SAS® GLOBAL FORUM 2019

USERS PROGRAM

APRIL 28 – MAY 1, 2019 | DALLAS, TX

Dr. Aric LaBarr

Aric is passionate about helping people solve challenges using their data. He is an Associate Professor of Analytics at the nation's first master of science in analytics degree program. There he helps design the innovative program to prepare a modern work force to wisely communicate and handle a data-driven future.

The Bayesians are Coming!
The Bayesians are Coming!
The Bayesians are Coming to Time Series!

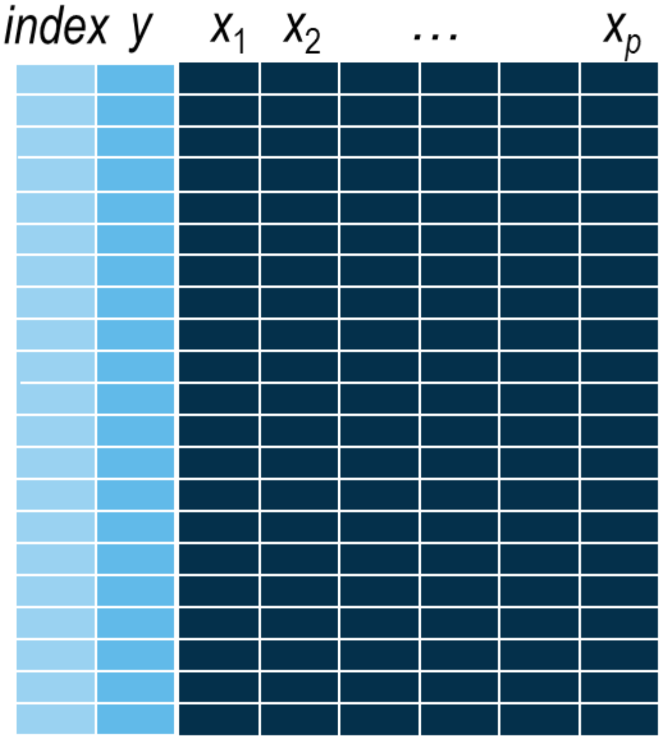
Dr. Aric LaBarr

Institute for Advanced Analytics at North Carolina State University

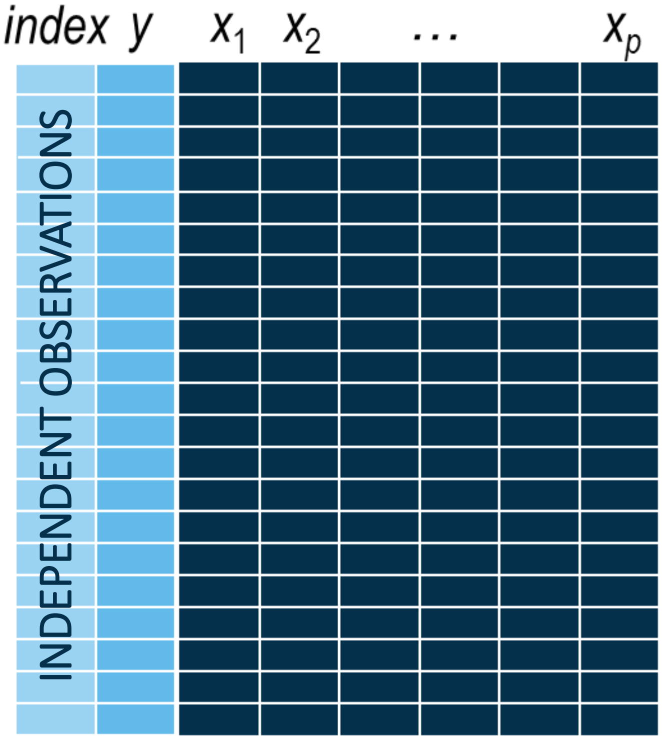
What is Time Series Again?

Asks nobody ever…

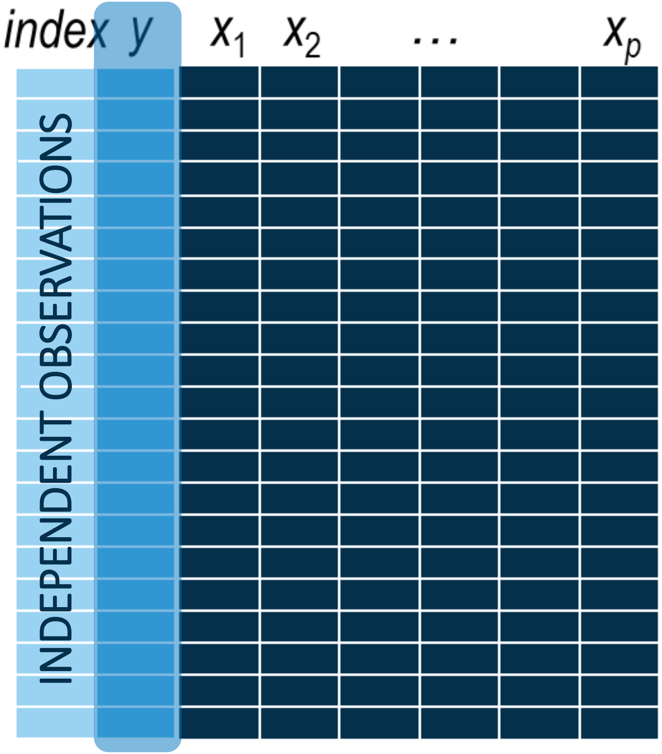
Cross-Sectional Data

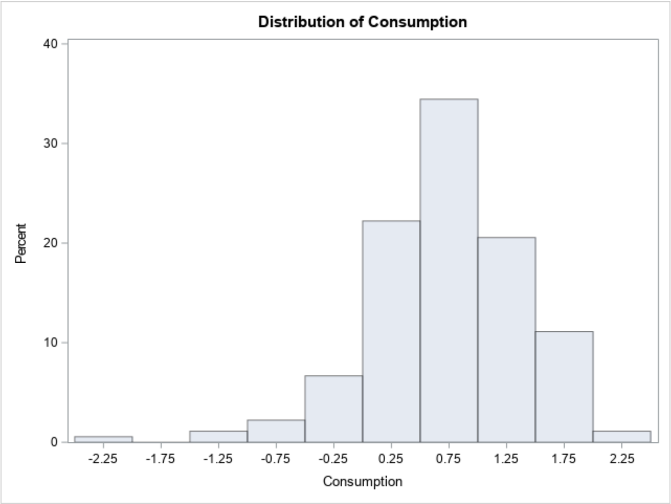


Cross-Sectional Data

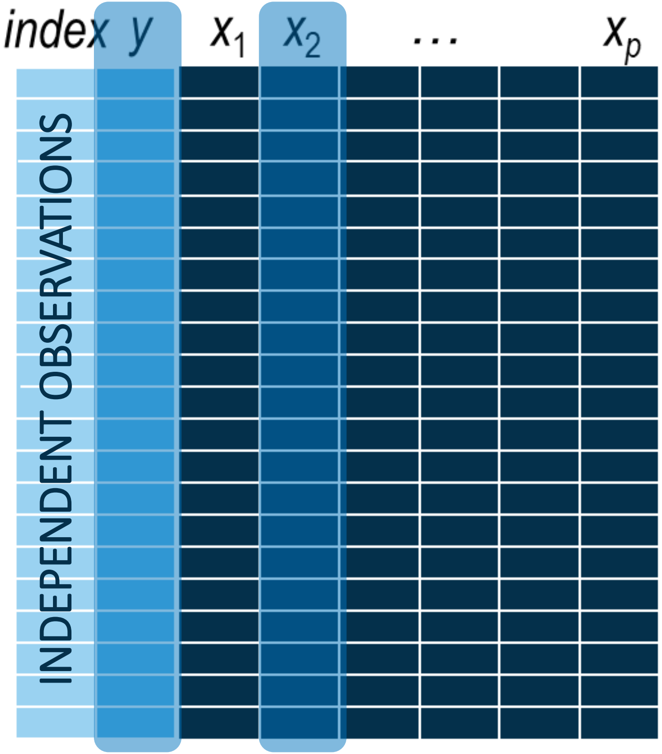


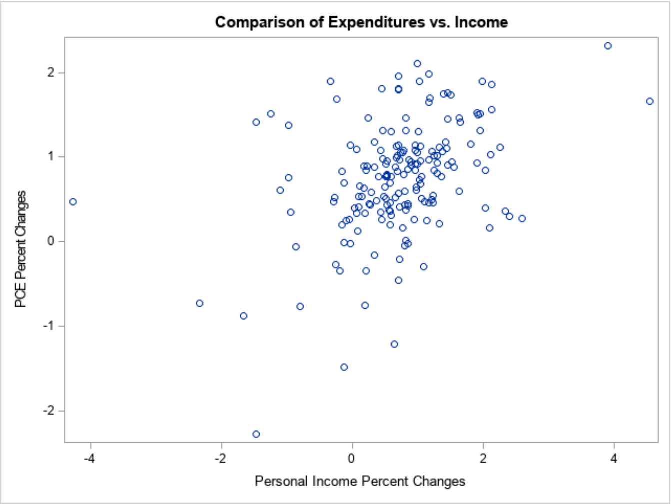
Cross-Sectional Data



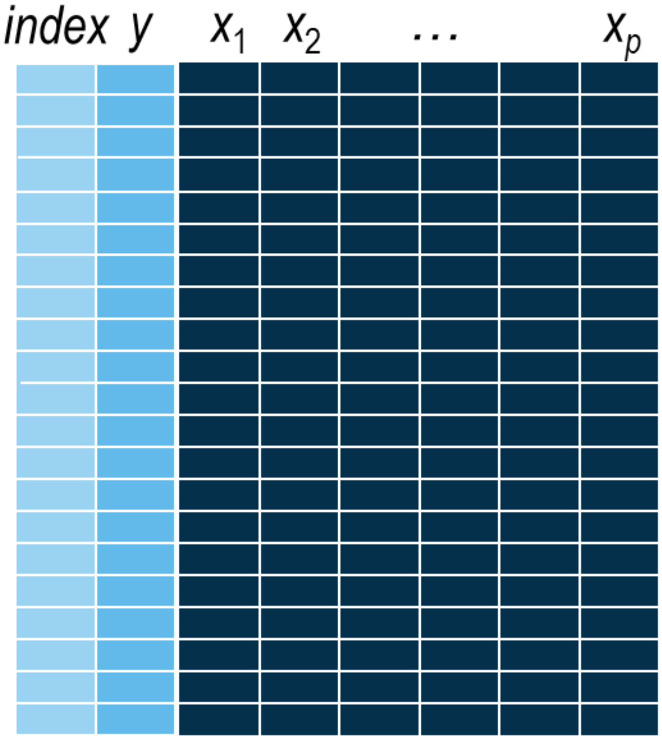


Cross-Sectional Data

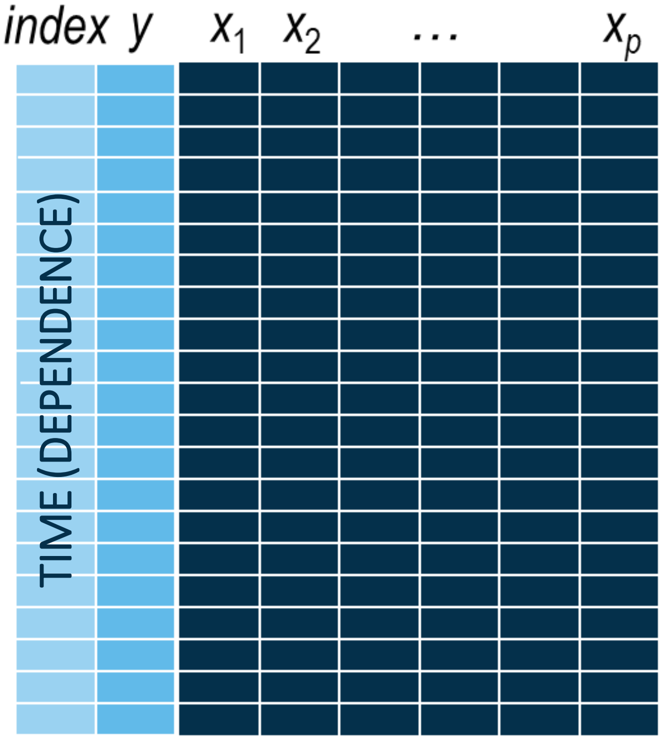




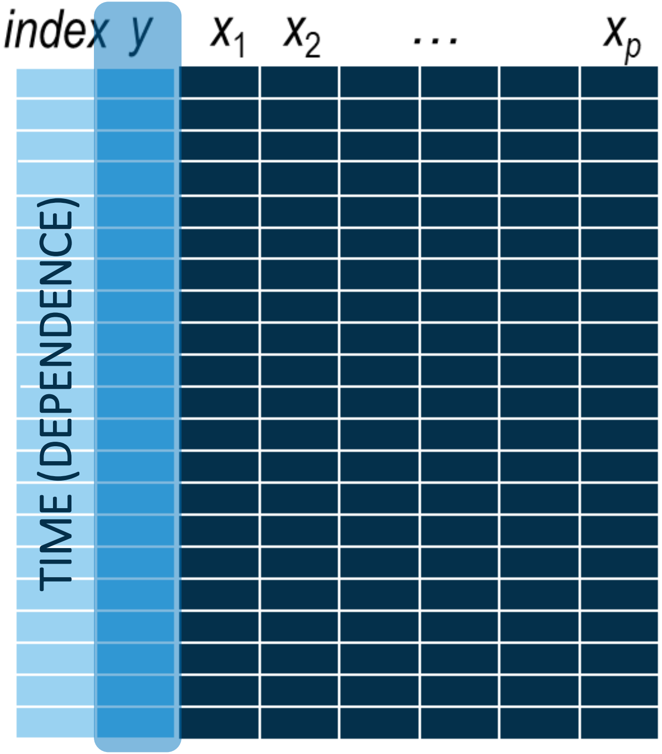
Time Series Data

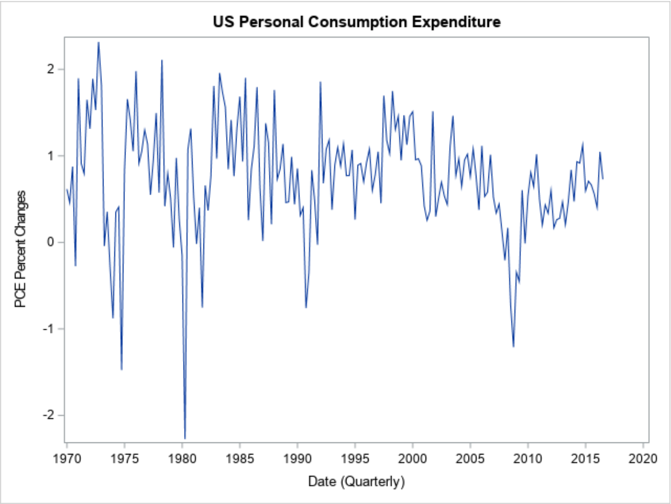


Time Series Data

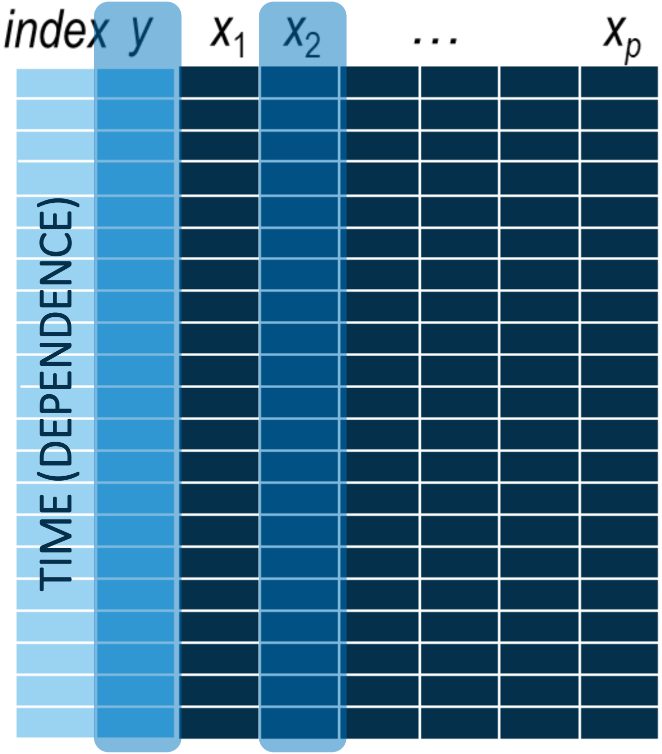


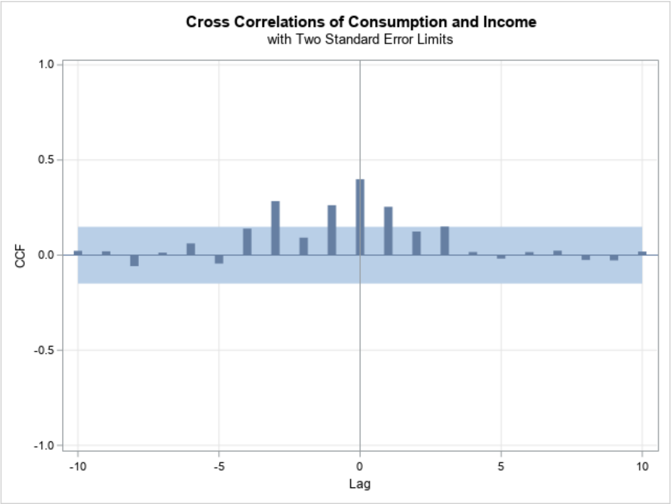
Time Series Data





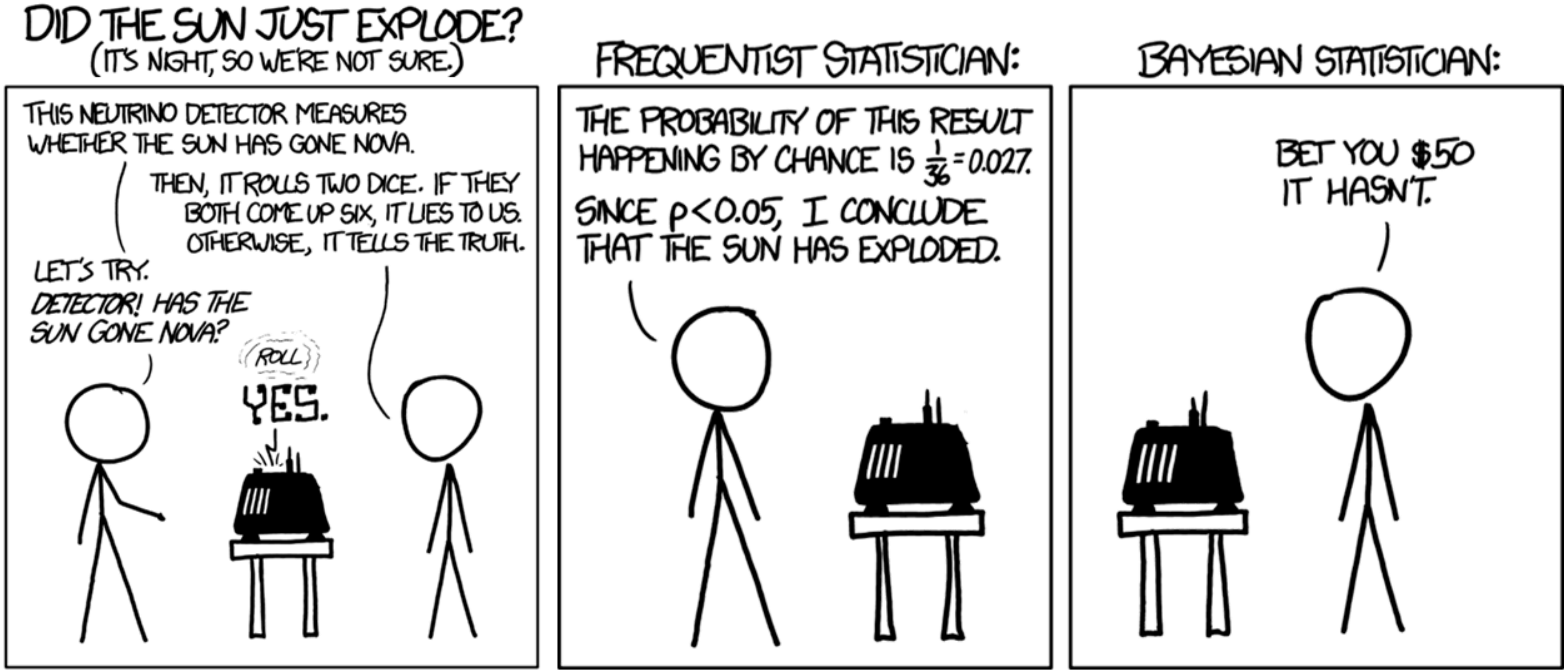
Time Series Data





Frequentist vs. Bayesian

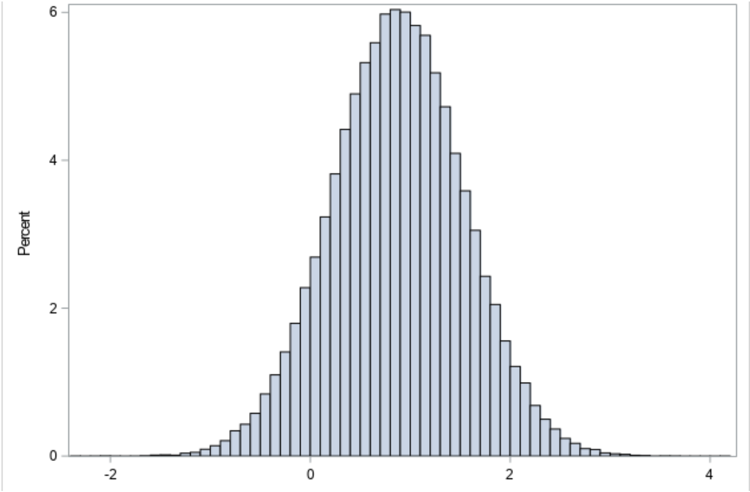
Quick Comparison from xkcd.com



Frequentist vs. Bayesian View

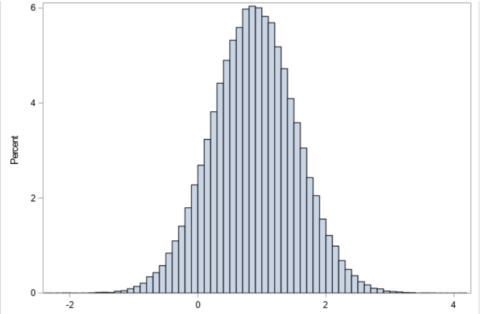
Frequentist view of μ

Bayesian view of μ



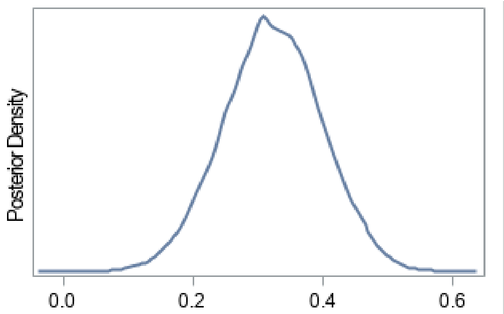
Priors & Posteriors

Prior Distribution

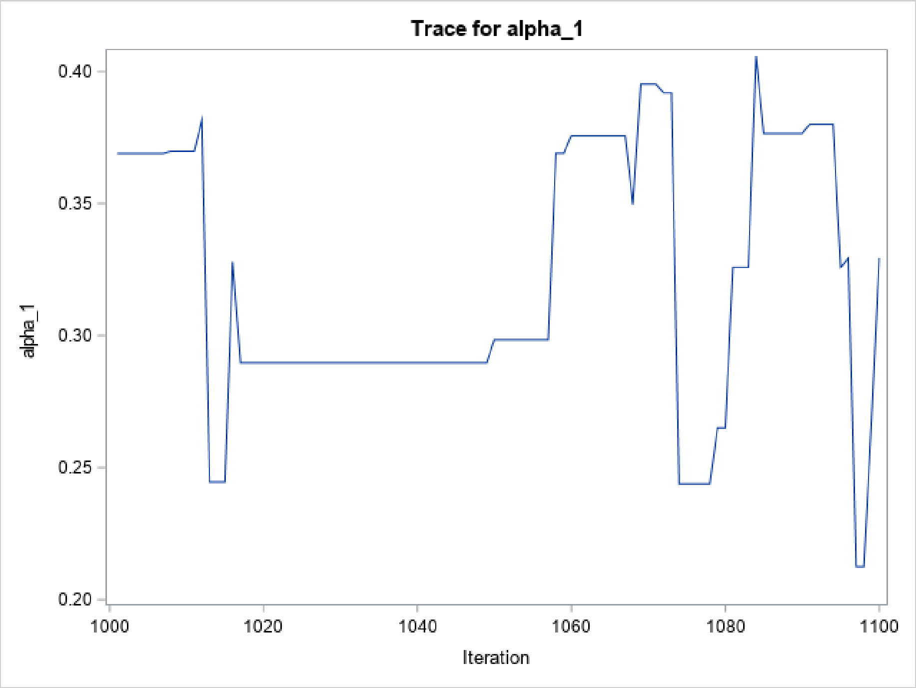


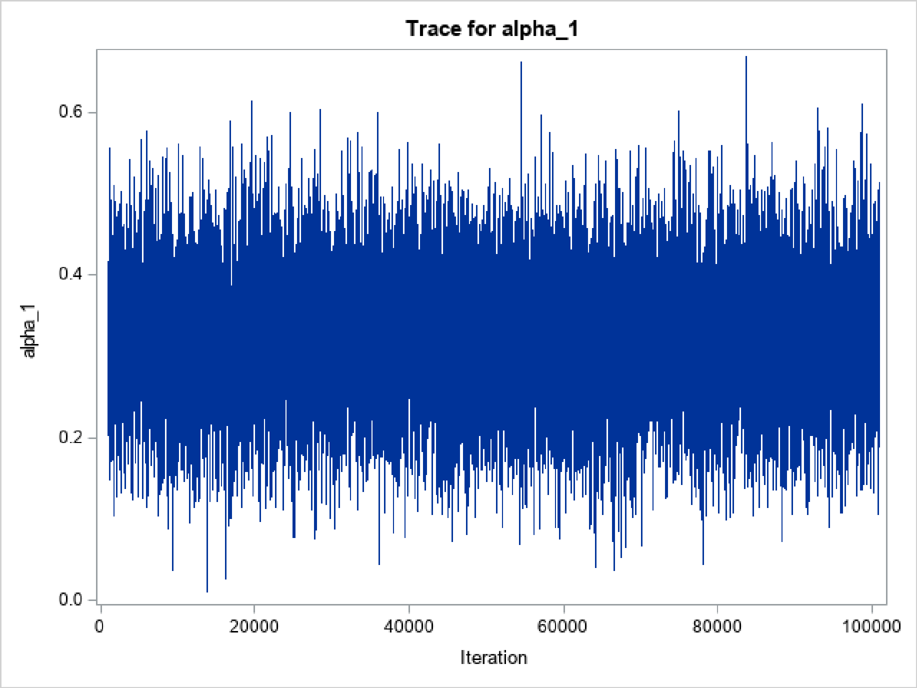
+ DATA =

Posterior Distribution



Markov Chain Monte Carlo (MCMC)





Classical Time Series Modeling

Statistical Forecasting

Time Series = Signal + Noise

Statistical Forecasting

Time Series = Signal + Noise

Forecasts extrapolate signal portion of model.

Confidence intervals account for uncertainty.

Time Dependencies

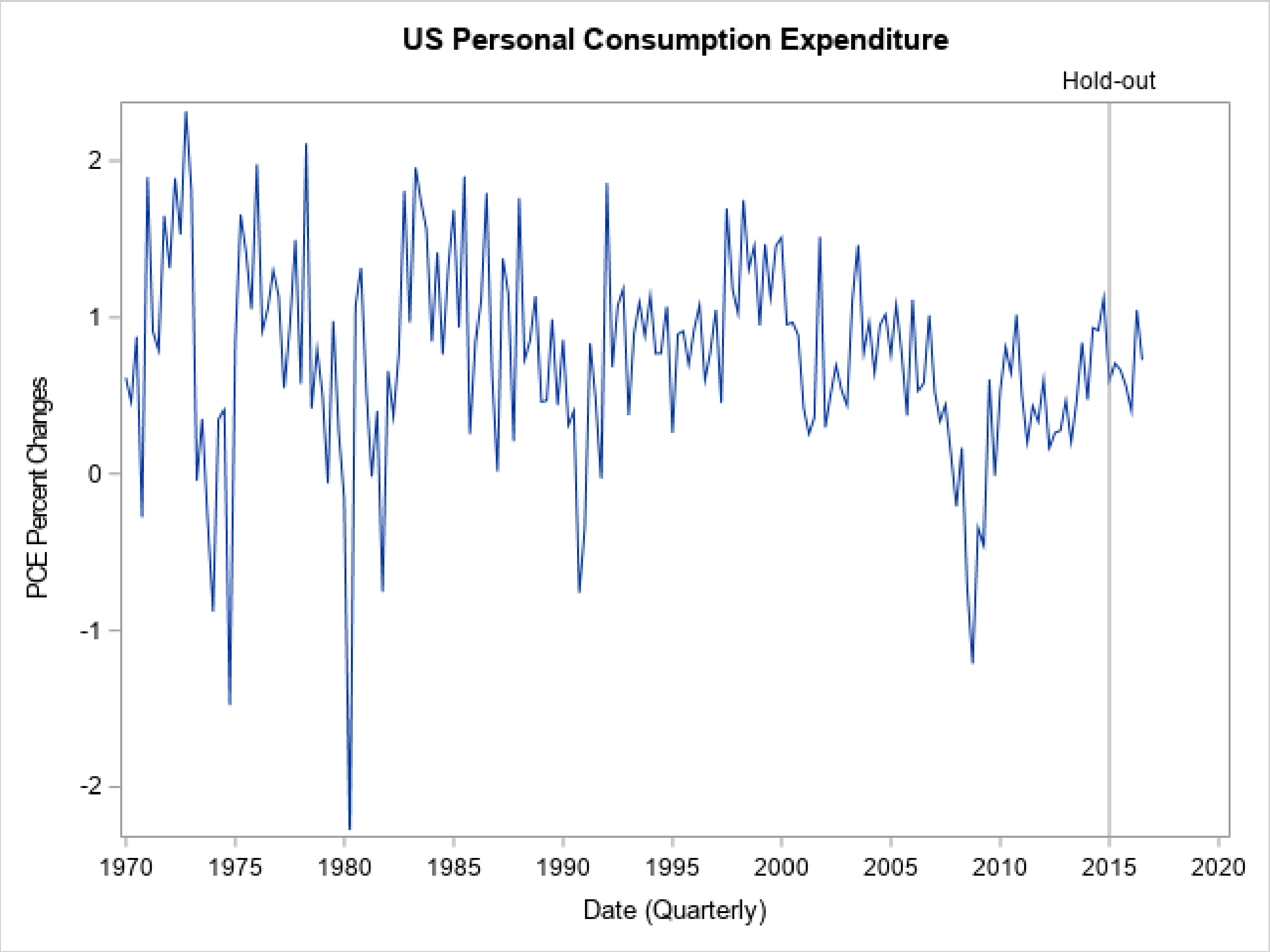
- Time series data relies on the assumption that the observations at a certain time point depend on previous observations in time.

Naïve Model:

$$\hat{Y}_{t+h} = Y_t$$

Average Model:

$$\hat{Y}_{t+h} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$



US Personal Consumption Expenditure
Hold-out
PCE Percent Changes
2
1
0
-1
-2
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Date (Quarterly)

Classical Time Series Modeling

Exponential Smoothing Model

Exponential Smoothing

- The weights should emphasize the most recent data.
- Forecasting should only require a few parameters.
- Forecast equations should be simple and easy to implement.

Single Exponential Smoothing

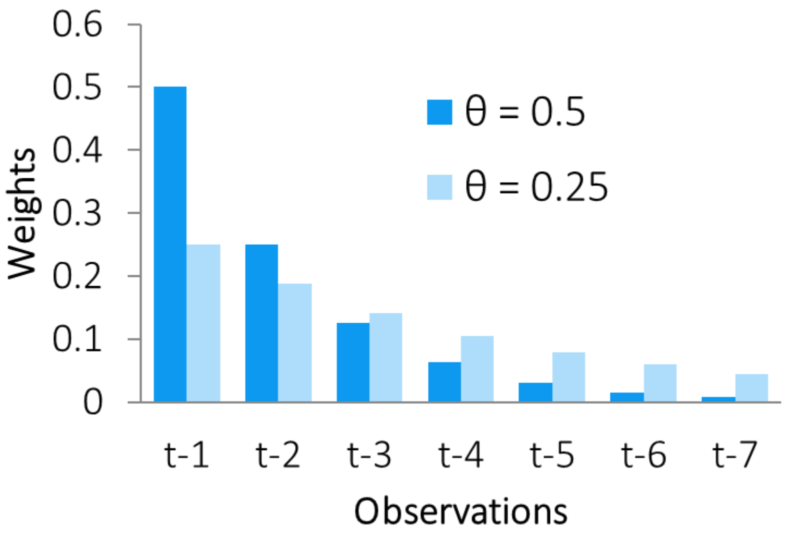
- We can apply a weighting scheme that decreases exponentially the further back in time we go.

$$\hat{Y}_{t+1} = \theta Y_t + \theta(1-\theta)Y_{t-1} + \theta(1-\theta)^2 Y_{t-2} + \theta(1-\theta)^3 Y_{t-3} + \theta(1-\theta)^4 Y_{t-4} + \cdots$$

$$0 \leq \theta \leq 1$$

Single Exponential Smoothing

- The larger the value of the θ, the more that the most recent observation is emphasized.



Single Exponential Smoothing

- The Single Exponential Smoothing model equates the predictions at time *t* equal to the weighted values of the previous time period along with the previous time period's prediction:

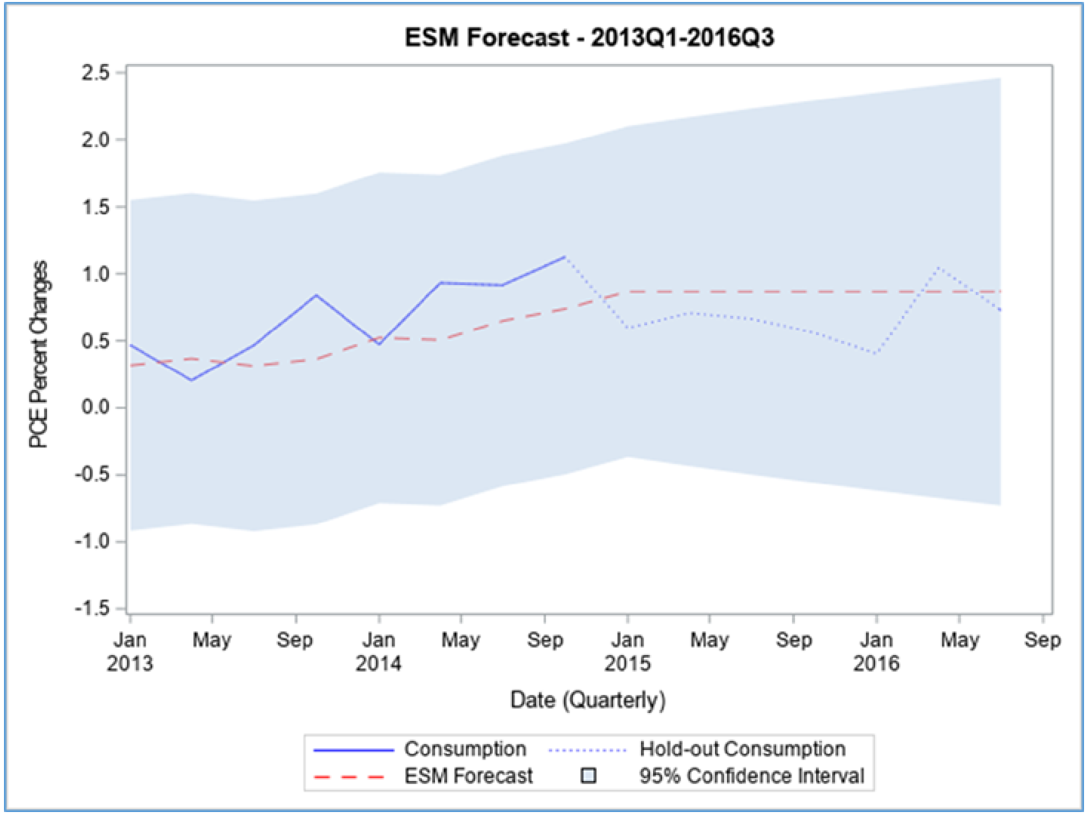
$$\hat{Y}_{t+1} = \theta Y_t + (1 - \theta)\hat{Y}_t$$

Parameter Estimation

- The typical method for calculating the optimal value of θ in the Exponential Smoothing model is through one-step ahead forecasts.
- The value of θ that minimizes the one-step ahead forecast errors is considered the optimal value.

$$SSE = \sum_{t=1}^{T} \left(Y_t - \hat{Y}_t\right)^2$$

Single Exponential Smoothing



Classical Time Series Modeling

ARIMA Model

Autoregressive (AR) Models

- Often you can forecast a series based solely on the past values.
- We are going to focus on the basic case – only lagged values of Y_t – called an AR(p) model:

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 Y_{t-1} + \hat{\alpha}_2 Y_{t-2} + \cdots + \hat{\alpha}_p Y_{t-p}$$

- Called long-memory models.

Moving Average (MA) Models

- You can also forecast a series based solely on the past *error* values.
- We are going to focus on the basic case – only lagged values of e_t, called an MA*(q)* model:

$$Y_t = \hat{\alpha}_0 + e_t - \hat{\beta}_1 e_{t-1} - \hat{\beta}_2 e_{t-2} - \cdots - \hat{\beta}_q e_{t-q}$$

- Called short-memory models.

ARIMA Models

- There is nothing to limit both an AR process and an MA process to be in the model simultaneously.
- These “mixed” models are typically used to help reduce the number of parameters needed for good estimation in the model.
- The “I” in ARIMA stands for **integrated.**
  - Differencing used for stationarity.

Automatic “Selection” Techniques

- SAS has three main automatic “selection” techniques for time series data – MINIC, SCAN, ESACF.

Automatic “Selection” Techniques

- SAS has three main automatic “selection” techniques for time series data – MINIC, SCAN, ESACF.

AR

		0	1	2	3	4
	0					
	1					
MA	2					
	3					
	4					

MINIC

AR

		0	1	2	3	4
	0					
	1					
MA	2					
	3					
	4					

SCAN

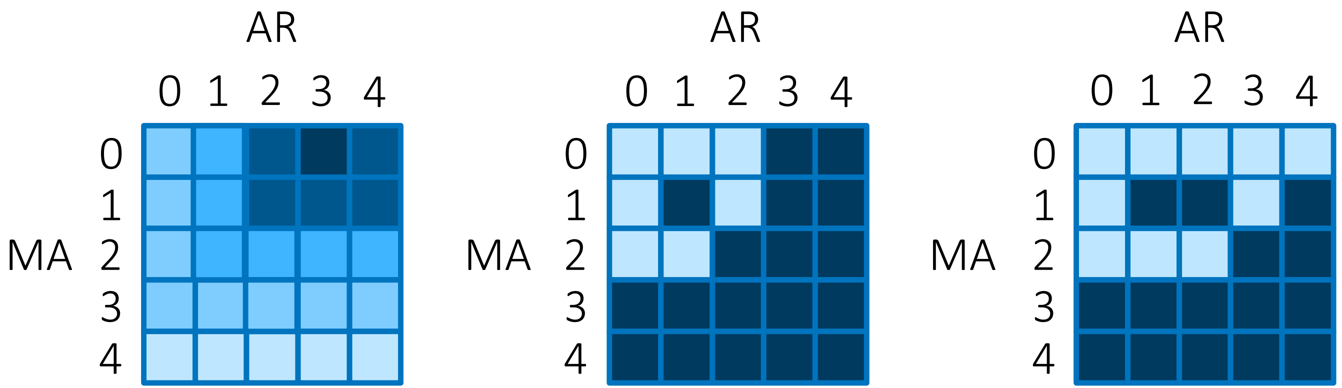
AR

		0	1	2	3	4
	0					
	1					
MA	2					
	3					
	4					

ESACF

Automatic “Selection” Techniques

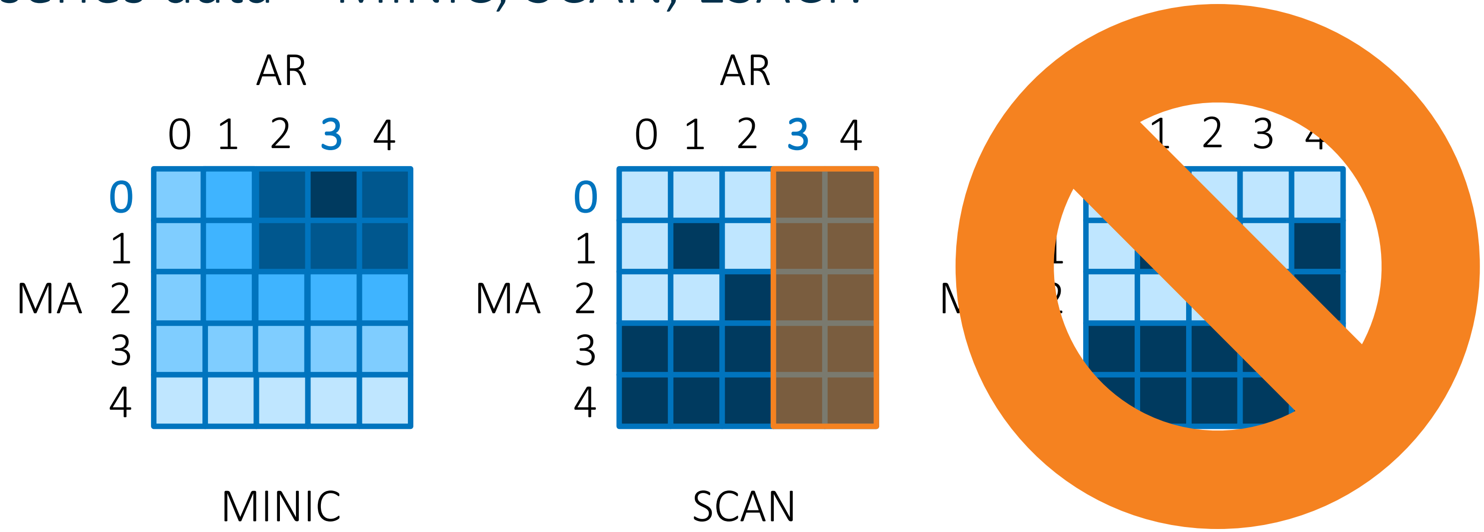
- SAS has three main automatic “selection” techniques for time series data – MINIC, SCAN, ESACF.



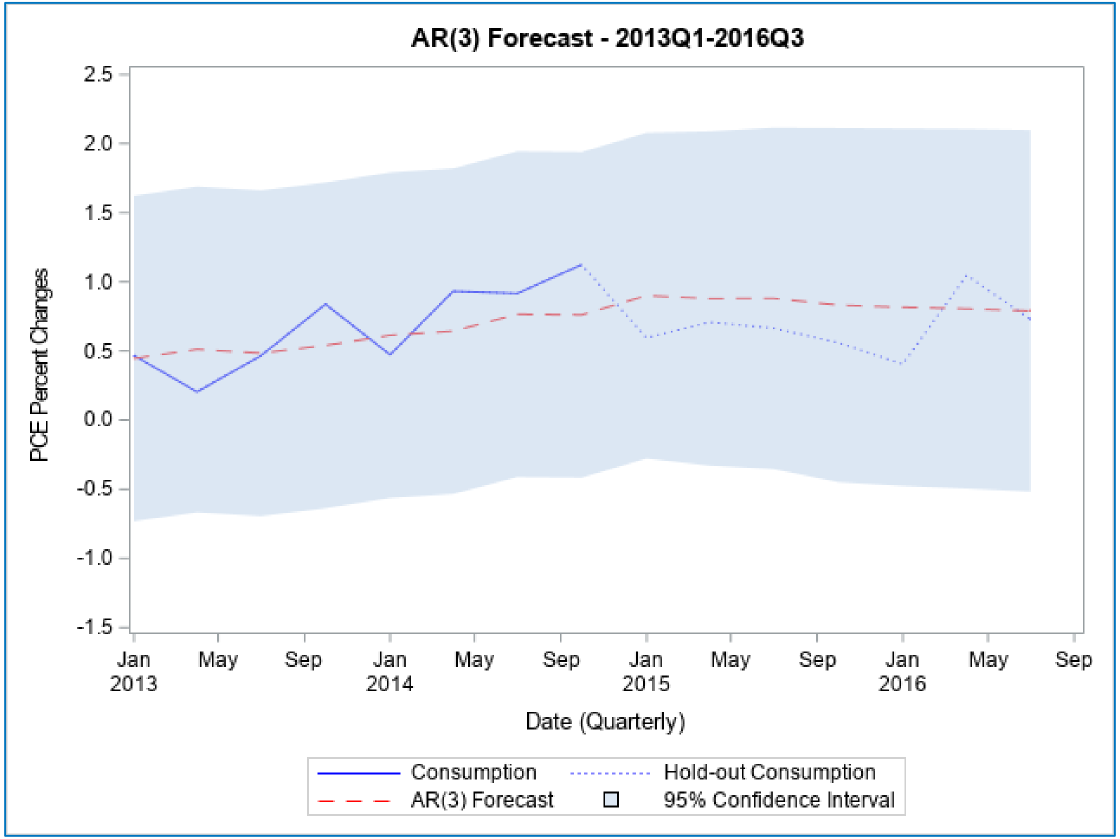
MINIC SCAN ESACF

Automatic "Selection" Techniques

- SAS has three main automatic "selection" techniques for time series data – MINIC, SCAN, ESACF.



Autoregressive Model



Classical Time Series Modeling

Vector Autoregressive Model

Multivariate Time Series

- Multivariate regression is **NOT** multiple regression.
- Trying to predict multiple **target** variables at the same time.

Vector Autoregressive Models

- Extension of ARMA models into multivariate models are VARMA models.
- In practice people typically use VAR instead of VARMA.
- Example of VAR(1):

$$\begin{bmatrix} Y_{t,1} \\ Y_{t,2} \end{bmatrix} = \begin{bmatrix} \alpha_{0,1} \\ \alpha_{0,2} \end{bmatrix} + \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} \begin{bmatrix} Y_{t-1,1} \\ Y_{t-1,2} \end{bmatrix} + \begin{bmatrix} e_{t,1} \\ e_{t,2} \end{bmatrix}$$

Vector Autoregressive Models

- Extension of ARMA models into multivariate models are VARMA models.
- In practice people typically use VAR instead of VARMA.
- Example of VAR(1) **expanded**:

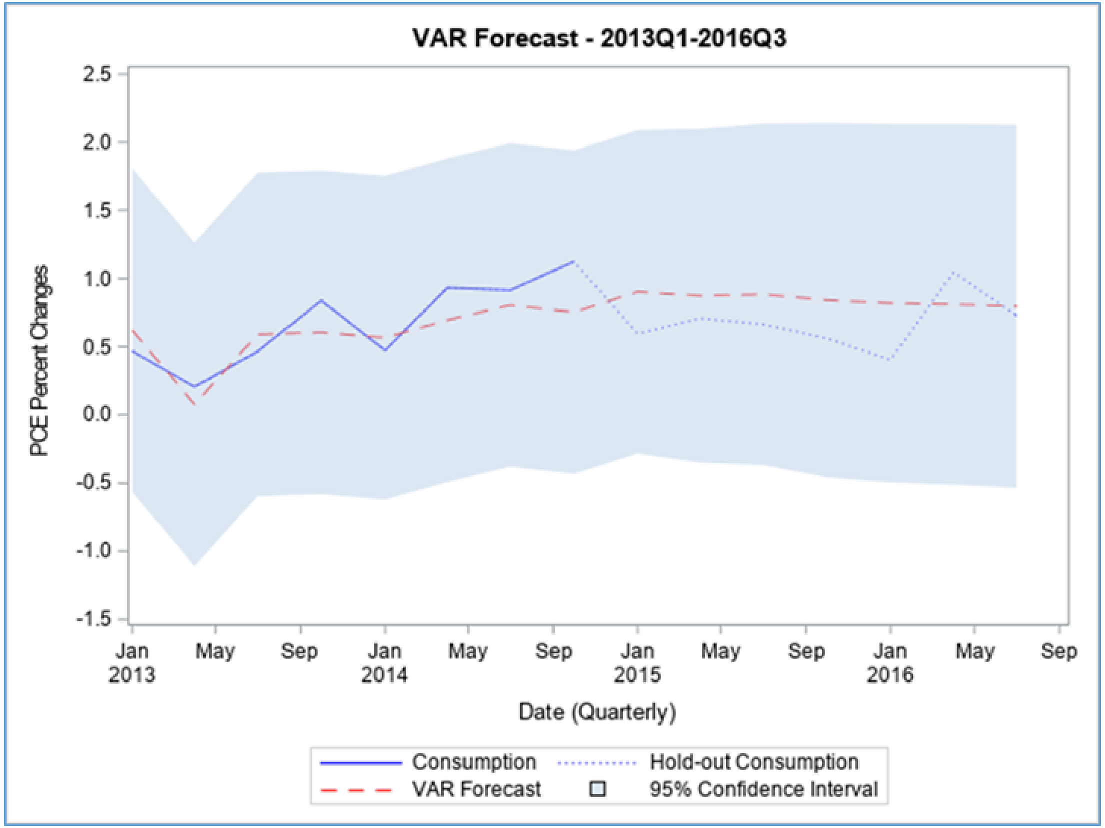
$$Y_{t,1} = \alpha_{0,1} + \alpha_{11,1}Y_{t-1,1} + \alpha_{12,1}Y_{t-1,2} + e_{t,1}$$
$$Y_{t,2} = \alpha_{0,2} + \alpha_{21,1}Y_{t-1,1} + \alpha_{22,1}Y_{t-1,2} + e_{t,2}$$

Consumption and Income

- What if predicting both consumption **and** income simultaneously helped with the prediction of consumption?
- Using VAR(3):

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 Y_{t-1} + \hat{\alpha}_2 Y_{t-2} + \hat{\alpha}_3 Y_{t-3}$$

Vector Autoregressive Model



Bayesian Time Series Modeling

Bayesian Autoregressive Model

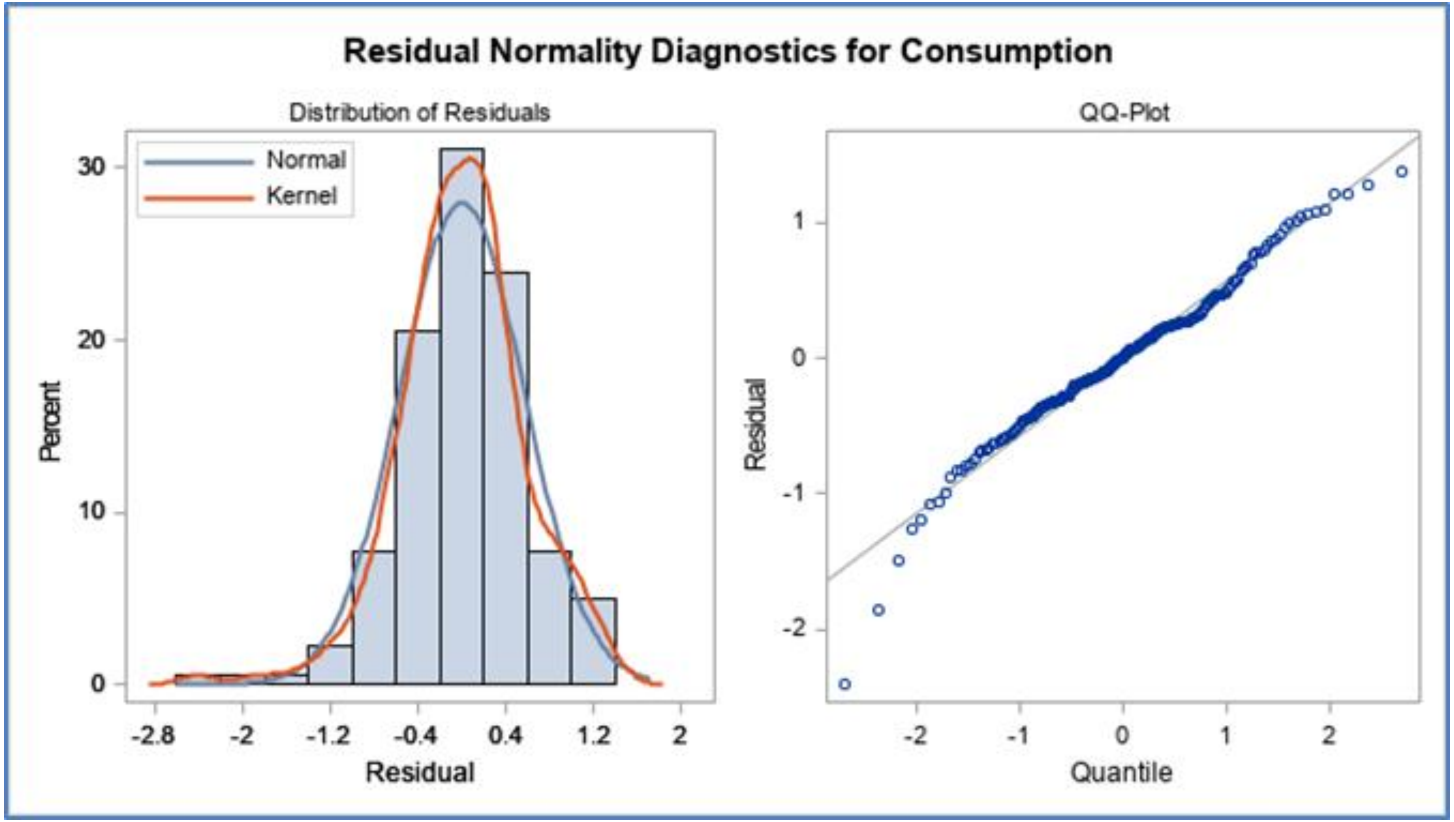
Bayesian Autoregressive (BAR) Models

- Model structure the same as AR model:

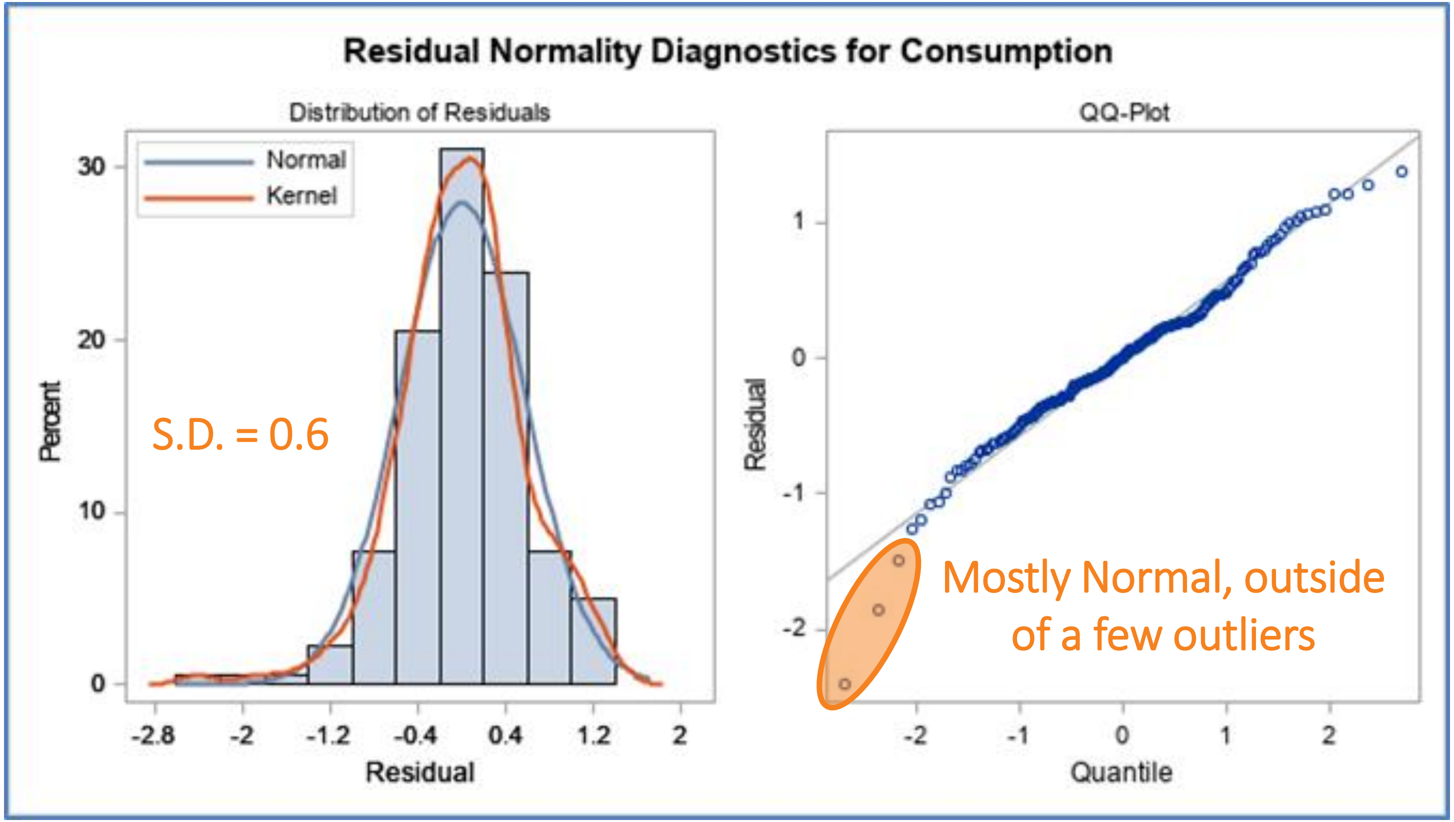
$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 Y_{t-1} + \hat{\alpha}_2 Y_{t-2} + \hat{\alpha}_3 Y_{t-3}$$

- Differences are in the assumptions!
  - Distribution of Y's?
  - Distribution of parameters?

Distribution of Consumption



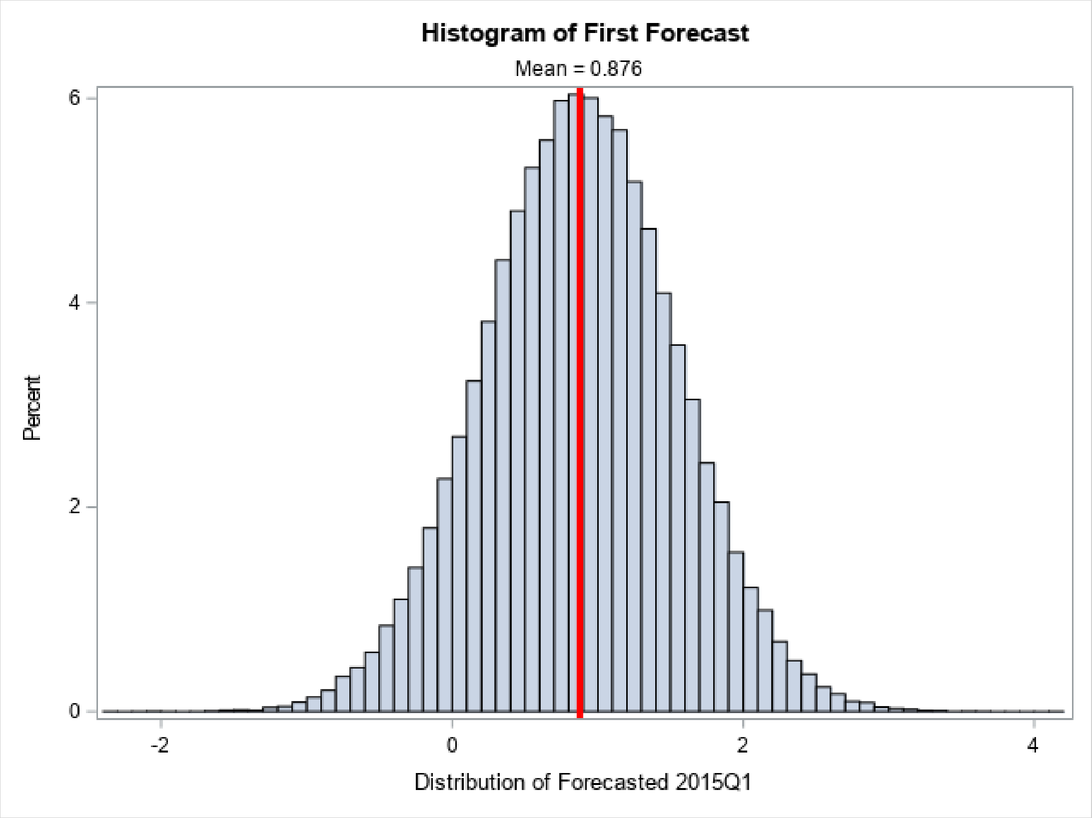
Distribution of Consumption



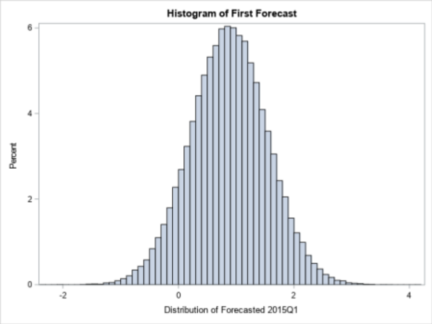
Assumed Distributions

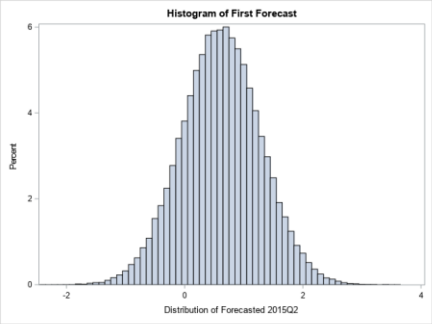
Random Variables	Assumed Distribution
Consumption	Normal Mean = $\hat{\alpha}_0 + \hat{\alpha}_1 \hat{Y}_{t-1} + \hat{\alpha}_2 \hat{Y}_{t-2} + \hat{\alpha}_3 \hat{Y}_{t-3}$ S.D. = σ^2
Consumption Initial Values Y_0, Y_1, Y_2	Normal Mean = 0, S.D. = 1
AR Parameters (α's)	Normal Mean = 0, S.D. = 1
Model Variance (σ^2)	Gamma Shape = 0.3, Scale = 3.33

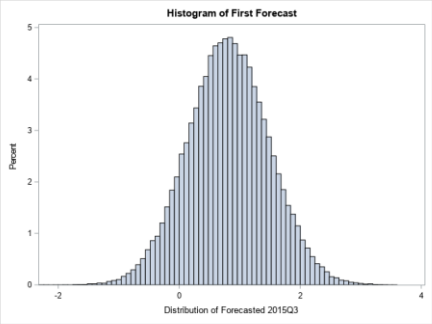
Posterior Distributions for Forecasts

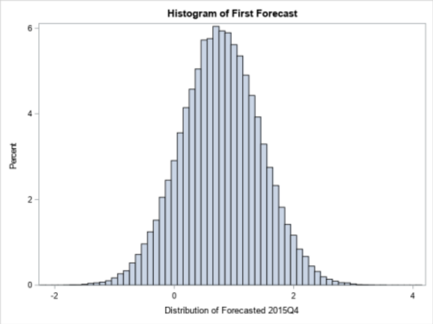


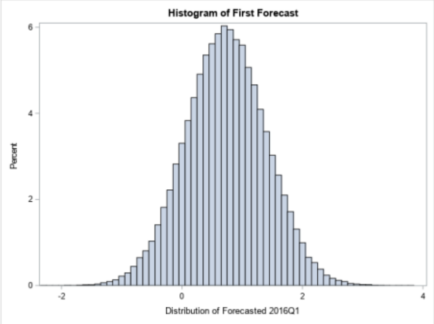
Posterior Distributions for Forecasts

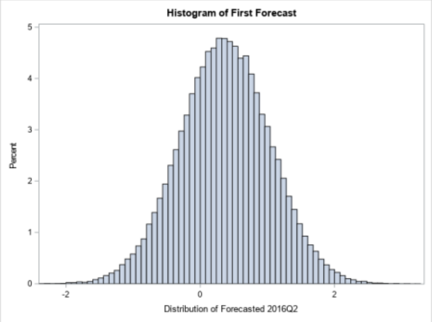


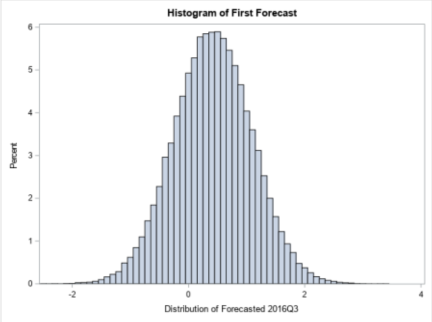




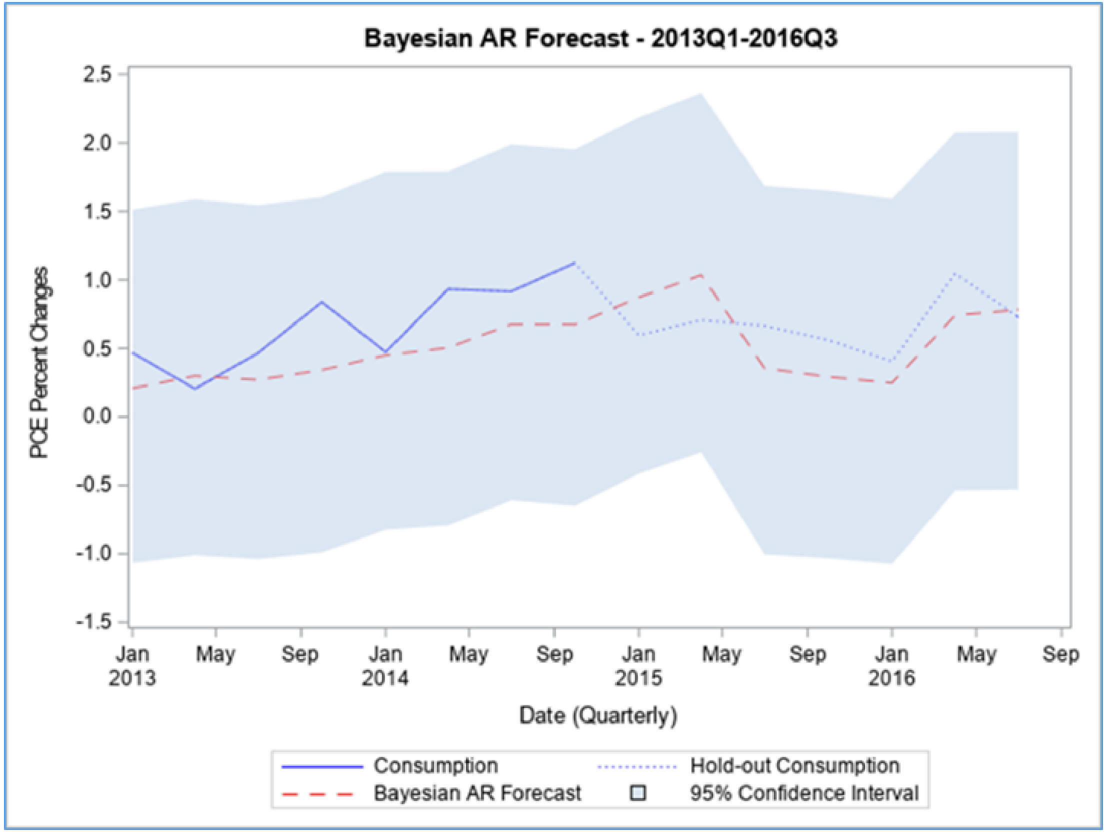








Bayesian Autoregressive Model



Bayesian Time Series Modeling

Bayesian Vector Autoregressive Model

Bayesian Vector Autoregressive (BVAR) Models

- Model structure the same as VAR model:

$$\hat{\boldsymbol{Y}}_{\boldsymbol{t}} = \hat{\boldsymbol{\alpha}}_{\boldsymbol{0}} + \hat{\boldsymbol{\alpha}}_{\boldsymbol{1}}\boldsymbol{Y}_{\boldsymbol{t-1}} + \hat{\boldsymbol{\alpha}}_{\boldsymbol{2}}\boldsymbol{Y}_{\boldsymbol{t-2}} + \hat{\boldsymbol{\alpha}}_{\boldsymbol{3}}\boldsymbol{Y}_{\boldsymbol{t-3}}$$

- Differences are in the assumptions!
  - Distribution of Y **vectors**?
  - Distribution of parameter **matrices**?

Bayesian Vector Autoregressive (BVAR) Models

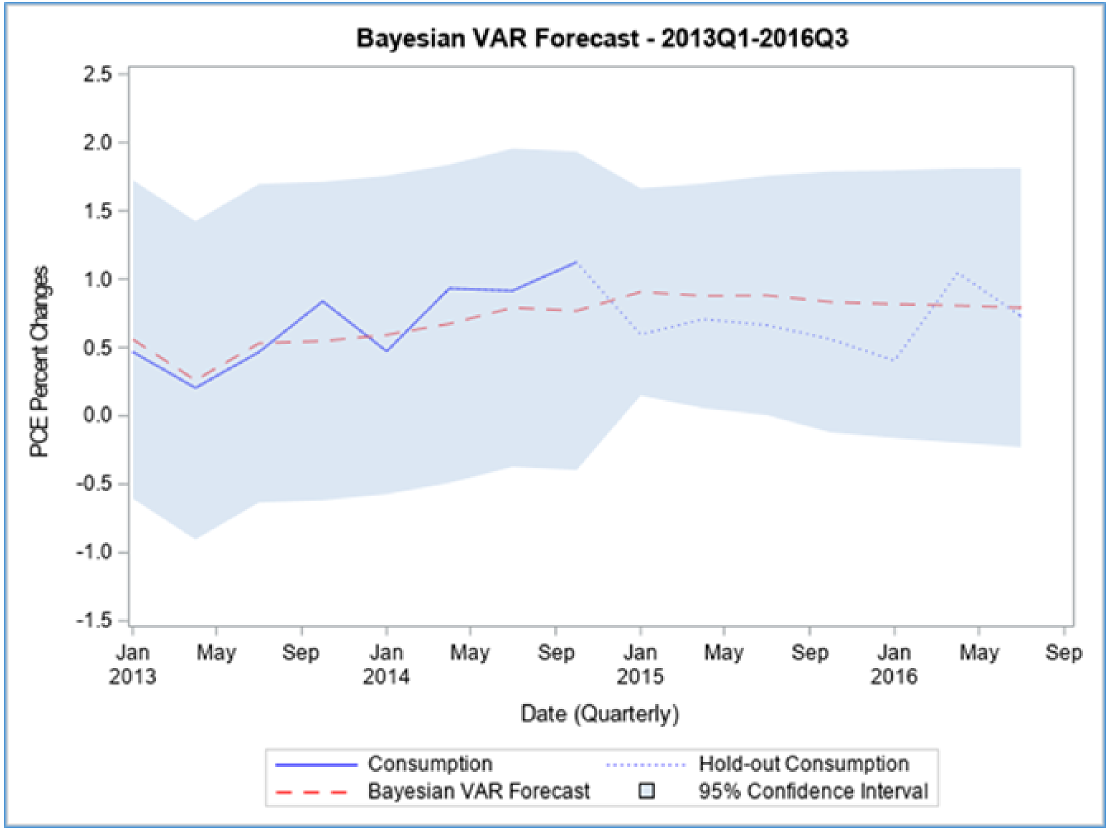
- Model structure the same as VAR model:

$$\widehat{Y}_t = \widehat{\alpha}_0 + \widehat{\alpha}_1 Y_{t-1} + \widehat{\alpha}_2 Y_{t-2} + \widehat{\alpha}_3 Y_{t-3}$$

- Differences are in the assumptions!
  - Distribution of Y vectors?
  - Distribution of parameter matrices?

Multivariate Normal

Bayesian Vector Autoregressive Model



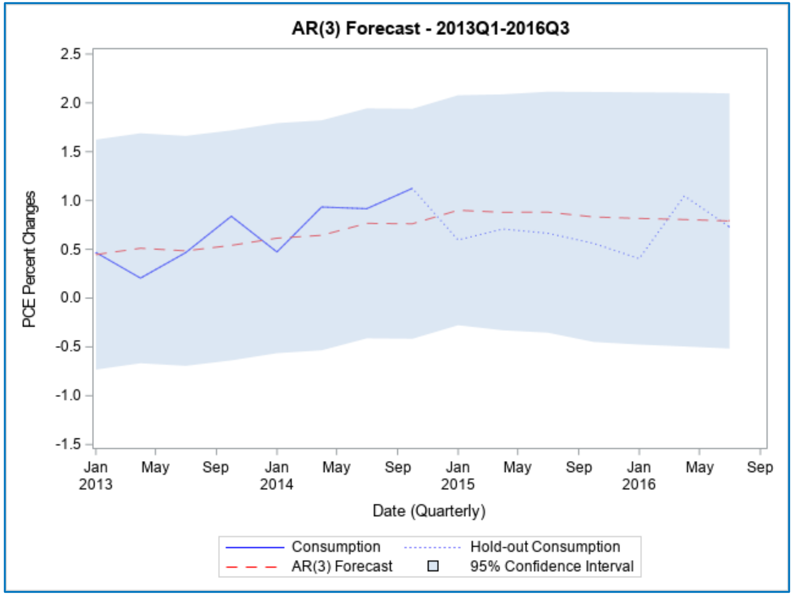
Comparison

Mean Absolute Percentage Error

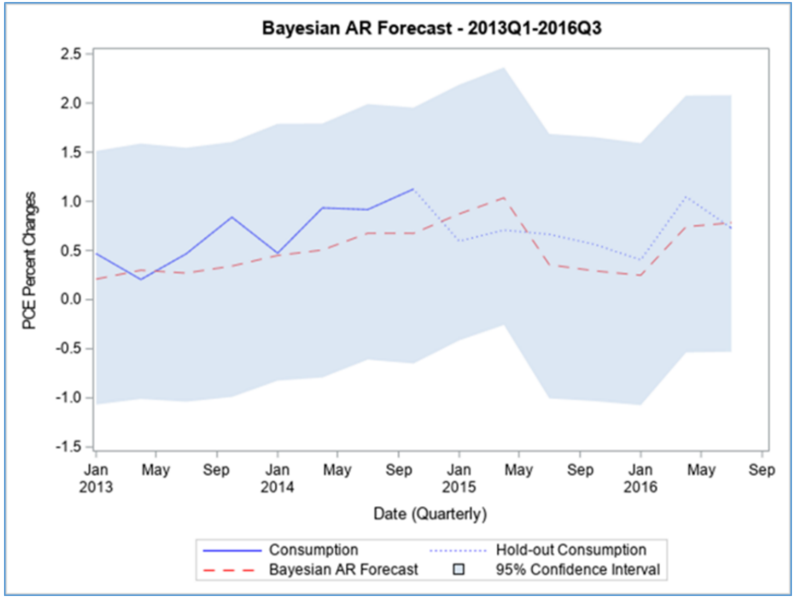
Model	Hold-out MAPE
Bayesian Autoregressive(3)	37.4%
Autoregressive(3)	41.3%
Bayesian Vector Autoregressive(3)	41.6%
Vector Autoregressive(3)	41.8%
Simple Exponential Smoothing Model	43.4%

Ensemble Model

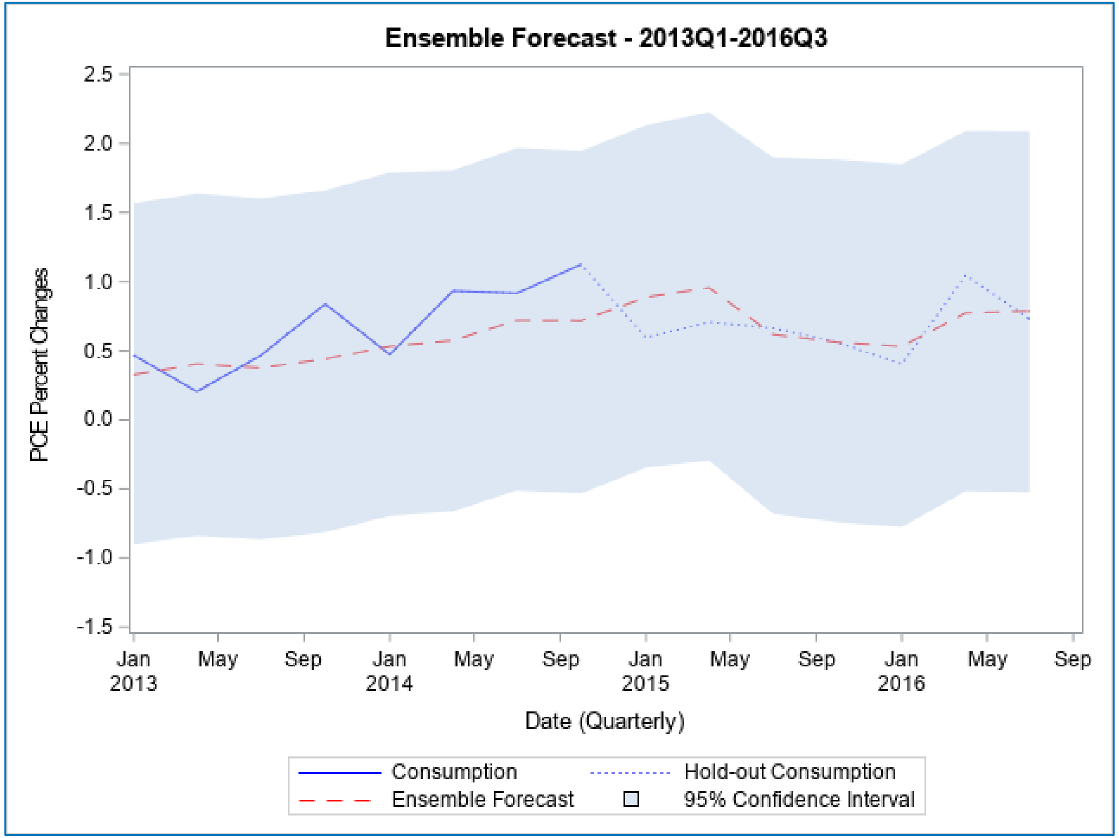
FREQUENTIST



BAYESIAN



Ensemble Model



Mean Absolute Percentage Error

Model	Hold-out MAPE
Ensemble (BAR & AR)	**22.4%**
Bayesian Autoregressive(3)	37.4%
Autoregressive(3)	41.3%
Bayesian Vector Autoregressive(3)	41.6%
Vector Autoregressive(3)	41.8%
Simple Exponential Smoothing Model	43.4%

Thank you!

Contact Information
aric_labarr@ncsu.edu

Reminder:

Complete your session survey in the conference mobile app.


#SASGF
SAS®
GLOBAL
FORUM
2019
APRIL 28 – MAY 1, 2019 | DALLAS, TX
Kay Bailey Hutchison Convention Center