Group Members: Gabi Huberman-Shlaes, Noah Schulhof, Faith Cramer

Technical Portion:

Paper #1: https://cran.r-project.org/web/packages/tscount/vignettes/tsglm.pdf

- The first paper discussed the tscount package in R for count times series data utilizing GLMs. It includes an in-depth discussion of the many specifications for the model.

Model: $$g(\lambda_t) = \beta_0 + \sum_{k=1}^{p} \beta_k \, \widetilde{g}(Y_{t-i_k}) + \sum_{\ell=1}^{q} \alpha_\ell g(\lambda_{t-j_\ell}) + \boldsymbol{\eta}^\top \boldsymbol{X}_t$$

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log p_t(y_t; \boldsymbol{\theta}) = \sum_{t=1}^{n} \Big(y_t \ln(\lambda_t(\boldsymbol{\theta})) - \lambda_t(\boldsymbol{\theta}) \Big), \qquad S_n(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \left(\frac{y_t}{\lambda_t(\boldsymbol{\theta})} - 1 \right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Paper #2: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1117187

- This paper discusses the special case of the GLM model from paper #1 when there are no covariates and when Y_t conditioned on past values is assumed to be Poisson distributed (we will explore this case further theoretically and for the applied portion).

Model (ACP): $$\lambda_t = \beta_0 + \sum_{k=1}^{p} \beta_k Y_{t-k} + \sum_{\ell=1}^{q} \alpha_\ell \lambda_{t-\ell}.$$

$$l_t(\theta) = N_t \ln(\mu_t) - \mu_t - \ln(N_t!) \qquad \frac{\partial l_t}{\partial \theta} = \left(\frac{N_t - \mu_t}{\mu_t} \right) \frac{\partial \mu_t}{\partial \theta}$$

where

$$\frac{\partial \mu_t}{\partial \theta} = z_t' + \sum_{s=1}^{p} \beta_s \frac{\partial \mu_{t-s}}{\partial \theta}, \tag{9.1}$$

and

$$z_t = [1, N_{t-1}, N_{t-2}, \ldots, N_{t-p}, \mu_{t-1}, \mu_{t-2}, \ldots, \mu_{t-q}]. \tag{9.2}$$

We will also consider various extensions of this model such as the double autoregressive conditional poisson and the generalized double autoregressive conditional poisson.

Objectives:

- We will walk through the derivations of the likelihood and score functions for the generalized linear model (paper #1).
    - We will then discuss their asymptotic behavior and inference in large samples.
- We will go over the derivations from the paper of some of the properties of the model, particularly the expected value, variance of the model, and derivation of log-likelihood and the score function for this specification.

Applied Portion:

- Dataset: https://www.aogr.com/web-exclusives/us-rig-count/2025
    - Monthly US Crude Oil and Natural Gas Rig Count data from 1973 to today.
- We will implement different extensions of the ACP time series model (ACP, DACP, GDACP) and compare their performances.
- We will also compare an ARIMA model as a baseline.
    - We will evaluate model performance by Rolling-window out-of-sample forecast.
- We will compare strengths/weaknesses of different modeling approaches.