Forecasting Firearm Innovation: A Time Series Analysis of U.S. Patent Trends and Macroeconomic Influencers, 1820-2025

Malcolm Hsu
Brandeis University ‘26 | M.S. Business Analytics
Wesleyan University ‘25 | B.A. Mathematics & Economics
9 October 2025

Research Questions & Focus

Q: How has firearm innovation developed in the United States since 1820?

Q: How do we see the future in firearms?

Q: How do fluctuations in the U.S. economy affect arms development?

Q: Does the U.S. Government/military expenditures have an impact on firearm innovation?

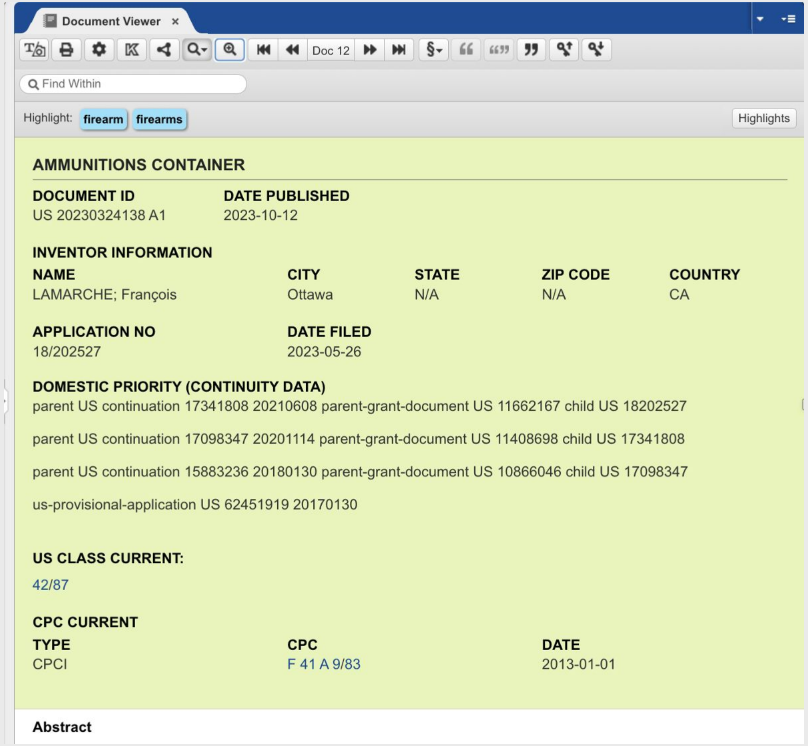
Goals:

1. Understand the best method to track patents
2. Test, model, and validate data to most accurately forecast the future
3. Identify exogenous macroeconomic variables that best explain the trends in the data
4. Explore U.S. Government/Military patents

An Overview of the F41 Patent

Patent classifications

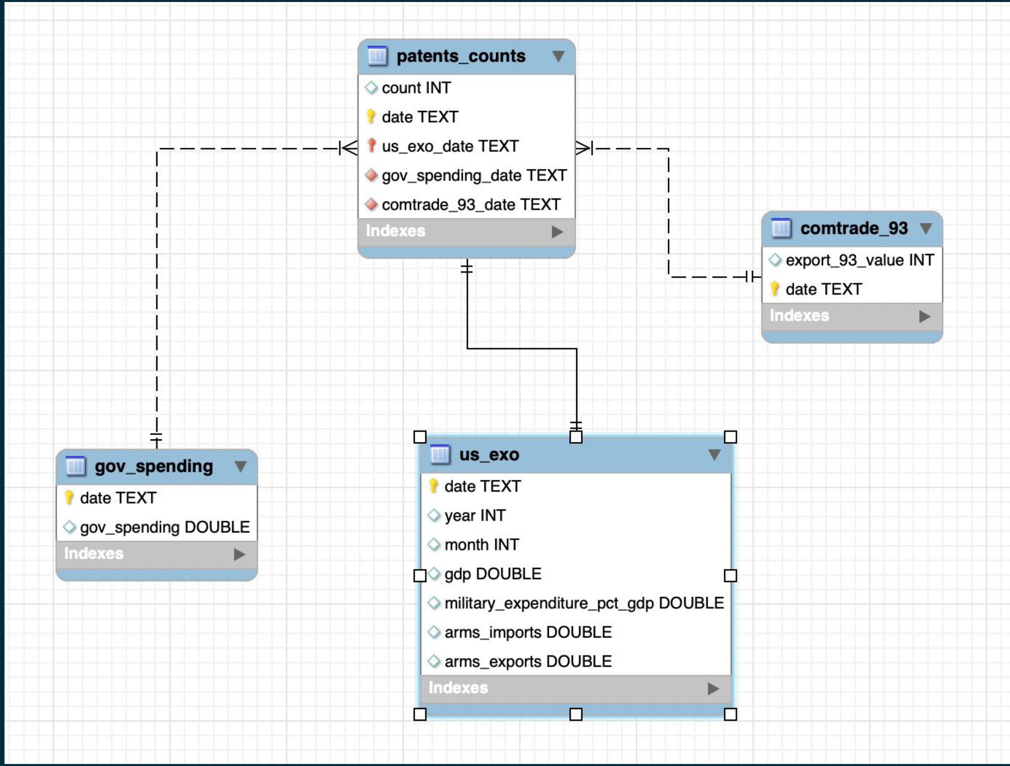
- F41 - Weapons
  - A: Functional features of details common to both smallarms and ordinance
  - B: Weapons for projecting missiles without the use of explosive/combustible propellant charge
  - C: Smallarms
  - F: Apparatus for launching projectiles/missiles from barrels
  - G: Weapon sights
  - H: Armour, armoured turrets, armoured or armed vehicles
  - J: Targets, target ranges, bullet catchers



Data & Methods

Methods

- Time series models
  - Naive models
  - ARIMA
  - TAR
  - ARIMAX
- Cross validation
- Forecasting
- Regression analysis



Workflow & R Packages

1. Data import & variable selection
2. Data cleaning, merging, & variable creation
3. Time series creation & testing
4. Run models
5. Cross validation
6. Results & discussion

Packages & use

Data importing

- ❑ readr → import CSV's
- ❑ readxl → import Excel files
- ❑ curl → powers API's
- ❑ comtradr → UN Comtrade API pull

Data manipulation

- ❑ dplyr → data manipulation
- ❑ tidyr → reshaping data
- ❑ stringr → string manipulation
- ❑ purrr → mapping functions for data manipulation
- ❑ zoo → handling irregular time series
- ❑ lubridate → date/time handling

Data visualization

- ❑ ggplot2 → visualization
- ❑ modelsummary → format tables
- ❑ gt → format tables

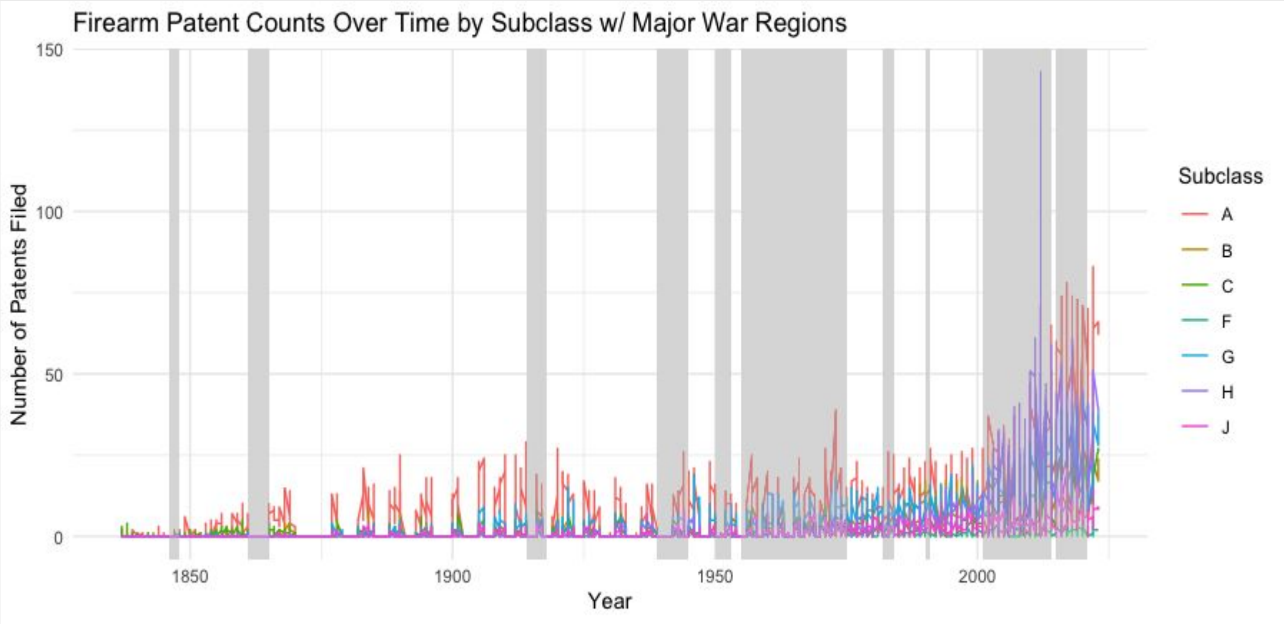
Time series modeling

- ❑ tseries → time series functions
- ❑ forecast → time series model forecasting functions
- ❑ TSA → functions for analysis of time series models
- ❑ tsDyn → TAR model functions

Regression analysis

- ❑ MASS → regression tools
- ❑ broom → handling regression objects
- ❑ fixest → econometric estimators

Time Series Data



Time series representation of firearm patent data, faceted by subclass, shaded by major U.S. war periods

Goals of time series analysis

- ❏ Make predictions & forecast data
- ❏ Understand underlying patterns in the data

Components of time series data

Time

- ❏ One variable measured sequentially at fixed intervals

Series

- ❏ Data is ordered and sequential

Assumptions

- ❏ Stationarity: no dependence between observations
- ❏ Errors are random

Time Series Methods

Benchmark: Naive Methods

Drift model

- ❏ Assumes future values follow a linear trend from the first and last observations

Mean model

- ❏ Assumes future values are dependent on the mean of the data

Naive model

- ❏ Assumes future values are dependent on the most recent observation

Autoregressive Integrated Moving Average Model (ARIMA)

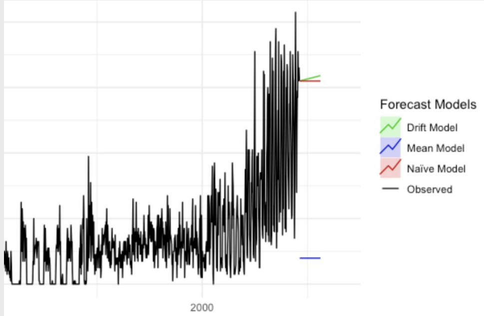
- ❏ Forecasts values based on past month values and past month errors

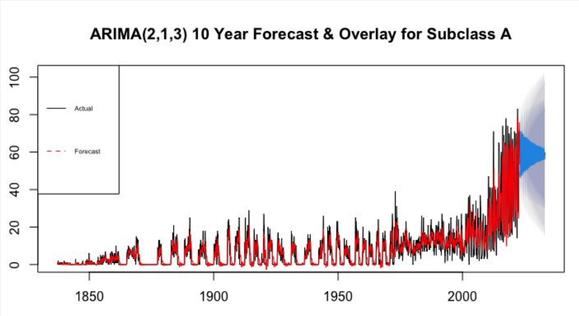
Threshold Autoregressive Model (TAR)

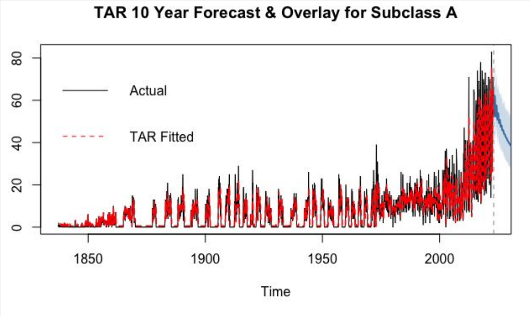
- ❏ Best choice to capture differences in pre/post- 21st century trends
- ❏ Similar to ARIMA model but allows for regime switching behavior
- ❏ Splits data into low & high regime

Time Series Plots

Model complexity







Drift, mean, naive
- ❏ Benchmark models

ARIMA
- ❏ Step up from benchmark
- ❏ AIC: 14221.7
- ❏ Number of patents is best explained by the past two months and past three month errors

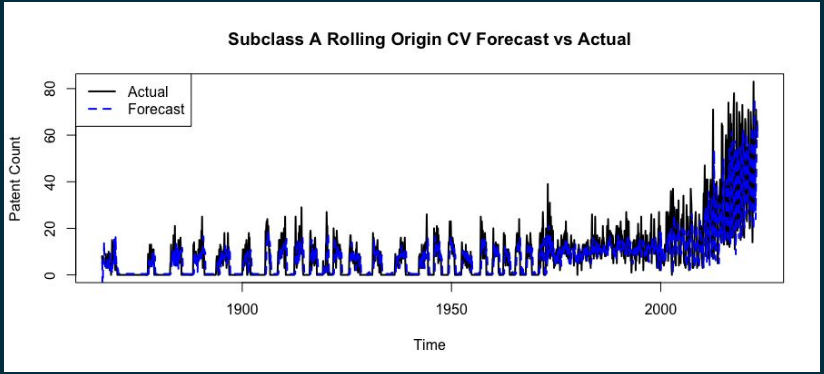
TAR
- ❏ Most complex model
- ❏ AIC: 7824.2

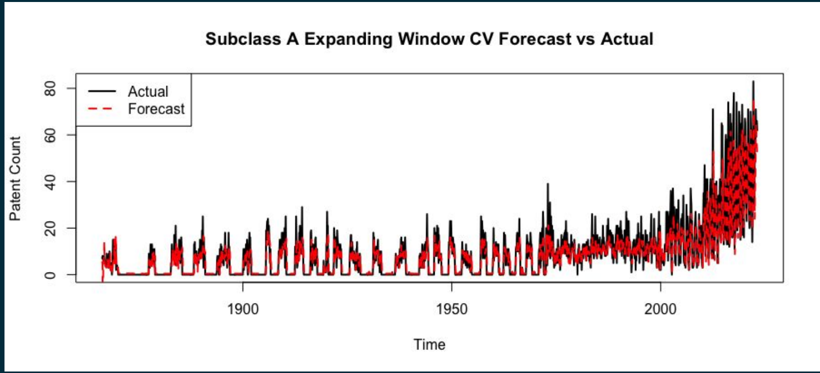
Cross Validation of TAR Models

Goals of cross validation
- ❑ Estimate how well the model fits data
- ❑ Testing the ability to forecast future values

Rolling Origin CV
- ❑ Fixed-size training windows (20 years) that move forward in time over each iteration
- ❑ Mean squared error (MSE): 87.23568

Expanding Window CV
- ❑ Starting with an initial training window (20 years), adding a new observation each iteration
- ❑ MSE: 42.4575





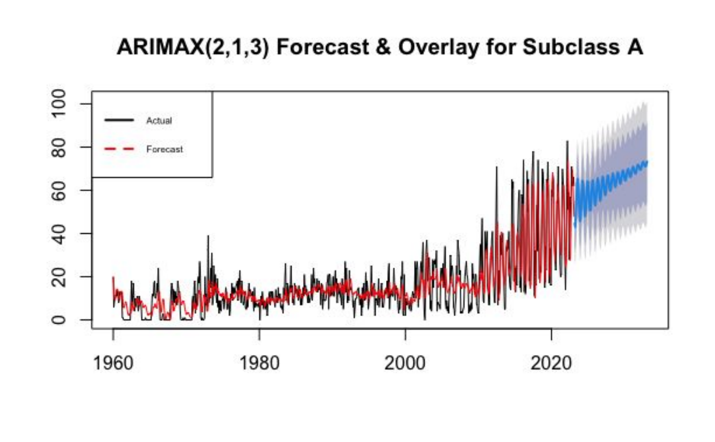
U.S. Government Patents

Keywords

- ❑ Any branch of the U.S. Government & military are included in "private" patents

Measures

- ❑ Subset data into U.S. government only
- ❑ Inclusion of macroeconomic variables:
  - ❑ Government spending
  - ❑ Military expenditure
  - ❑ Arms imports
  - ❑ Arms exports
  - ❑ Proportion of U.S. patents/month
- ❑ AIC: 5399.04



Measuring the Effect of War Using Count Regressions

Poisson Regression

$$count_i \mid war_i, cluster_i \sim Poisson(\mu_i)$$

$$Var(count_i) = \mu_i$$

- Best suited to deal with count data
- Assumes variance = mean

Interpretation

- Predictors influence the log of expected count

$$log(\mu_i) = \alpha_0 + \beta_1 war_i + \sum_{j=1}^{N} \delta_j C_{ij} + \sum_{j=1}^{N} \gamma_j (war_i \cdot C_{ij})$$

Negative Binomial Regression

$$count_i \mid war_i, cluster_i \sim NegBin(\mu_i, \theta)$$

$$Var(count_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- Generalization of Poisson
- Allows for overdispersion
  - Variance ≠ mean

Interpretation

- Predictors influence the log of expected count, allows variance

$$\mu_i = E[count_i] = e^{\alpha_0 + \beta_1 war_i + \sum_{j=1}^{N} \delta_j C_{ij} + \sum_{j=1}^{N} \gamma_j (war_i \cdot C_{ij})}$$

Count Regressions Summary

subclass	model	term	estimate	std.error	p.value	effect_type	significance
A	km_cluster	war1	1.07	0.11	0.00	main_war_effect	significant
A	km_cluster	war1:km_cluster2	0.58	0.92	0.53	interaction	insignificant
A	km_cluster	war1:km_cluster3	0.44	0.12	0.00	interaction	significant
A	km_cluster	war1:km_cluster4	0.80	0.15	0.00	interaction	significant
A	km_cluster	war1:km_cluster5	0.15	0.12	0.19	interaction	insignificant
A	km_cluster	war1:km_cluster6	0.07	0.12	0.56	interaction	insignificant
A	hc_cluster	war1	1.22	0.04	0.00	main_war_effect	significant
A	hc_cluster	war1:hc_cluster2	−0.22	0.09	0.02	interaction	significant
A	hc_cluster	war1:hc_cluster3	0.31	0.08	0.00	interaction	significant
A	hc_cluster	war1:hc_cluster4	−0.04	0.07	0.60	interaction	insignificant
A	hc_cluster	war1:hc_cluster5	0.29	0.07	0.00	interaction	significant
A	hc_cluster	war1:hc_cluster6	0.43	0.92	0.64	interaction	insignificant

Effect of war regressions

- ❑ Each subclass has two types of clustering: k-means and hierarchical
- ❑ Each model has war and interaction terms

Results

- ❑ Times of war are always significant

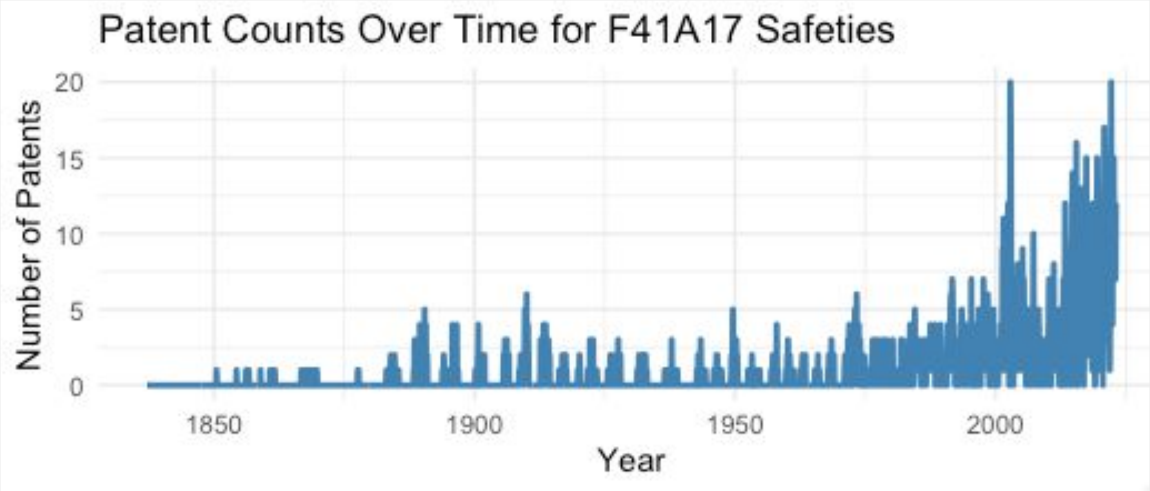
Case Study: F41A17 Safety Arrangements

Studying F41A17 classification

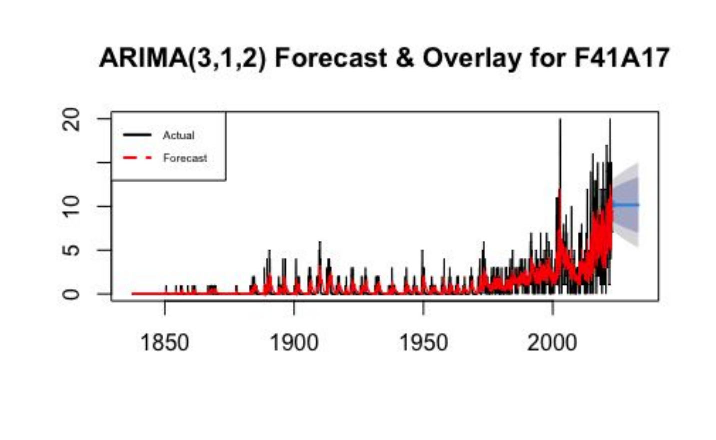
- ❑ Focus on modeling safety arrangements

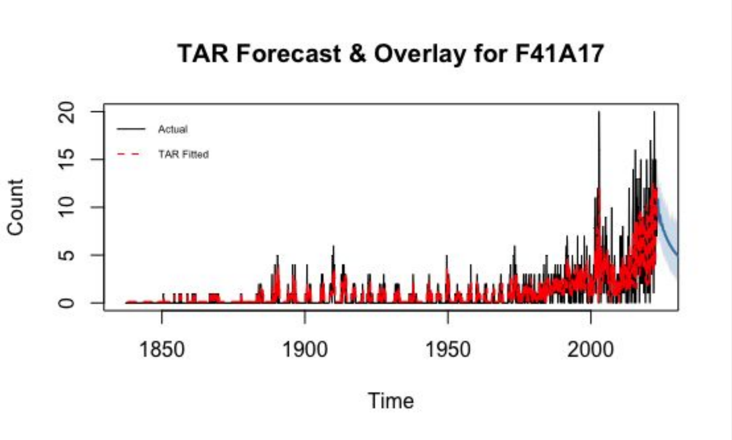
Models

- ❑ ARIMA
- ❑ TAR
- ❑ Regressions



Case Study: F41A17 Time Series Plots





ARIMA

- AIC: 7844.837
- Number of patents is best explained by the past three months and past two month errors
- Forecasts a steady trend in safety patents

TAR

- AIC: 1565.21
- Forecasts a decrease in safety patents

Case Study: F41A17 Regressions

Linear & Polynomial Regressions

Linear

$$count_i = \alpha_0 + \beta_1 gov_i + \beta_2 war_i + \beta_3 gdp_i + \beta_4 military_i + \beta_5 imports_i + \beta_6 exports_i + \delta_1 t_i + \epsilon_i$$

Polynomial

- ❑ Inclusion of time and time squared effects

$$count_i = \alpha_0 + \beta_1 gov_i + \beta_2 war_i + \beta_3 gdp_i + \beta_4 military_i + \beta_5 imports_i + \beta_6 exports_i + \delta_1 t_i + \delta_2 t_i^2 + \epsilon_i$$

Results

- ❑ Some significance across GDP, military expenditure, arms imports, and time
- ❑ Moderate explanatory power
  - ❑ R-squared: 0.383/0.394

	linear_1960	poly_1960
(Intercept)	2.77e+00*	−1.83e+00
	(1.12e+00)	(1.92e+00)
gov_spending	1.23e−04	2.37e−04
	(2.55e−04)	(2.69e−04)
war	−2.74e−01	−1.30e−01
	(2.56e−01)	(2.78e−01)
gdp	4.32e−13***	6.64e−13**
	(1.11e−13)	(2.08e−13)
military_expenditure_pct_gdp	−3.40e−01*	−3.45e−01**
	(1.32e−01)	(1.32e−01)
arms_imports	−3.01e−09	−3.51e−09
	(3.34e−09)	(3.36e−09)
arms_exports	2.46e−10	2.31e−10
	(4.73e−10)	(4.72e−10)
date	−2.95e−04***	
	(8.57e−05)	
poly(date, 2)1		−1.06e+02*
		(4.21e+01)
poly(date, 2)2		−1.63e+01
		(1.24e+01)
Num.Obs.	758	758
R2	0.393	0.394
R2 Adj.	0.387	0.388
AIC	3516.2	3516.5
BIC	3557.9	3562.8
Log.Lik.	−1749.120	−1748.237
RMSE	2.43	2.43

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Contributions

Time Series Analysis

- ❏ Comparing model fits between ARIMA and TAR models and forecast 10 years into the future
- ❏ Model improvement with the inclusion of economic variables

Regression Analysis

- ❏ Understanding the effects of times of war on the number of patents published
- ❏ Statistical significance of war across all patents

U.S. Government

- ❏ Better model fit when restricting to only “private” patents and including economic variables

Safety Arrangements Case Study

- ❏ Effectively model safety arrangement patents using time series models
- ❏ We expect a zero or decreasing trend
- ❏ Statistical significance of GDP, military expenditure, arms imports, and time