Bayesian Analysis of the Effects of Colorado's Salary Transparency Law on Hiring Trends

Will Johnson (hmd9tv) and Kaia Lindberg (pkx2ec)
DS 6040 - Bayesian Inference

Introduction

As states and cities in the United States enact or consider enacting salary transparency laws, it's possible that such laws have an impact on hiring rates for the local job market, for better or worse. Salary transparency laws require employers to disclose salary ranges for new positions in the job posting, information that is normally withheld from the jobseeker. One such state on the cutting edge of requiring employers to divulge salary ranges is Colorado. Effective January 2021, Colorado's Equal Pay for Equal Work Act "requires employers to include compensation to potential candidates, notify employees of promotional opportunities, and keep job description and wage rate records." [1]. Starting in November 2022, New York City is also expected to implement a similar law requiring employers to include salary ranges for every job posting [2]. Several other states don't require salary ranges to be listed on job postings, but do require employers to provide salary ranges during the hiring process to prospective employees and/or to disclose pay scales to current employees [3]. With such laws making pay information available to the public, these states are seeking to help address pay equity in reducing gender and race wage gaps.

With these new and forthcoming salary transparency laws, and in light of the current employment environment deemed the Great Resignation, the impact of such laws is a pertinent question to explore. In this paper we seek to explore how salary transparency impacts hiring rates for all non-agricultural positions. To answer this question we will examine the effect of Colorado's Equal Pay for Equal Work Act on hiring patterns. The analysis will use a Bayesian approach to explore how seasonality, short/medium term fluctuations, and long term trends help explain hiring trends in Colorado prior to the law in January 2021. Then, when applying this model to predict hiring trends in Colorado post January 2021 we'll seek to understand whether the observed hiring trends can still be explained by the model or whether there seems to be a meaningful change due to Colorado's Equal Pay for Equal Work Act.

Methodology

Data Preparation

The data for this analysis come from the US Bureau of Labor Statistics. In particular, the Job Openings and Labor Turnover Survey [4] offers historical data including job opening, hires, and separation by month, state, industry, etc. From this source Colorado hires data is obtained in order to evaluate the effect of the law. In particular, this analysis used data from the BLS with the following search conditions: Industry: total nonfarm, State: Colorado, Areas: all, Data Elements: hires, Size Classes: all, Rate and/or Level: level (in thousands), Seasonal Adjustment: not seasonally adjusted. Similar search criteria is used for Washington in order to provide a comparison to a similar state to use as a control (non-treatment) group. This data includes monthly hire counts (in thousands) from January 2021 to May 2022.

Prior to modeling, some data preparation was performed in order to normalize the hire counts and to convert the times to a numeric representation (as opposed to datetime type). The datetimes were converted to numbers by calculating the difference between each date and the reference time (first date in the dataset) and then dividing by 365 in order to represent the time (in years) since/relative to the initial time in the dataset. The hires data is normalized by calculating the difference between each month's hire count and the initial value and dividing by the standard deviation of the hire counts. Finally, the data was split to include data pre January 2021 in model training and leaving the post data (January 2021 to May 2022) for holdout analysis of the effect of the law.

Model

For our modeling method, we chose to use Gaussian Processes to estimate what would have happened without the law in place. Gaussian Processes are a modeling technique where the goal, instead of estimating posterior distributions of parameters, is to estimate a function. Since there are potentially infinite possible functions that could match a given dataset, Gaussian Processes work by providing probabilities to each one and using the means of those probabilities to estimate the most appropriate function possible. This way we can predict the most likely outcomes given our data, and it also means we can give estimates of our uncertainty for these means as well.

These possible functions are modeled by…

$$f \sim GP(\mu, k)$$

Where our function (f) is estimated by a Gaussian process with mean function $\mu(x)$ and with covariance function $k(x,x')$. In order to understand Gaussian Processes, we first note two important characteristics of a multivariate Gaussian distribution.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim N(\mu, \Sigma)$$

A set of random variables can be described as Gaussian distributed with mean of vector $[\mu]$ and covariance matrix $[\mathbf{\Sigma}]$ where each value in $[\mu]$ is the mean of each dimension and covariance matrix $[\mathbf{\Sigma}]$ describes the shape of the distribution around those means. What's handy about this kind of distribution is that we can both marginalize one of these dimensions and condition a dimension with a given input of the others and the probability distributions of both cases will also be Gaussian (notes on marginalizing and conditioning in **Appendix A**).

In Gaussian Processes, our possible functions for f(x) have probabilities we can model using this distribution. We aim to model the true mean vector $[\mu]$ and covariance matrix $[\mathbf{\Sigma}]$ with our training data, which together have a multivariate Gaussian distribution. We start by choosing a kernel function k(x,x') that will help us estimate the covariance matrix. This covariance matrix describes the shape of our distribution and how one variable influences the other, so this is where we will spend most of our time tuning our model. Choosing one kernel to populate a covariance matrix allows us to randomly sample a set of possible functions which will eventually center around a mean value and have a Gaussian distribution.

At this point, the range of possible functions across our prediction range is quite wide, so we introduce training data as a means of conditioning our prior distributions on observed data. This allows us to find the probability of a function f(x) given known outputs y, or P[f(x)|f(y)]. The observed data constrain the possible functions by forcing all functions to pass through each observed point. Combining our kernel function with observed data points and updating our prior beliefs about what functions should be possible is how we arrive at our posterior distribution.

> A big benefit that kernels provide is that they can be combined together, resulting in a more specialized kernel. The decision which kernel to use is highly dependent on prior knowledge about the data, e.g. if certain characteristics are expected.[5]

In our Gaussian Process, we will use our own industry knowledge about the annual hiring lifecycle and macroeconomic trends to inform our kernel choice and use those as a means to create informed priors. The resulting kernels added together will allow us to model specific behaviors in hiring we expect to see both in our training set and test set. Our choices of model and priors are based on this paper [6], but we've made several adjustments to the priors to reflect our knowledge of hiring practices as applicable to our data and problem of interest.

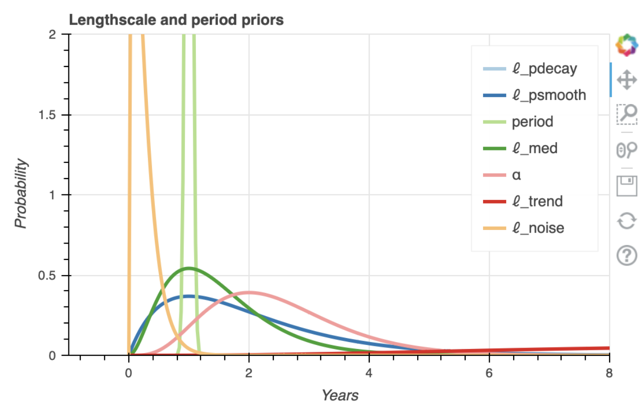
Priors

$$f(t) \sim GP_{long}(0,\ k_1(t,t')) + GP_{med}(0,\ k_2(t,t')) + GP_{per}(0,\ k_3(t,t')) + GP_{noise}(0,\ k_4(t,t'))$$

Each of the kernels above represent some prior-held belief about the nature of hiring trends along with a noise kernel to simulate error in our prediction values. They can be organized by prior this way:

1. A long term trend modeled by an exponential quadratic kernel.
    - This kernel represents our prior understanding of the macroeconomic business cycle which typically lasts between 7 to 12 years. This kernel will help us model overall economic growth during our 10 years of data.
2. A medium term irregularities modeled by a rational quadratic kernel.
    - This kernel will give our model some flexibility to react to shorter term growth or retraction rates as localized hiring markets tend to grow at different rates from year to year.
3. A short term period trend modeled by a periodic stationary kernel.
    - This will be our seasonality prior based on industry experience. This is an annual cycle that we will discuss in more detail in subsequent sections.
4. An error term modeled by the sum of a Matern32 kernel and a WhiteNoise kernel.
    - This term will model the error term that exists in all regression models.

Figure 1:
Lengthscale and Periodic Hyperparameter Prior Distributions. As an example, we can look at the "Period" parameter used in the seasonality adjustment kernel. This parameter decides how frequently the seasonality should occur, and based on industry experience we've decided that a strong prior on 1 year was appropriate as hiring tends to rise in the first half of the year, plateau in summer, then drop in the fall for the holidays.



Algorithm

A Markov Chain Monte Caro approach is used to obtain samples to estimate the posterior distribution using pymc3. The default sampling algorithm in pymc3, the No U-Turn Sampler (NUTS) is a variant of Hamiltonian Monte Carlo that automatically selects the tunable parameters of HMC. These methods use gradient information to converge faster than other sampling methods and generally work well to sample complex posterior distributions.

Application

The intention we have is to train a multi-kernel GP model on Colorado hiring data prior to the implementation of the new salary transparency law. The predictions will be made on data after this law went into effect, which we will call our "intervention period". Similar to other time-series based methods used in industry, we will attempt to measure the difference between our predicted values for Colorado hires and the actual values observed to estimate the impact this law had on the number of jobseekers getting employed in the state. We will also use a similar model on a separate state that did *not* have a similar law enacted over this time period as a way to adjust for macroeconomic trends not captured by our model.

Results

Posterior Distribution Plots

In **Figure 2** we've chosen to deep dive into just a few of our model's posteriors, see **Appendix B** to see the distribution of all of the components of our model.

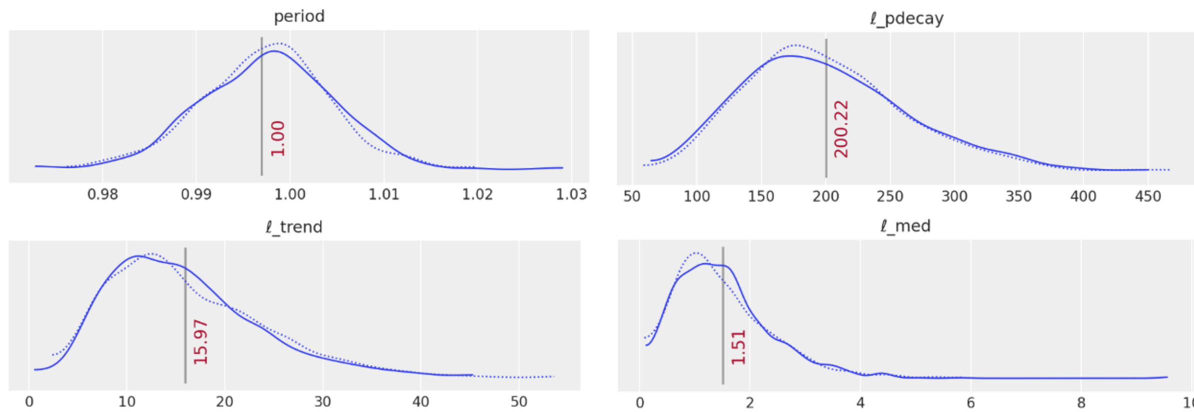


Figure 2:
Posterior distribution plots for period (period), periodic decay (ℓ_pdecay), long term trend lengthscale (ℓ_trend), and medium term variation lengthscale (ℓ_med).

The posterior distribution for period is centered closely around 1, reflecting our strongly informed choice of prior. The closeness of this estimation to 1 indicates that the time period for the seasonality component is 1 year, which aligns with our expectation and prior knowledge that the typical hiring cycle is annual. Period decay's posterior distribution had the largest density from 100 to 400. A large value for this parameter, like we obtained here, suggests that the periodicity of the seasonal component is unlikely to decay in the near future. This result is consistent with our choice of priors for this parameter and our expectation that the seasonality of the hiring cycle is unlikely to disappear.

The next two posterior distributions focus on the lengthscale parameters, one for the long term trend and the other for medium term variations. For medium term fluctuations the posterior density is greatest from about 0.5 years to 3 years, suggesting that medium term fluctuations impact hiring trends for a period of several months to a couple of years. In the application of hiring changes we'd expect this range of lengthscales to account for medium term macroeconomic or social events, for example the impact of recessions and pandemic related impacts to the job market. The lengthscale for the long term trend has a mean of about 16 years with the greatest posterior density from 5 years to 30 years. The range of this lengthscale parameter is much larger than that of the short/medium term parameter, the function governing the long term trend will change less frequently. This time frame of the lengthscale for the long term trend aligns with the business cycle in which we'd expect larger macroeconomic events and cycles to continue to impact hiring trends.

Decomposition of Trends

Now that we have a model and understand the posterior distributions of each component, we can use it to predict hiring levels and see how well the model can reflect variation in hiring in our data. **Figure 3** shows the observed and fitted observations during our pre-intervention training. The total fit (solid red line) shows the total fit (MAP predictions) with a confidence band to reflect the uncertainty in these predictions. We see that this model fits the observed data fairly well as it picks up on both longer term trends as well as intra-year fluctuations. Other than a handful of outlying observations, most points fall close to the total fit line suggesting that the model does a fairly good job of explaining hiring trends during our training period.

Figure 3:
Plot of the observed data, total fit, and components of the Gaussian Process including long term trend, medium range variation, and seasonality.

Since we've modeled this data using a GP with three components, we can break down the total fit to better understand how the long term trends, medium term variation, and seasonal processes are contributing to the total fit. The long term trend, shown in blue, suggests that in the longer term (~10 to 15 year time frame) hiring in Colorado tends to increase, but there is some leveling off towards the end of this range. This long term cyclical trend reflects our expectation of the rise and fall of macroeconomic trends including a rise in hiring post the Great Recession and then more of a leveling off. The green line reflects short to medium range variation in hiring and is able to reflect hiring trends over the timeframe of a year or two. The seasonality trends, in orange, show a roughly sinusoidal pattern as we'd expect due to annual hiring cyclicality. The discussion below regarding **Figure 4** goes into further detail explaining this seasonality as it relates to our knowledge of the annual hiring lifecycle.

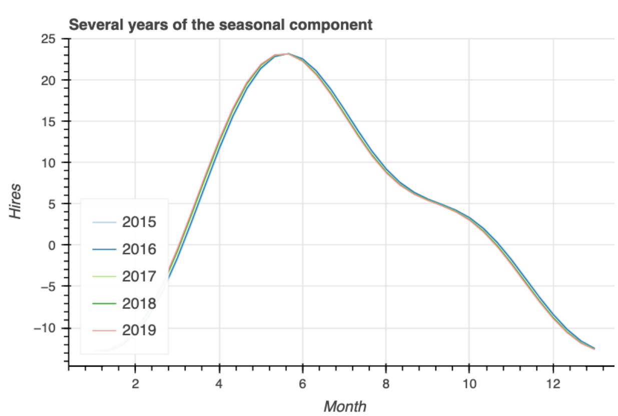


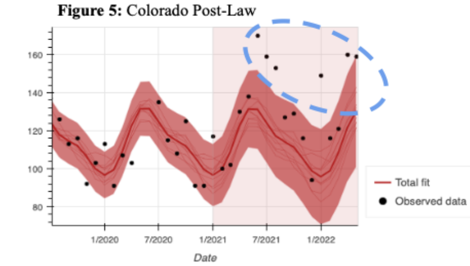
Figure 4:
Plot of the seasonal component over several years highlights the seasonality trends in hiring.

These curves reflect the annual hiring lifecycle with a 12 week delay caused by BLS reporting lags. The typical hiring wave happens early in the year with lots of new jobs and jobseekers looking to find work all at once. After this initial spike, it will typically plateau in the summer time before falling around Thanksgiving and the holidays. We'd expect a more closely fitted curve to have two marked decreases in hires made in November and December with a short spike in between the holidays as well, but our smoothing processes we applied to the seasonal GP is likely why we don't see that in these curves.

Post Intervention Observations

When comparing predicted hiring levels to the actual hiring levels after the transparency law was enacted, we immediately notice how our model underestimated the actual number of hires by significant levels. **Figure 5** shows Colorado's predicted range in red and actual monthly hiring levels in black. This might suggest the law had a positive impact on hiring in the state. To control for additional unknown macroeconomic forces, we consider another model with identical priors for Washington State, shown in **Figure 6**. In this chart, we see a similar underestimation of actual hiring trends, though more muted in comparison.

Figure 5: Colorado Post-Law



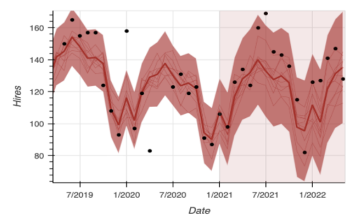


Figure 6: Washington's predicted hires over actual (control group)

To quantify these underestimations and compare our test to the control group directly, we observe a summation of error terms in **Table 1** below.

Table 1: Model Performance Metrics

	RMSE	**MAPE**	**Sum of Errors**	**∆ Net Errors Post Law**
Colorado	Train: 12.14 Test: 25.30	Train: 9.0% Test: 14.1%	Train: -11.4 Test: 302.7	**314.1**
Washington	Train: 12.6 Test: 16.7	Train: 8.7% Test: 10.9%	Train: 7.2 Test: 176.3	**169.1**

Note how the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) of our regression line increased more for Colorado than it did for our control group. The Sum of Errors also increased to resoundingly positive (meaning underestimations) for both states as well. Given the above, we observe a stronger aggregate increase in hiring activity over our expected values in Colorado than we observed in Washington.

Conclusion

The evidence we present in this paper do not prove causality between the enactment of the salary transparency law and increases in hiring rates definitively. There are still potential confounding variables uncaptured in this initial research, such as population migration trends or the impact of long term remote work options becoming more widely available.

However, we believe that the evidence found here suggests that more research is warranted towards the impact salary transparency has on local hiring economies. One hypothesis is that the availability of salary information in the job listing actually increases the speed at which roles get filled and jobseekers get hired. If true, the potential benefits of doing so could have long lasting effects on Colorado's economy. Potential follow-up research could be to look for similar increases in hiring activity after a similar law goes into effect in Washington State in 2023.

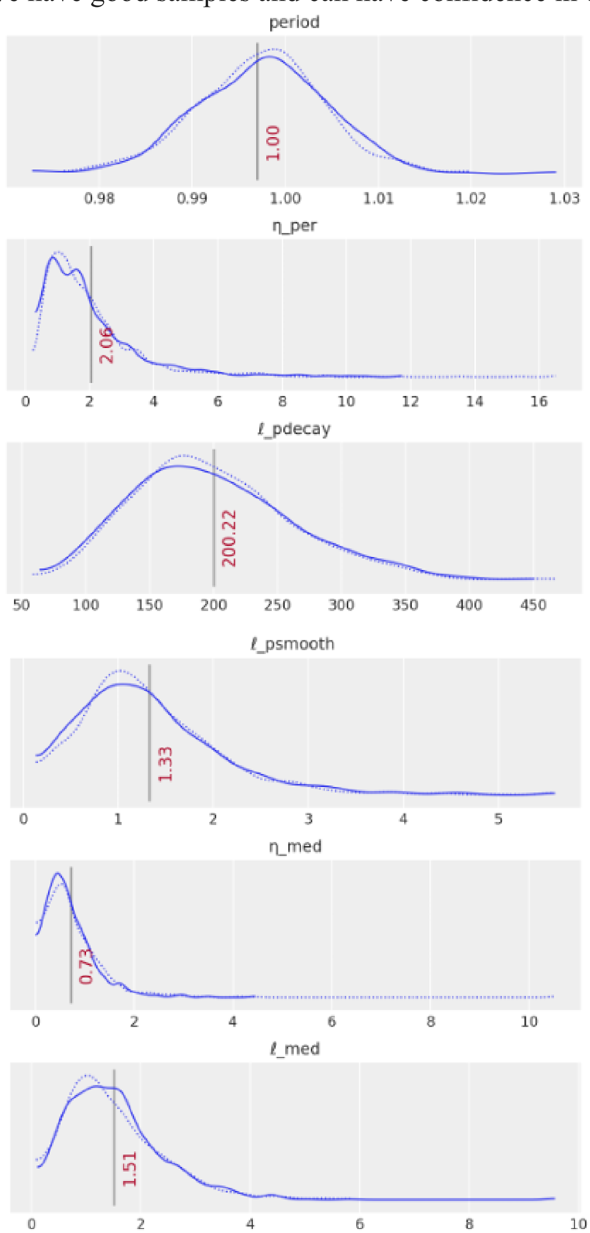
Appendix

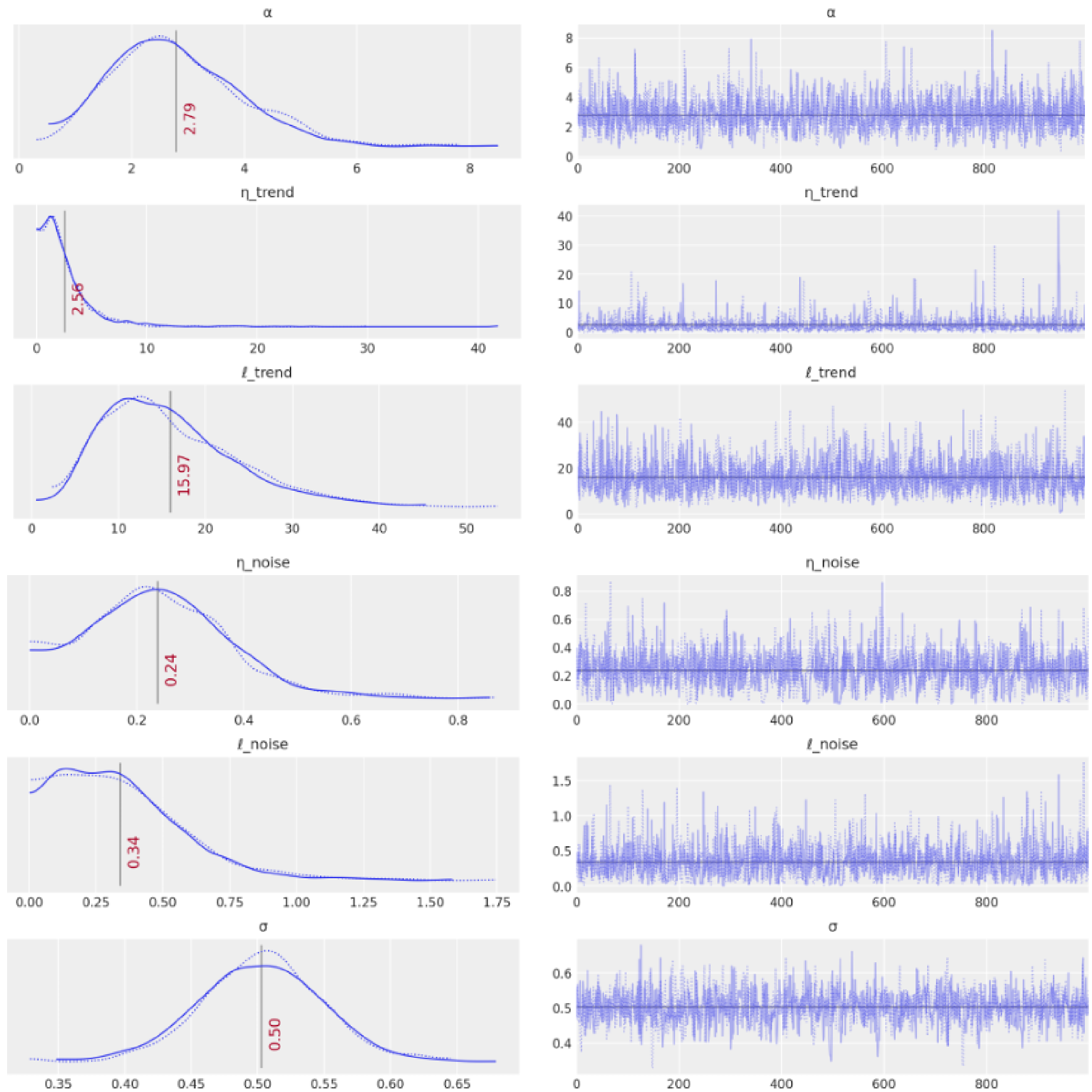
A - Marginalizing and Conditioning

If we marginalize X2, we essentially integrate X over all the other variables till we're left with the Gaussian distribution (normal curve of X2). If we want to find the distribution of X2 n given specific values of X1…n , the resulting distribution of will also be Gaussian and is essentially a "slice" of the multivariate distribution where the mean of X2 will be affected by the covariance of X2 and all the other variables.

B - Posterior Distribution Plots and Traceplots

The plots below show the posterior distributions and trace plots for each of the components of the GP for Colorado. From these plots it looks like our posterior distributions have converged. The trace plots look like hairy caterpillars and the posterior distributions look reasonable. Because these have converged we have good samples and can have confidence in our posterior estimates.





References

[1] https://cdle.colorado.gov/equalpaytransparency
[2] https://www1.nyc.gov/assets/cchr/downloads/pdf/publications/Salary-Transparency-Factsheet.pdf
[3] https://greyjournal.net/news/every-state-that-has-passed-pay-transparency-laws-so-far-2/
[4] https://www.bls.gov/jlt/#data
[5] https://distill.pub/2019/visual-exploration-gaussian-processes/#PriorFigure
[6] https://docs.pymc.io/en/v3/pymc-examples/examples/gaussian_processes/GP-MaunaLoa.html