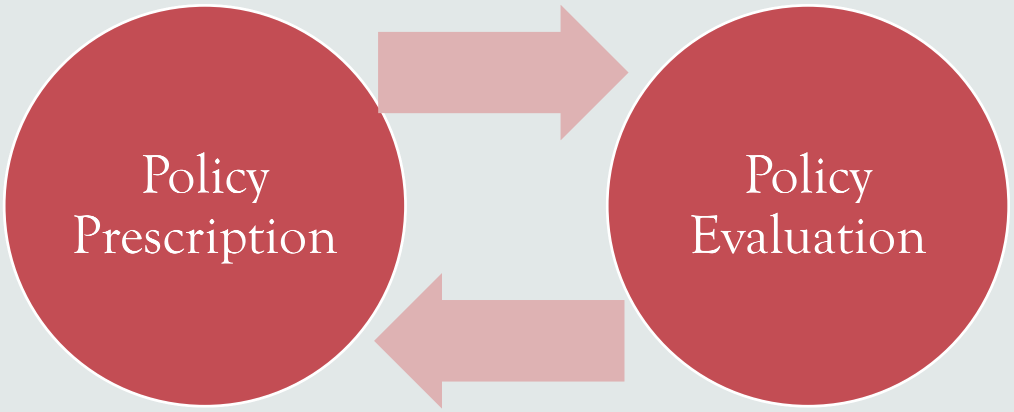
"Predicting Individual Treatment Effects: Increasing Social Efficiency of Public Policy"

Dr. Matthew B. Kidder
Newberry College

MY MOST ADMIRED POLICY ANALYST: THE OPTOMETRIST

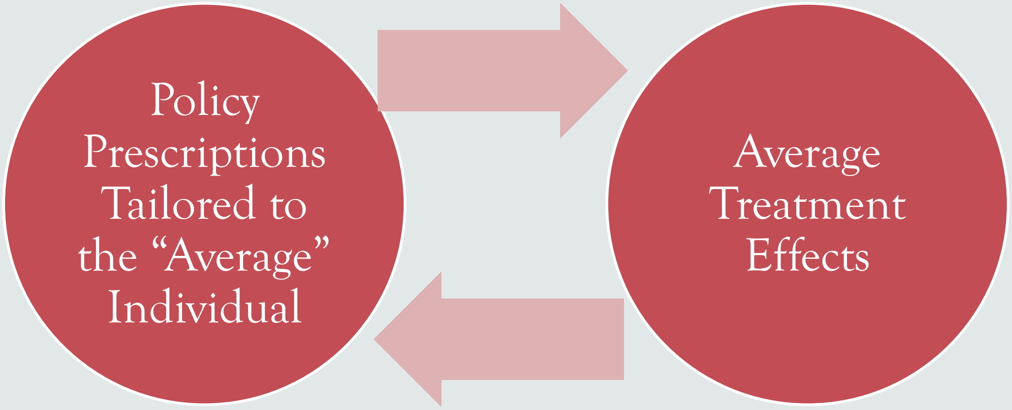
RETHINKING POLICY EFFICIENCY

- Policy is here defined, as an intervention for the purpose of moving an outcome towards a target
- Efficiency is here defined, by how close the post-policy outcome is to the targeted outcome
- Policy Analysis is measuring policy efficiency so that subsequent policy can be modified such that outcomes are closer to the target.



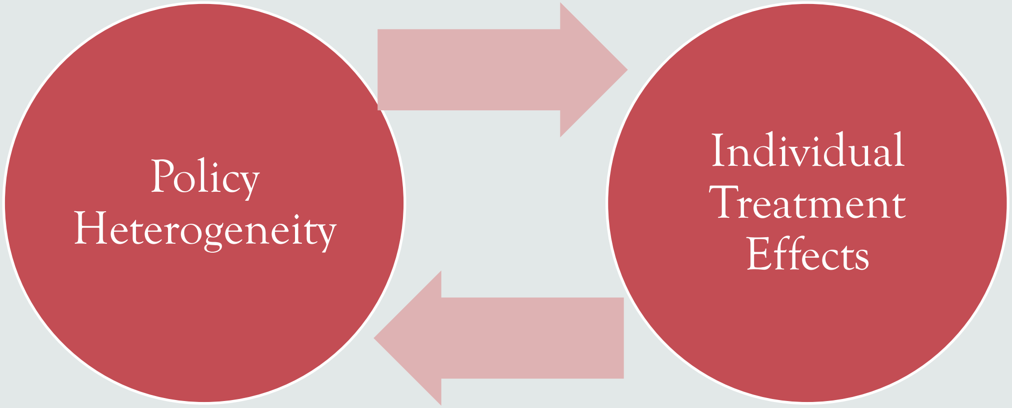
RETHINKING POLICY EFFICIENCY

- Policy is here defined, as an intervention for the purpose of moving an outcome towards a target
- Efficiency is here defined, by how close the post-policy outcome is to the targeted outcome
- Policy Analysis is measuring policy efficiency so that subsequent policy can be modified such that outcomes are closer to the target.



RETHINKING POLICY EFFICIENCY

- Policy is here defined, as an intervention for the purpose of moving an outcome towards a target
- Efficiency is here defined, by how close the post-policy outcome is to the targeted outcome
- Policy Analysis is measuring policy efficiency so that subsequent policy can be modified such that outcomes are closer to the target.



Individual Treatment Effect
Challenges in the Cross-Section

Treatment Effect Estimation

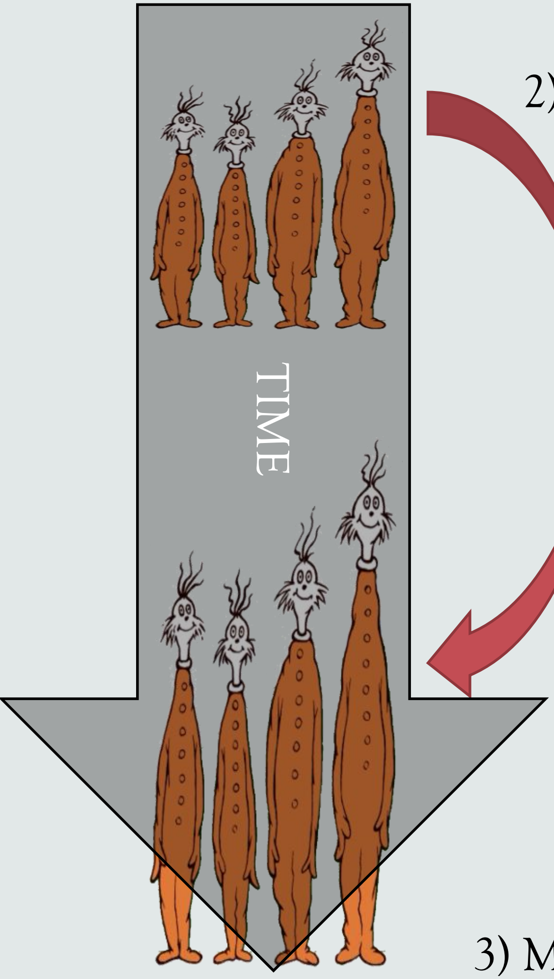
- Y= $\beta_0 + \beta_x X + \ \beta_\tau D_\tau + \epsilon$
  - β_τ *can be biased because of selection bias, confounders etc.*
  - $E[Y^1|X, D_\tau = 1] - E[Y^1|X, D_\tau = 0]$ *where superscript denotes treatment group and* D_τ *denotes treatment status*
  - $E[Y^1|X, D_\tau = 0]$ *is never observed (counter-factual) and must be predicted*
  - Classical Experiment: *Proxy* $E[Y^1|X, D_\tau = 0]$ *using* $E[Y^0|X, D_\tau = 0]$
    - Random Group Assignment makes treatment and control group statistically comparable.
  - *Similarity based methods such as PSM: Proxy* $E[Y^1|\mathbf{X} = \boldsymbol{x}, D_\tau = 0]$ *using* $E[Y^0|\mathbf{X} = \boldsymbol{x}, D_\tau = 0]$
    - Strong Ignorability: $\{(Y^1, Y^0) \perp \ T\} \mid X$
    - ***Many "ITE" models are CATE models:***
      - ***CATE Learners: X-learners, DR-learner, T-learner, R-learner***
- Weaknesses:
  - *Assumptions will always need to be asserted rather than tested*
  - *Estimating parameters is not consistent with the main objective: to predict* $E[Y^1|X, D_\tau = 0]$

THE IDEAL EXPERIMENT

A TIME MACHINE

Hypothetical Experiment (below)
to study if a drug increases a person's height

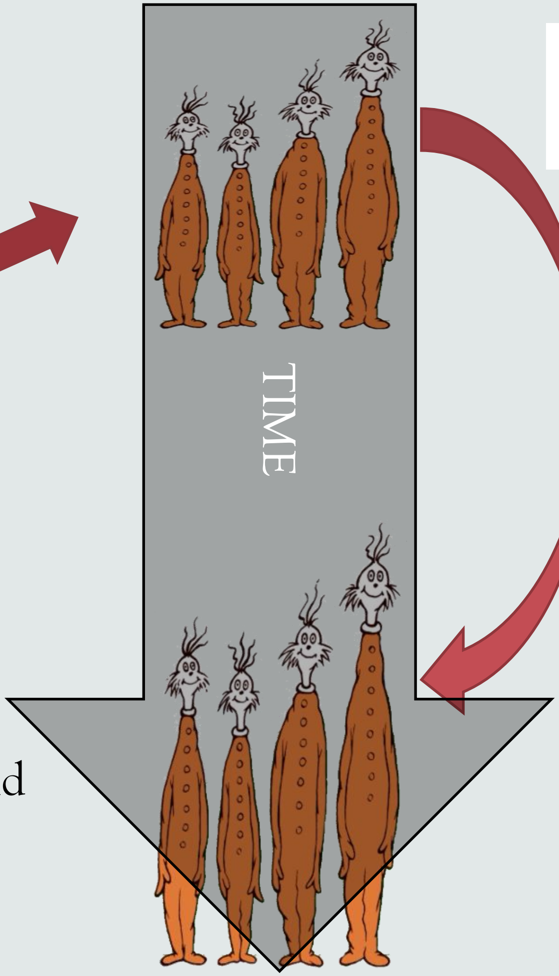
1) Measure heights



2) Give Drug

4) Hop in Time Machine and Travel back in time

5) Don't Give Drug



6) Measure heights and compare with (3) on individual basis

3) Measure heights after

LET'S BUILD A CAUSAL TIME MACHINE

THE TIME DIMENSION IS THE KEY

A CAUSAL TIME MACHINE

1. **Split Panel of *N* individuals into a collection of time series**
2. **Select one treated time series and split** into training, test, validation and post-treatment sets.
3. **Model Search and Training:** $Y_t^1 = f(Z_{i,t \in T}, t \in \text{training set})$. Search for $f(\cdot)$ and $Z_{i,t \in T}$
   - $f(\cdot)$ *can represent any ML or Statistical model and can be different for each* i *in* N.
   - $Z_{i,t \in T}$ *is* ***<u>any</u>*** *instrument that is not impacted by treatment and can be from* ***<u>any</u>*** *time period.* $Z_{i,t \in T}$ *can be different different for each* i *in* N.
4. **Predict:** $\hat{Y}_{t \geq T}^1 = E[Y_t^1 | Z_{i,t \in T}, t \geq T]$ where T is the time period when treatment begins
5. **Simple Mandate:** Make accurate predictions of $\hat{Y}_{t \geq T}^1$ on data the model has never seen! (test and validation sets)
6. ***Repeat steps 3 thru 5*** until optimal training/test error is achieved.
7. ***Repeat steps 2 thru 6*** for all N individuals that are treated and/or non-treated to study spillovers:
8. **Confidence Intervals:** $Y_t^1 \pm 1.96\ RMSFE$ Calculate RMSFE using $\epsilon_t\ \ t \in test\ set$:
9. **Individual Treatment Effect:** $\hat{\tau}_t = Y_t^1 - \hat{Y}_t^1$ if $\hat{\tau}_t$ significant and $\hat{\tau}_t = 0$ otherwise.
10. **Study:**

    $E[\hat{\tau}_{i,t} \mid X_{i,t}]$ where $X_{i,t}$ is some characteristics of the individual being treated.

    Time Dynamics of Treatment conditional on X

A CAUSAL TIME MACHINE

<u>Causal Time Machine Conditions:</u>

- **Accurate predictions of $\hat{Y}^{1}_{i,t}$ - model specific**
  - *MUST Test accuracy using data the model has never seen. Machine Learning: training, test, validation methods.*
    - Similar methods are missing validation step. Ref: HCW - Hsiao, Ching and Wan (2012) and Synthetic Controls. Leads to overfitting data which leads to inaccurate treatment estimates.
  - *Use Root Mean Squared Forecast Error to form Confidence Intervals (RMSE on the Test Set).*
- **$Z_{i,t}$ is not impacted by treatment**
  - *Does not need to be structurally related to anything. Causality trees not needed.*

MODEL AGNOSTICISM

- Any predictive modeling method that can fit a time series is suitable
- **Which model(s) makes the most accurate predictions for a specific case?**
- **Predictive Accuracy is measurable so there is no need for philosophical debates about Asymptotics**

Pure Statistical Methods

Time Series Models

- ARIMA
- VAR-SVAR
- Kalman Filter
- GMM
- ARCH
- Regression
- Etc.

Machine Learning

Neural Nets and Random Forests

- RNN
- LSTM
- Temporal Convolutional Nets
- Spatiotemporal Convolutional Nets (Chaotic Systems)
- Ensemble Methods
- Bagging and Boosting
- Etc.

CASE: US-CHINA TRADE WAR IN 2018

Policy homogeneity is a clear sign of poor understanding of treatment heterogeneity

<u>$50 Billion List</u>

- April 3rd 2018 US Announces Tariffs 1,333 Chinese Products Flat 25% increase
- June 15 US Revises List
- July 6 First Phase of June list becomes effective
- August 23 US and China impose Second Phase
- September 17-18 Tariffs become effective

<u>$200 Billion List</u>

- July 10 US Announces Additional List to be Tariffed at Flat 10%
- August 7 US Revises Second Phase to be a Flat 25% instead of a Flat 10%
- September 17-18 Tariffs become effective

POLICY QUESTION

- **Simple Objective of War:** Harm the other side more than you harm yourself.
- **In Context:** For a given product, what effect does the tariff have on US imports?
  - $Y_{i,t}^{1}$ *is the US imports of Chinese product i (HS-6 digit) at time t, for a product that is subject to a tariff (treated unit)*
  - Objective: $\hat{Y}_{t\geq T}^{1} = E[Y_t^1 | Z_{i,t\in T}, t \geq T]$
- **Choices for $Z_{i,t}$: Curse of Dimensionality turns into a Blessing**
  - *Potential* $Z_{i,t}$ *include all monthly HS-6 digit EU imports from India from 2010 to 2018*
  - *Can be expanded to include non-trade data of any potential variety not impacted by US-China trade war*
  - *We do not need the "best" model. We only need a model that performs well enough and performance is measurable*
    - A Curse becomes a Blessing

DATA

1) Pablo D Fajgelbaum, Pinelopi K Goldberg, Patrick J Kennedy, Amit K Khandelwal, The Return to Protectionism, *The Quarterly Journal of Economics*, Volume 135, Issue 1, February 2020, Pages 1–55, https://doi.org/10.1093/qje/qjz036

Second Thanks to their Research Assistants: Huifeng Chang, Jett Pettus, and Brian Pustilnik

Main Data Sets Borrowed in Primary Analysis

* HS-10 Digit Import Data from US Census (Jan 2017 – April 2019), Limited to US

* Import Tariffs from U.S. International Trade Commission documents,

* NAICS-4 to HS Crosswalk by Pierce and Schott (2012)

2) HS-6 Digit Trade Data from Comtrade (Jan 2010 – Dec 2018)

* Data covers most global trade relationships

* Extends the coverage of my US data but US Census would be better in this dimension

Instrument Selection

Iterative Selection (below)
Other methods can also work

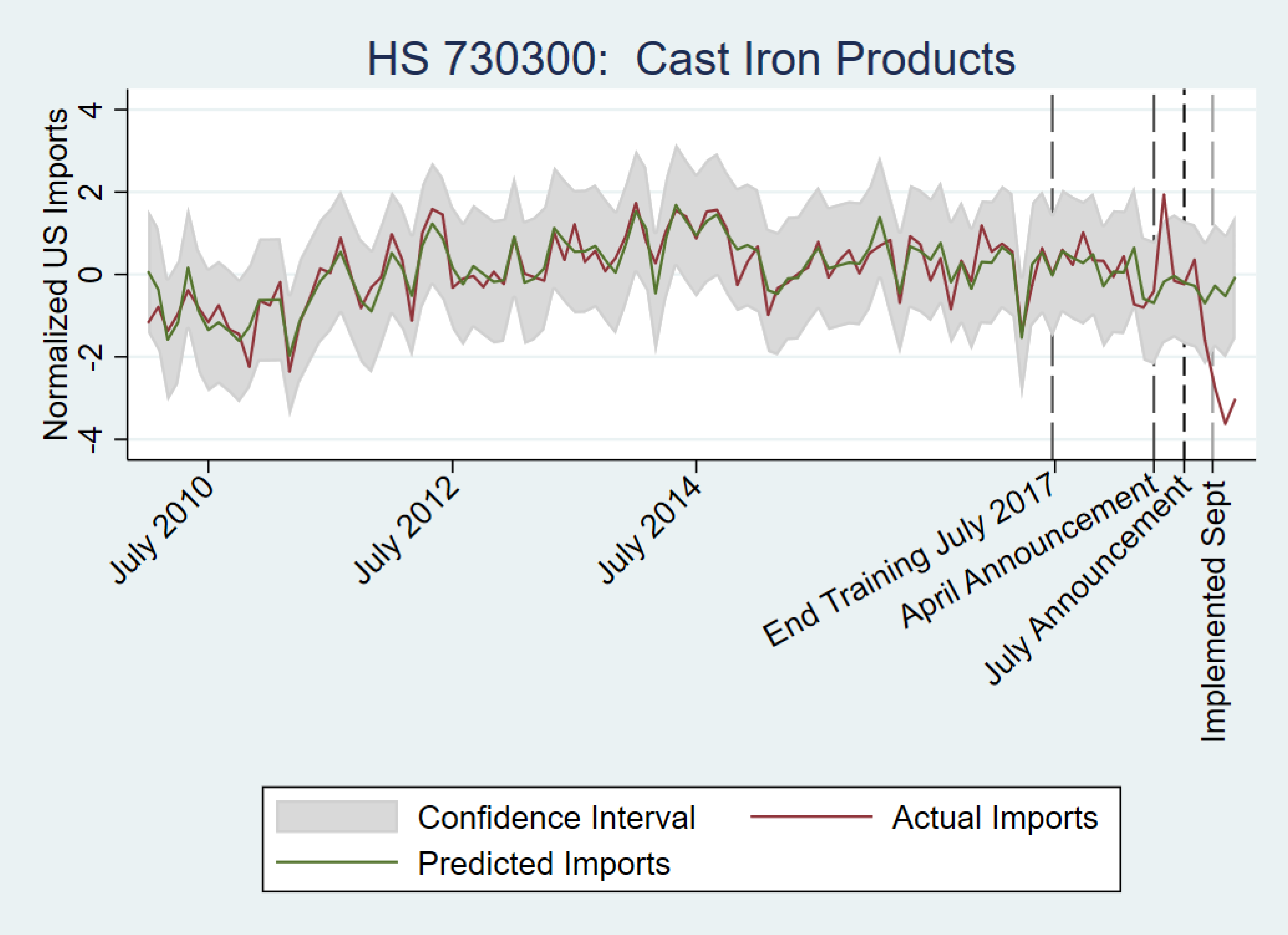
Iterative Selection: Perform the iterations below for each 1,630 HS-6 products that were “treated”

- Select the trade values for a single US Product $(i = 1)$ that was subjected to a Tariff: $Y_{i=1,t}^{treated}$
  - *Split time series into the following sets*
    - Training Set (Jan 2010 – May 2017)
    - Test Set (June 2017 – Mar 2018)
    - Treated Set (April 2018 – Dec 2018)
  - *Iteratively add features from $1{,}639$ possible HS-6 codes (EU imports of Indian Products)*
    - Estimate all 1 feature models. $Y_{i=1,t}^{treated} = \beta_0 + \beta_1 X_{N,i,t} + \epsilon_{i,t} \quad \forall \ N \in [1, 1{,}639]$
    - Calculate MSE individually for training and test sets MSE_{train} and MSE_{test}. Select k^{th} model that $\min(MSE_{train})$
    - Estimate all 2 feature models. $Y_{i=1,t}^{treated} = \beta_0 + \beta_1 X_{N=k,i,t} + \beta_2 X_{N,i,t} + \epsilon_{i,t} \quad \forall \ N \in [1, 1{,}638]$
    - Stop when MSE_{test} shows evidence of an upward trend. Save model predictions for entire sample (training, test, treated)
- Stopping Criteria is CRITICAL. Simply maximizing a fitness criteria like R^2 or BIC results in gross overfitting and poor performance out of sample.

EXPLORE PRELIMINARY RESULTS

- **Behavioral Patterns:**
  - ***Anticipatory Effects*** *– stocking up or immediately moving production*
  - ***Post Tariff Effects –*** *some are positive and some are negative*
- **Spillovers** – Effects on goods that without news of a direct tariff increase
  - Complement/Substitution effects with a good that is being tariffed
  - Risk Aversion

POLICY WINNER



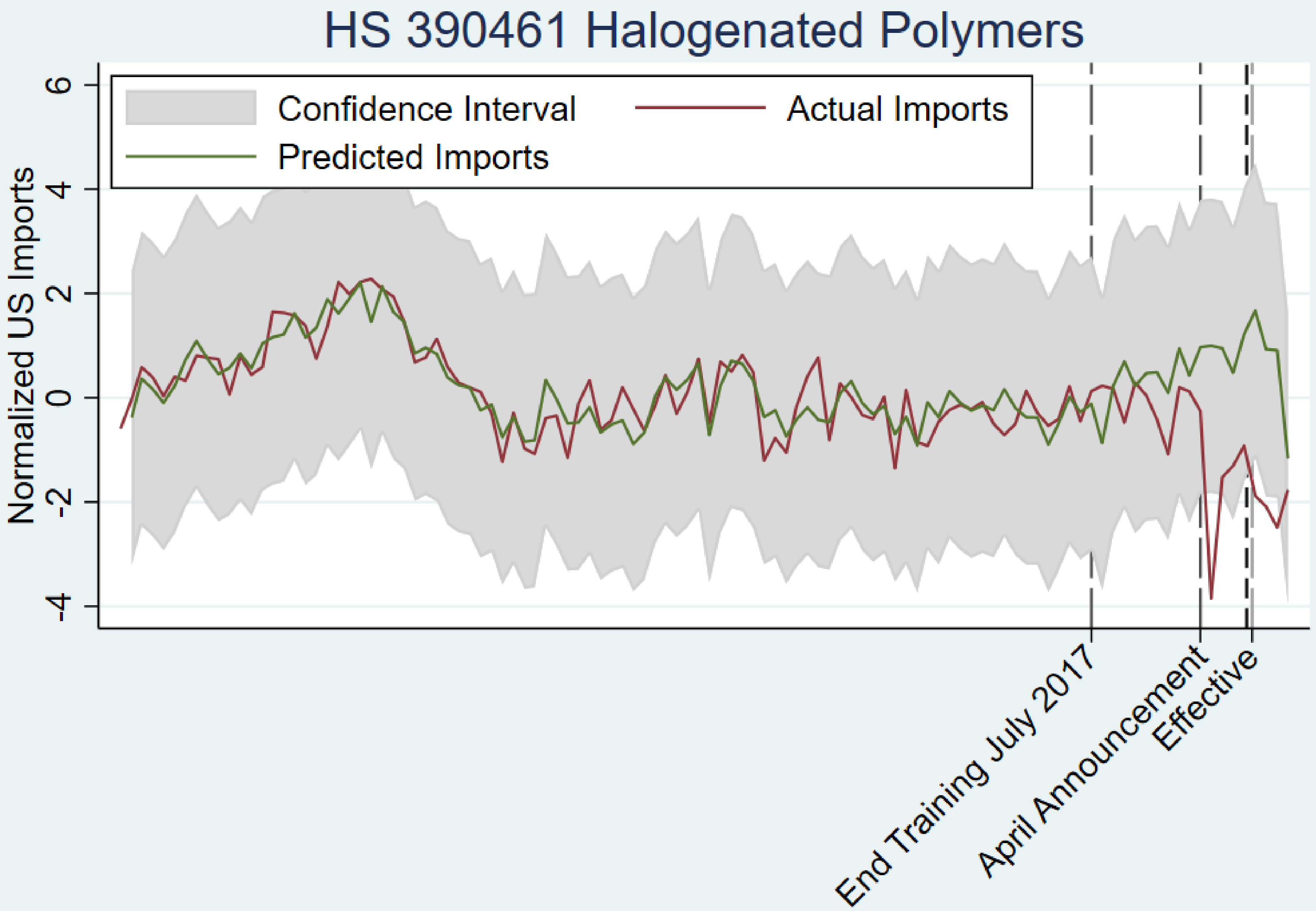
Pattern (Classic):

- Stock up after announcement. Decline after effective date.
- Both are record setting
- Predicted by theory

Notes:

- Not included in April Announcement
- Tariffs were first announced in July for this product.

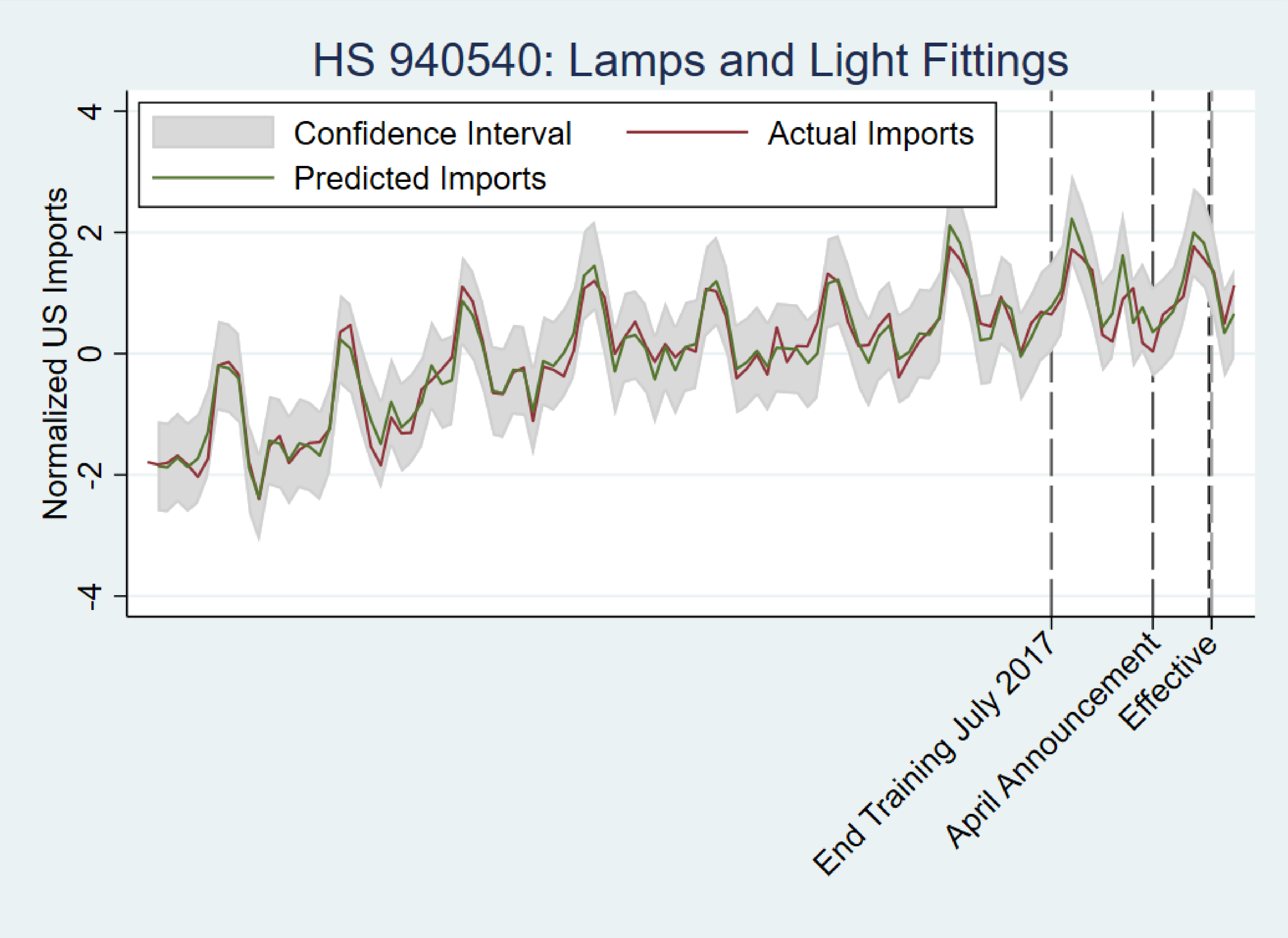
Policy Winner



Pattern (Contagion):

- Drops to historical lows after April announcement.
- Contagion driven: Tariffs on this product were first Announced August 8

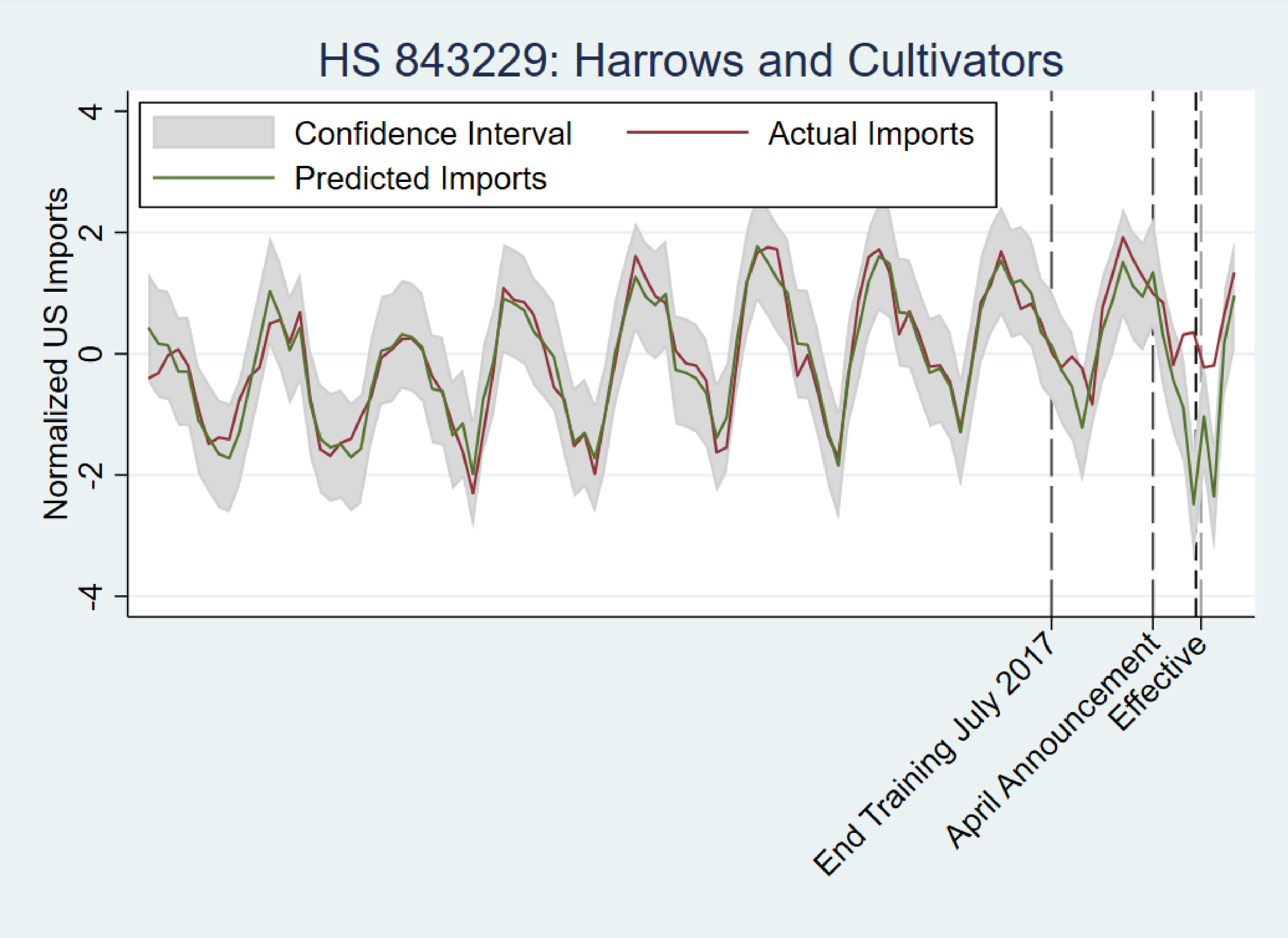
Policy Ineffective



Pattern (None):

- Tariffs did not have any impact on import quantities

POLICY LOSER



Pattern (Moderate Stock-up):

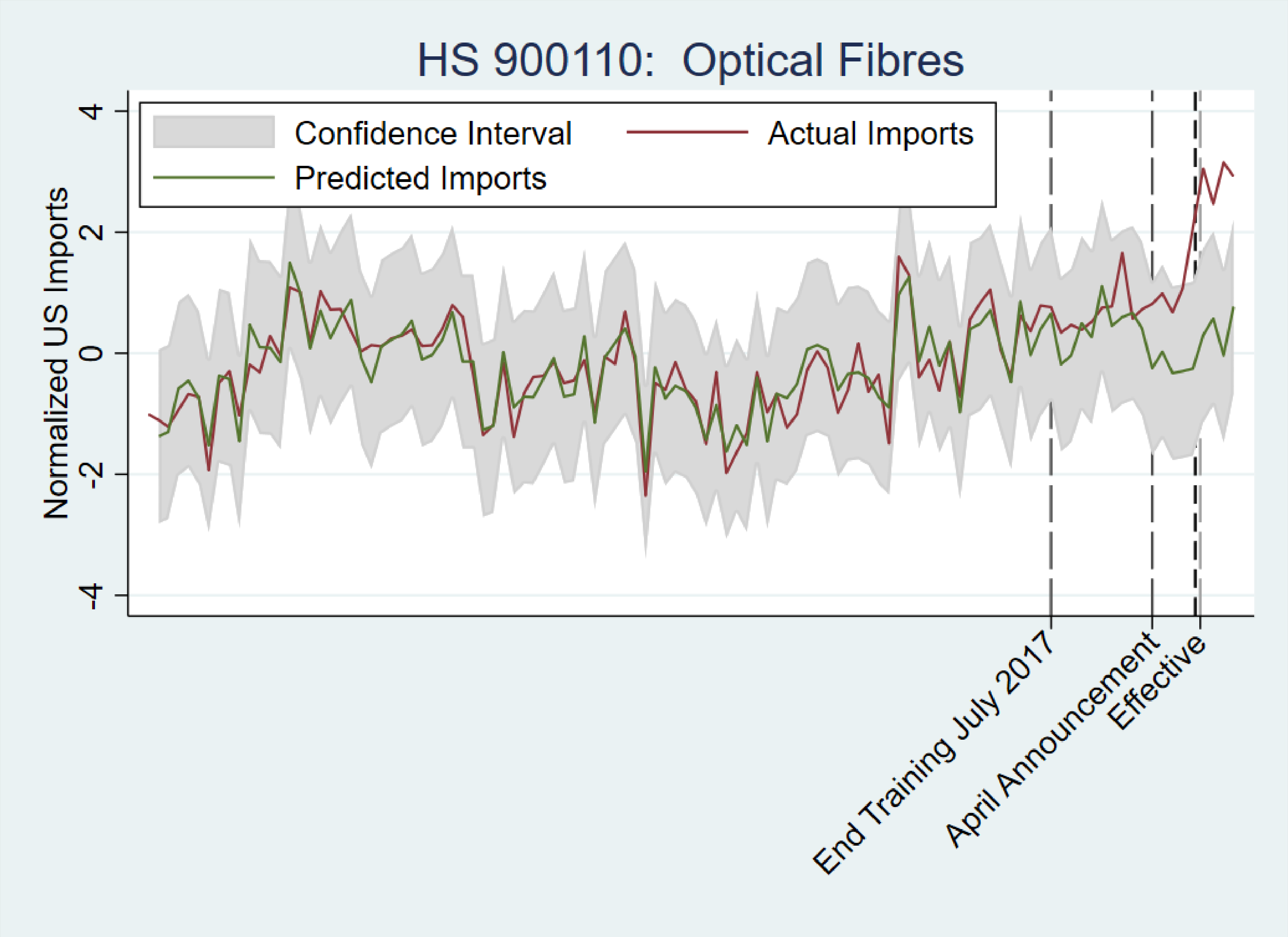
- Stock up prior to effective date without immediate drop after date

Notes:

- Not included in April Announcement

- Tariffs were first announced in July for this product.

Policy Disaster



Pattern (Prolong Stock-up):

- Imports rise in anticipation of tariff hikes and then **rise after tariffs are implemented!**
- Effect is counter policy intention and this case is thus a great example of policy inefficiency under homogeneous policy.

Notes:

- Actual Imports reach record levels after April announcement

WORKING PAPER STATUS:

- I am still sharpening the predictive model.
  - *Need more accurate predictions*
    - Current results are from a simple linear model.
    - Have more accurate results using more sophisticated Machine Learning frameworks and I am working to automate model tuning
    - Possible to add to the set of controls
  - *Poorly behaved models vs. no evidence of a treatment effect*
    - Set benchmark for predictive accuracy. Adjust models that do not meet benchmark

RESEARCH ADVANTAGES OF THE CAUSAL TIME MACHINE

CAUSAL TIME MACHINE

- **Causal Moderators:** What characteristics moderate the treatment patterns (below)? Upstreamness, Contract Intensity, Capital Intensity, Comparative Advantage, etc?
- **Behavioral Patterns:**
  - ***Anticipatory Effects*** *– stocking up or immediately moving production.*
  - ***Post Tariff Effects*** *– some are positive and some are negative*
- **Spillovers** – Effects on goods that without news of a direct tariff increase
  - Complement/Substitution effects with a good that is being tariffed
  - Risk Aversion

STANDARD METHODS

- **ATEs: I**s there a positive or negative effect on average? Analysis stops here.
- **CATE "ITEs"**
  - *Can find Causal Moderators*
  - *Lean on heroic assumptions*
  - *Farther from the "first best" experimental ideal: a time machine*
  - *Unclear how to study spillovers*
  - *Usually does not have time dynamics for treatment*

Imagine a world where the Optometrist prescribes only the average corrective lens to all patients. This is the inefficient world that we live in today. We suffer from profound policy inefficacy because we are stuck in the statistical cross-section and thus stuck with average metrics of success such as Average Treatment Effects.

Dr. Matthew B. Kidder