# Time Series Forecasting Real World Data: Auto Regression Using a Neural Network Forecaster with Weighted Windows

By

Brad Morantz Ph.D.

# Time Series Forecasting

- Used when don't have causal variables to build model
  - Don't know all of the variables
  - Can't measure all of the variables
  - Do not have metrics to describe variables
  - Do not know the model
- Assume whatever has been happening will continue (from near past to near future)
- Effect of causal variables reflected in criterion/dependent lagged values
- Use lagged values to predict future

# Problems

- Nonlinearity
  - The real world is not linear
  - Usually not even Gaussian
- Changing Causal Factors
  - Time of occurrence is important
  - Underlying factors change
    - Slowly
    - Quickly
    - Combination of both

# Changing Causal Factors

- When there is a system shock
  - Example: Sept 11 effect on the stock market
- When there is a slow change over time in causal factors
  - Example: the price of computer memory
    - New processes and methods
    - New standards
    - Economies of scale
    - New or different needs
    - Competition

# Tracker

- Passive object
  - Behaves according to projectile theory
  - Other items & gravity act upon it
- Active object
  - Fins move, motors turn on
  - Engine causes maneuver
  - Sudden changes in direction
  - Sudden changes in acceleration

# Remediation

- Transform Data
  - Square
  - Square root
  - Trig function
  - Exponential
  - Log
  - etc.
- Must convert back in order to interpret

# Pros and Cons of Transforms

## Pros

- More linear
- Better fit
- Less error

## Cons

- Need to convert back before can assess error or fit or apply results
- Does not always work
- Requires Time
- Requires operator skill

# Current Methods

- OLS autoregression
  - Best fit linear model incorporating all data
- Random Walk
  - Naïve forecast expects no changes
- ARIMA/Box-Jenkins
  - Requires model
  - Incorporates all data
- Moving average

# Methods Cont'd

- Exponential smoothing
- Specialized non-linear
  - ARCH, GARCH, Bilinear, TAR, STAR
- Kalman Filter, extended Kalman
- Batch Least Squares
- Artificial neural network
- Many more

# Neural Network

- Can fit non-linear models
- Not affected by multi-colinearity
- Can solve multiple step horizon
- Learns input output relationship
- More immune to noise
- The general approximator
- Data driven - few a priori rules
- Successfully used in forecasting

# Ways to Cope with Changes

- Short forecast horizons
  - use only more recent data
- Weighted regression
  - NIST states effective for heteroscedasticity
- Moving versus Rolling Windows
  - Control number of observations used for training/building model

# Training Data

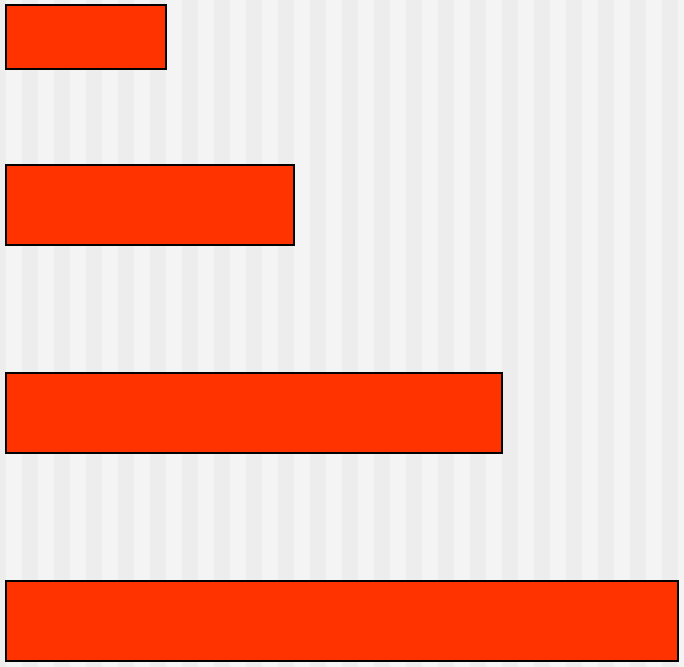
- Used to build model
- Must clean data before use
  - Garbage in = garbage out
- Methods to use data
  - Rolling Window
  - Moving Window
  - Weighted Window (WW)
- Must maintain positive degrees of freedom

# Rolling vs. Moving Window

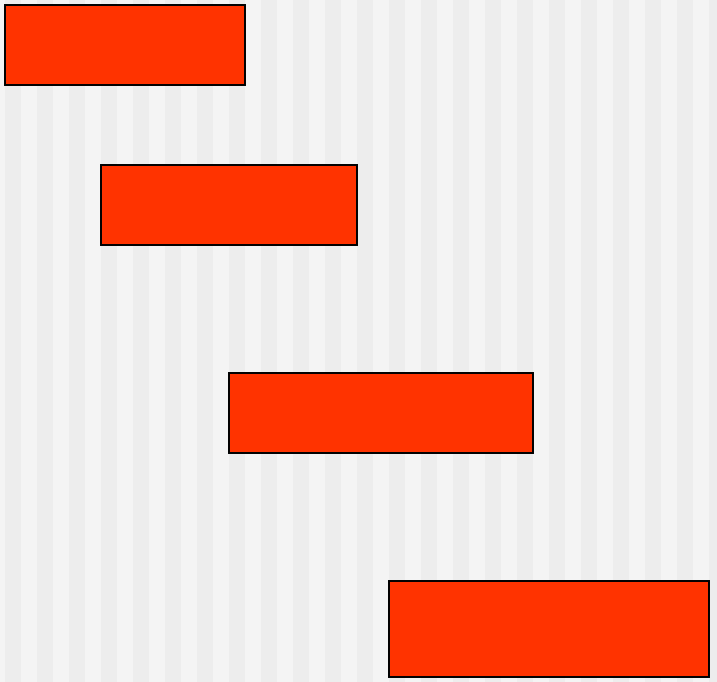
- Rolling Window
  - Rolls forward including all data behind
  - Constant starting point with ever increasing size
- Moving Window
  - Deletes the oldest as it adds the newest
  - Constant size with ever increasing starting point

# Rolling vs. Moving Window

- Rolling Window



- Moving Window



# Advantages

- Rolling Window
  - Learns from all
  - Doesn’t forget
  - Everything is equally important
  - Can reduce variance and confidence interval
- Moving Window
  - Not affected by causal changes
  - Not as affected by system shocks

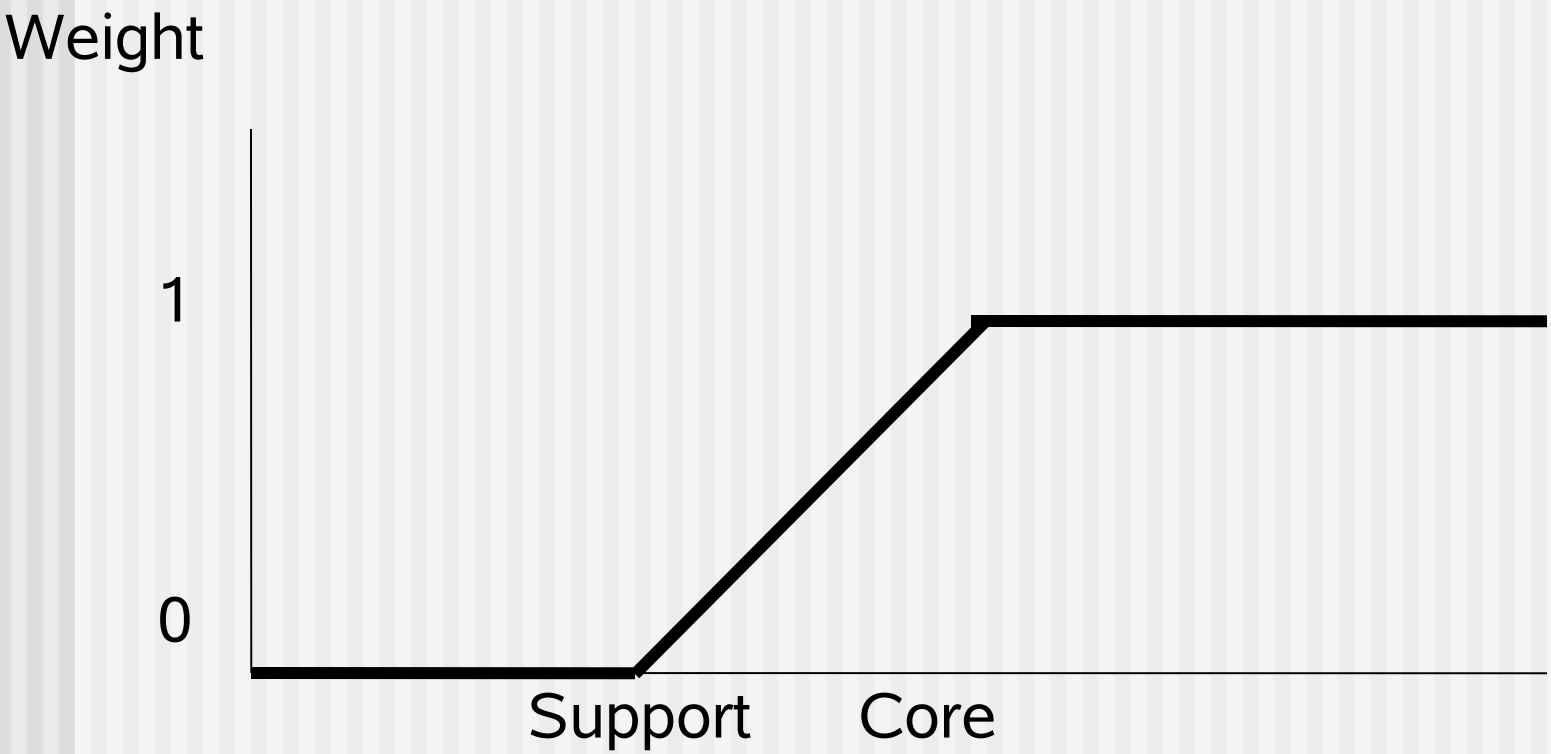
# Disadvantages

- Rolling Window
  - Might include too much information
  - May include rule that now has smaller affect
  - May include rule that has no affect
- Moving Window
  - May not include enough data
  - May have lost valuable information
  - Might have deleted important rule

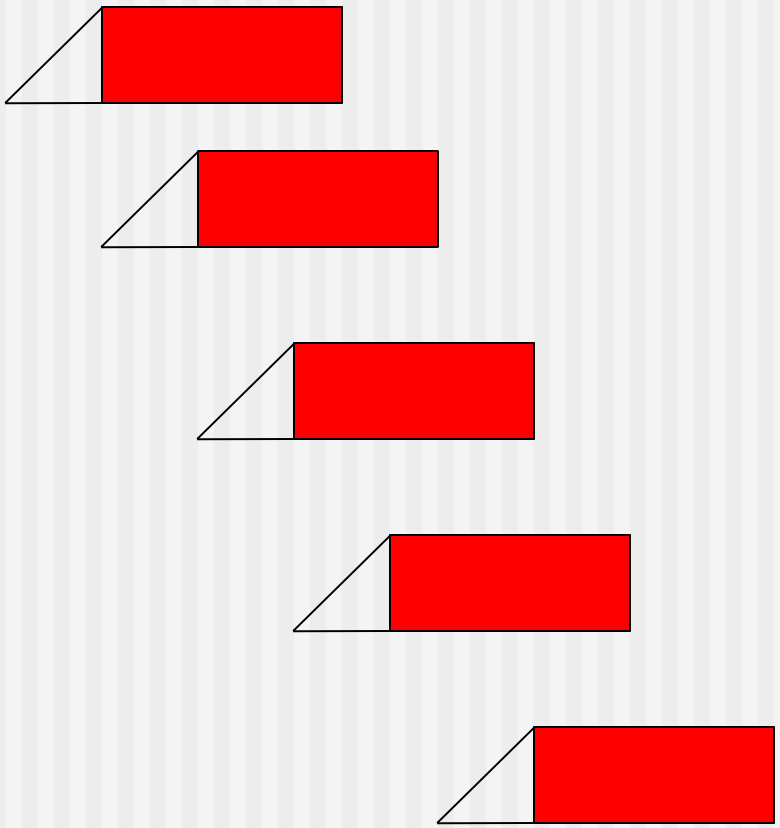
# Weighted Window (WW)

- Currently uses linear ramp
  - Until improvement developed
- Deletes information that is too old
- The newest information is the most important (recency rule)
- Older information is retained but is not as important as new data
- Superset of rolling & moving windows

# Linear Ramp



# Weighted Window



# Selecting Window Size

- Trial and Error
  - Try many
  - Select the one with least error
- A priori knowledge
- Heuristics or formulae
- Application of Chaos theory

# Macro View

- Neural Network Forecaster with Weighted window (WW)
- Learns from history
  - Newest information is most important
  - Older knowledge is not lost, it is just not as important as the most recent
  - Can fit non-linear applications

# Results

| Data set | ANN* | WW* | Change |
|---|---|---|---|
| CD Rates | 11.4% | 5.36% | 52.98% |
| Eurodollar | 4.98% | 4.13% | 17.17% |
| Fed Funds | 4.41% | 2.49% | 43.39% |
| N Hse Sals | 20.12% | 11.19% | 44.38% |
| FR Franc | 3.26% | 2.79% | 14.48% |

*Mean Absolute Percentage Error (MAPE)(best of rolling or moving window)

Change is reduction in MAPE

# Potential Applications

- Time series with changing causal variables
- Time series that has system shocks
- Examples:
  - Projectile with some active control
  - Stock market
  - Sales of popcorn

# Future

- Optimizing Window size
  - Application of Chaos Theory [Frank et al]
  - Selecting similar data [Shimodiara]
  - Use of genetic algorithm (GA)
- Optimizing weighting pattern
  - A priori knowledge
  - Heuristics & formulae

# Future Cont'd

- Weighted window with neural network classifier
- Optimizing Architecture of ANN
  - Genetic Algorithm
  - Embedding theorem [Frank et al]
  - A Priori knowledge
  - Combination of above

# Home Project

- Just built quad core machine
- Projectile simulation
  - Active
  - Passive
- Compare accuracy, precision, computation time
- Use Kalman vs Neural
- Analyze with multi-factor ANOVA

# Appendix
# Statistical Analysis Methods

- Two factor ANOVA
  - Time of forecast
  - Treatment (type of window)
- MAPE derived from 31 forecasts
- One unit (month) horizon
- ANOVA performed
- Tukey's HST performed
- Longer horizons also forecast with comparable results

# Results of Two Factor ANOVA

- Statistically Significant
  - CD rates
  - Federal Funds Rate
  - New House Sales
- Not significant - excessive variance within treatment
  - Eurodollar
  - French Franc

# More Statistical Analysis

- Three Factor ANOVA
  - Treatment (type of window)
  - Time of forecast
  - Data set
- Results
  - All treatments were statistically significant

# Final Note

- This was part of a more exhaustive study.
- Academic references are available upon request
- Yes, more research is continuing in this area

# References

- www.machine-cognition.com
- IEEE computational intelligence society
- TIMS/ORSA
- *Times Series Forecasting* Bauerman & Occonel

# My Contact Information


- Brad Morantz PhD
- www.machine-cognition.com
- bradscientist@ieee.org
- www.cognitive-decisions.com

# Questions?