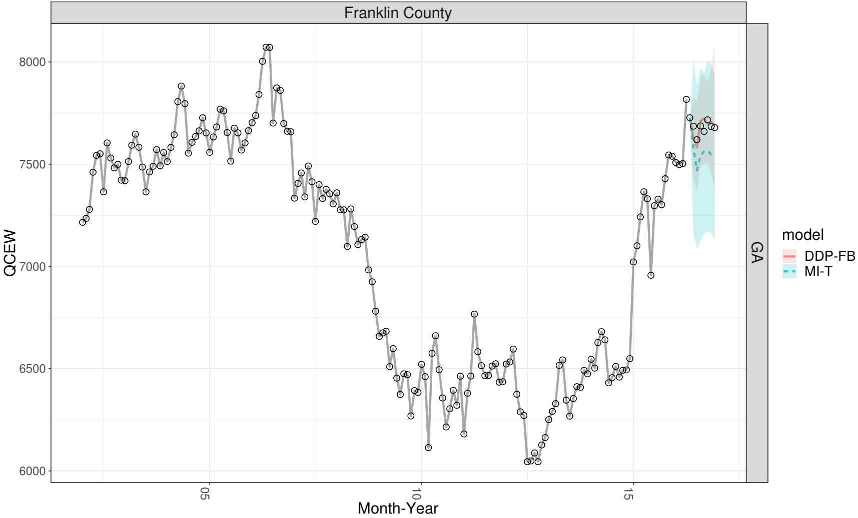

Figure: Predictor Assist (pink). Spatial process (turquoise).

Compare Prediction Errors of Models

Model	RMSPE	MAPE-C
Predictor Ast. Fourier (DDP - FB)	919	1.29%
Unsupervised Fourier (DP - FB)	1570	2.11%
Predictor Ast. Global (DDP - PCAR)	1688	2.45%
Predictor Ast. Local (DDP - ICAR)	2103	2.71%
Spatial Model (MI-t)	2987	3.37%
LAUS Production (SAEE)	–	2.49%

Comments:

- The models differentiated on seasonality
- DDP-FB performs best
- SAEE is the current production model

Summary

Bayesian Analysis, Advance Publication 1-25 2021.
https://doi.org/10.1214/21-BA1274

- Heterogeneity between counties for seasonal structures is a challenge
- The Fourier Basis shows marked improvement over Autoregressive Smoothers
- The Predictor Assisted clustering (DDP) shows marked improvement over unsupervised clustering (DP)
- Co-modelling time series leads to better prediction vs. modelling time series separately
- Clustering based on similar economic indices improves performance.
- Modelling a spatially-varying time series was much more effective than modelling a time-varying spatial process

Thanks to:

- Garret Schmitt, Tyler Bohnsack, Nic Aakre, Andrew Bean, Walter Sylva

CONTACT INFORMATION

Savitsky.Terrance@bls.gov
mrwilliams@rti.org

Bayesian Analysis, Advance Publication 1-25 2021.
https://doi.org/10.1214/21-BA1274