

# Forecast Adjustment Under Shocks: Similarity-based Solutions to Unprecedented Events

## Doctoral Defense

David Lundquist<sup>1</sup>, Daniel Eck<sup>2</sup> (advisor)

December 19th, 2024

---

<sup>1</sup>davidl11@illinois.edu

<sup>2</sup>dje13@illinois.edu

# A High-level Summary Uniting All My Projects

Forecasting under non-ideal conditions like

- 1 a rupture in the DGP
- 2 lack of past information for units we wish to forecast

What should we do?

Central technique explored: *correction of model parameters using an aggregation strategy*, where aggregation means that we make use of external data.

We began with particular models; then built upwards, establishing an abstract, general framework for correcting forecasts that is widely applicable for forecasting and beyond.

# Why is this valuable?

We are weighing-in on a puzzle:

conflict between the intuitive notion that more relevant information should help in forecasting, and the hard reality that attempts to make it do so have not been uniformly successful (Clements and Hendry [2005](#))

## Why is this valuable?

We provide a precise proposal to a well-specified challenge:

incomplete information by itself is unlikely to play a key role in forecast failure (except if that information would forecast breaks). Consequently, using large amounts of data may not correct one of the main problems confronting forecasters, namely location shifts, unless that additional information is directly pertinent to forecasting breaks (Castle, Clements, and Hendry [2013](#))

## Why is this valuable?

- Canonical problems: predicting based on a limited feature space and/or small  $n$
- The framework we've offered is something others can 'plug-in' to easily and add to.
- Software developed

# Questions guiding these slides (remove later)

- 1 Whatcha learned? **Forecasting is difficult; Heterogeneity of DGP**
- 2 Why did you choose this topic? **Although it seems narrow, the question of “what if you had a new unit and had very little past information to go on” is a common phenomenon in statistics.** Examples: Shock to a time series Unscheduled scheduled  
A new seller joins amazon’s platform, i.e. cold starts (Fatemi et al. [2023](#))
- 3 How would you improve your work? **Many ideas here. There were more directions than I could pursue.**
- 4 What is the importance of your study or how will it contribute or add up to the existing body of knowledge? **Two entirely separate perspectives: (1) Post-shock forecasting is a novel research framework. (2) Post-shock forecasting builds on intercept corrections and other canonical questions.**
- 5 What type of background research have you done for the study? **In no particular order: synthetic control; convex geometry; convex optimization; p-value combination; FWER/FDR; fixed/mixed/random effects; panel data; linear time series models; RNN; econometrics**

# Questions guiding these slides (remove later)

- 1 What are the limitations you have faced? A ton of hyperparameters
- 2 What more to add? LLM-generated donors and covariates
- 3 What are the recommendations of your study? Relatively easy to answer: for an unprecedented event, locate it in the space of previous events.
- 4 What was your hypothesis and how did you frame it? Signal to noise
- 5 If given a chance, would like to do something different with your work?
- 6 What are the limitations you faced? Realized volatility is something that can be estimated with HF data
- 7 How did you relate your study to the existing theories?
- 8 What is the future scope of this study?
- 9 How did you evaluate your work? Simulations and real data examples
- 10 How would you improve your work? By design, synthetic control cannot extrapolate. In the causal inference context, that may very well be a virtue. However, in the prediction context, it may hinder us. Note that random forest has the same problem.

A seemingly unprecedented event might make one ask

- 1 What does it resemble from the past?



# A seemingly unprecedented event might make one ask

- 1 What does it resemble from the past?
- 2 What past events are most relevant for our objectives?

# A seemingly unprecedented event might make one ask

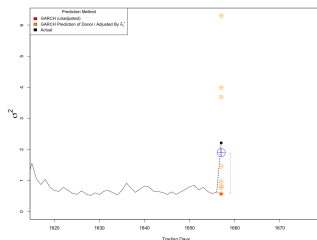
- 1 What does it resemble from the past?
- 2 What past events are most relevant for our objectives?
- 3 Can we incorporate past events in a systematic, principled manner?

## Punchline of the paper

Credible forecasting is possible under news shocks, so long as we incorporate external information to account for the **nonzero errors**.

# Punchline of the paper

Credible forecasting is possible under news shocks, so long as we incorporate external information to account for the **nonzero errors**.



**Figure:** Adjusting our One-Step-Ahead Forecast Using Only Arithmetic Mean of Donors

# Outline

- 1 Introduction
- 2 The Idea and Methodology
- 3 Formal Results
- 4 Applications
- 5 Software and LLMs
- 6 How can we trust this?
- 7 Future directions for Forecasting Amid Shocks
- 8 Directions and Limitations

Premise: There is a breaking news at some fractional lag  $T-\epsilon$

- After-hours trading provides a poor forum in which to digest news
- News constitutes public, material information for one or more traded assets
- The **qualitative aspects** of the news provide a basis upon which to
  - match to past news shocks
  - match in a  $p$ -dimensional covariate space

# Model Preliminaries

Let  $I(\cdot)$  be an indicator function.

Let  $T_i$  denote the time length of the time series  $i$  for  $i = 1, \dots, n + 1$ .

Let  $T_i^*$  denote the largest time index prior to news shock, with  $T_i^* < T_i$  (i.e. we assume at least one post-shock observation).

Let  $\delta, \mathbf{v}_{i,t} \in \mathbb{R}^p, \mathbf{x}_{i,t} \in \mathbb{R}^d$ .

# Model Setup

For  $t = 1, \dots, T_i$  and  $i = 1, \dots, n + 1$ , the model  $\mathcal{M}_1$  is defined as

$$\begin{aligned}\sigma_{i,t}^2 &= \omega_i + \omega_{i,t}^* + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \\ \mathcal{M}_1: \quad a_{i,t} &= \sigma_{i,t}((1 - D_{i,t}^{\text{return}})\epsilon_{i,t} + D_{i,t}^{\text{return}}\epsilon_i^*) \\ \omega_{i,t}^* &= D_{i,t}^{\text{vol}}[\mu_{\omega^*} + \delta' \mathbf{v}_{i,t} + u_{i,t}],\end{aligned}$$

with error structure

$$\begin{aligned}\epsilon_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_\epsilon \text{ with } \mathbb{E}_{\mathcal{F}_\epsilon}(\epsilon) = 0, \text{Var}_{\mathcal{F}_\epsilon}(\epsilon) = 1 \\ \epsilon_{i,t}^* &\stackrel{iid}{\sim} \mathcal{F}_{\epsilon^*} \text{ with } \mathbb{E}_{\mathcal{F}_{\epsilon^*}}(\epsilon) = \mu_{\epsilon^*}, \text{Var}_{\mathcal{F}_{\epsilon^*}}(\epsilon^*) = \sigma_{\epsilon^*}^2 \\ u_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_u \text{ with } \mathbb{E}_{\mathcal{F}_u}(u) = 0, \text{Var}_{\mathcal{F}_u}(u) = \sigma_u^2 \\ \epsilon_{i,t} &\perp\!\!\!\perp \epsilon_{i,t}^* \perp\!\!\!\perp u_{i,t}\end{aligned}$$

where  $D_{i,t}^{\text{return}} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,\text{return}}\})$  and  $D_{i,t}^{\text{vol}} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,\text{vol}}\})$  and  $L_{i,\text{return}}, L_{i,\text{vol}}$  denote lengths of log return and volatility shocks, respectively.

Note: we will be looking GARCH(1,1) only in this presentation.



# Significance of the Covariates

Covariates chosen for inclusion may be any  $\mathcal{F}_t$ -measurable function, for example

- levels
- differences in levels
- log returns
- percentage returns
- measurable transformations of the above

Key criterion for inclusion: how plausible is the covariate as a **proxy for risk conditions** for the volatility series to be forecasted?

# Forecasting

We present two forecasts:

$$\text{Forecast 1: } \hat{\sigma}_{unadjusted}^2 = \hat{\mathbb{E}}[\sigma_{\mathbf{1}, T_{\mathbf{1}}^*+1}^2 | \mathcal{F}_{T^*}] = \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \sigma_{i,t-j}^2 + \hat{\gamma}_i^T \mathbf{x}_{i,t}$$

$$\text{Forecast 2: } \hat{\sigma}_{adjusted}^2 = \hat{\mathbb{E}}[\sigma_{\mathbf{1}, T_{\mathbf{1}}^*+1}^2 | \mathcal{F}_{T^*}] + \hat{\omega}^* = \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \sigma_{i,t-j}^2 + \hat{\gamma}_i^T \mathbf{x}_{i,t} + \hat{\omega}^* .$$

# Distance-based Weighting in Action

- Observe the pair  $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ .

# Distance-based Weighting in Action

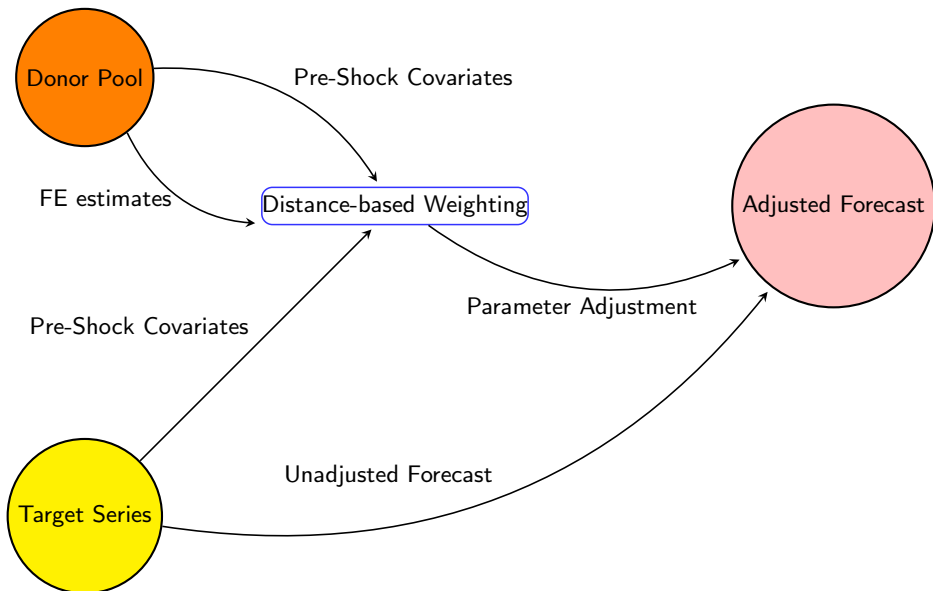
- Observe the pair  $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ .
- Goal: recover weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$  and compute  $\hat{\omega}^* := \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ , our forecast adjustment term.

# Distance-based Weighting in Action

- Observe the pair  $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ .
- Goal: recover weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$  and compute  $\hat{\omega}^* := \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ , our forecast adjustment term.
- Following Abadie and Gardeazabal [2003](#), Abadie, Diamond, and Hainmueller [2010](#), let  $\|\cdot\|_S$  denote any semi-norm on  $\mathbb{R}^P$ , and define

$$\{\pi\}_{i=2}^{n+1} = \arg \min_{\pi} \|\mathbf{v}_{1,T^*} - \mathbf{V}_{T^*} \pi\|_S .$$

## Visuals That Tell The Story



# Visuals That Tell The Story

## The forecaster's decision tree

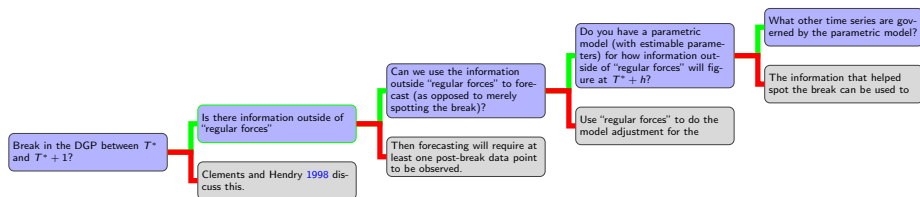
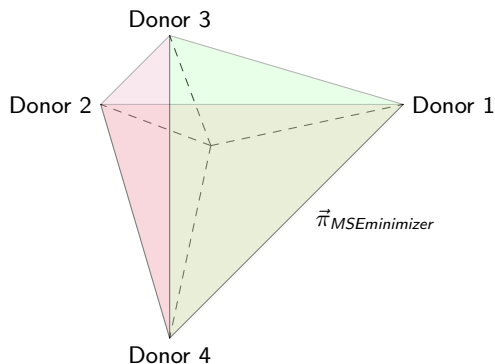


Figure: Forecast Model Adjustment: A Decision Tree

# Visuals That Tell The Story



**Figure:** The 3-Simplex,  $\Delta^3$ , where hypothetical minimizer is a convex combination of Donors 1 and 4.



# Global Overview

- 1 **Object-to-predict** Most fundamentally, the method requires a random object (indexed over time and possibly space, as well) that obeys a specification with additive errors, or, at the very least, a specification that can be transformed to have additive errors.
- 2 **Common Model Family on the Shocks** The method requires that residuals be governed by a model that is shared across all units. This ensures that in the estimation of news shocks in the donor pool, the estimators will enjoy similar properties that will produce a good aggregated shock estimator. This condition is satisfied by the parametric shock distributions found in Lin and Eck 2021; Lundquist and Eck 2024.
- 3 **Reliable and Shared Model-Fitting Procedure** There must exist a reliable model-fitting procedure for the  $n + 1$  units.
- 4 **Reliable Correction Term Estimation**
- 5 **Reliable Correction Function Estimation** There must exist a correction function (presumably based on the correction term) that maps data from the donor pool to the *predicted* correction term in the time series under study based on some notion of similarity. In some cases, there may exist a posited DGP that the correction term estimates. In other cases, there may be no posited DGP.

# Formal Results

## Proposition

Assume

- 1 For each  $i$ ,  $1 \leq i \leq n+1$ , let  $\{y_t\}_{i=1}^{T_i}$  follow an  $AR(p)$ -X as laid out in Section ??.
- 2 Assume for each  $i$ ,  $1 \leq i \leq n+1$ , the shocks  $\alpha_{i,t}$  are uncorrelated across donors.
- 3 Assume for each  $i$ ,  $1 \leq i \leq n+1$ , the shocks  $\alpha_{i,t}$  are uncorrelated with  $\alpha_{i,T_i^*+1}$

Then the tuple of estimators  $(\hat{p}_{i,1}, \dots, \hat{p}_{i,p}, \hat{\alpha}_{i,T_i^*+1})$  is consistent as  $t \rightarrow \infty$ .

# Formal Results

## Proposition

Assume

- ❶ All conditions listed in Proposition 1.
- ❷ There exist weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^{n-1}$  such that  $\mathbf{v}_{1,T_1^*} = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T_i^*}$ .
- ❸ For all  $i$ , the  $\{u_{i,t}\}$  are equal in distribution.

Then the aggregated estimator  $\alpha_{T_1^*+1}^{\hat{}}$  converges in distribution to  $\alpha_{T_1^*+1}$  as  $t \rightarrow \infty$ . Furthermore, if the  $\{u_{i,t}\}$  are constant with probability 1, the convergence is in probability.

# Formal Results

## Proposition

Let  $\{\hat{y}_{1, T_1^*+r}\}_{r=1}^h$  denote the vector of adjusted predictions (adjusted through  $h$  steps ahead) in the time series under study. Assume all conditions listed in Propositions 1 and 2. Then  $\{\hat{y}_{1, T_1^*+r}\}_{r=1}^h \xrightarrow{d} \{y_{1, T_1^*+r}\}_{r=1}^h$ . Furthermore, if the  $\{u_{i,t}\}$  are constant with probability 1, the convergence is in probability.

# My Prelim: An Incredibly Brief Recap

- Paper is under review at the *International Journal of Forecasting*

## My Prelim: An Incredibly Brief Recap

- Paper is under review at the *International Journal of Forecasting*
- Event-driven investing strategies (unscheduled news shock)

# My Prelim: An Incredibly Brief Recap

- Paper is under review at the *International Journal of Forecasting*
- Event-driven investing strategies (unscheduled news shock)
- Scheduled macroeconomic news possibly pre-empted by a news leak

ECONOMY

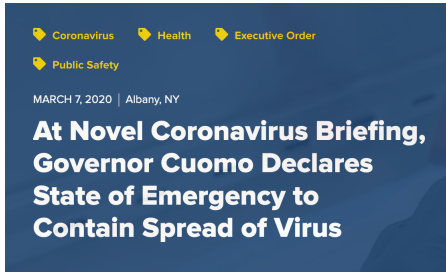
## Fed Likely to Consider 0.75-Percentage-Point Rate Rise This Week

Officials had signaled plans to raise interest rates in half-point increments before recent deterioration in data

By Nick Timiraos [Follow](#)

Updated June 13, 2022 7:47 pm ET

Example (Weekend of March 6th - 8th, 2020)



# Oil nose-dives as Saudi Arabia and Russia set off 'scorched earth' price war

PUBLISHED SUN, MAR 8 2020-9:01 AM EDT | UPDATED MON, MAR 9 2020-5:33 PM EDT

## Oil crashes by most since 1991 as Saudi Arabia launches price war



By [Matt Egan](#), CNN Business

🕒 3 minute read · Updated 3:21 PM EDT, Mon March 9, 2020



## Why apply our method to the 2016 US Election?

- You can win the US Presidency without a majority.
- No incumbent candidate
- Donald J. Trump espoused unorthodox, populist positions on healthcare, trade, foreign policy
- Donald J. Trump had no record to assess or criticize
- It was not predicted – hence it delivered news.

IYG

**iShares U.S. Financial Services ETF**

Figure: IYG includes JPM, BAC, WF, CITI, among other financial majors

- 1 **Model choice** GARCH(1,1) on the daily log return series of IYG in each donor
- 2 **Covariate Choice**
  - previous 30 log returns of IYG (large pre-treatment period, in the language of SC)
  - log return Crude Oil (CL.F)
  - VIX
  - log return of the VIX
  - log returns of the 3-month, 5-year, 10-year, and 30-year US Treasuries
  - return of the most recently available monthly spread between AAA and BAA corporate debt
  - log return in the trading volume of the ETF IYG itself
- 3 **Donor pool construction** US Elections from 2004, 2008, 2012
- 4 **Choice of estimator for volatility** Sum of 77 squared five-minute returns generated between 9:35am and 4pm on November 9th, 2016.

# 2016 Election

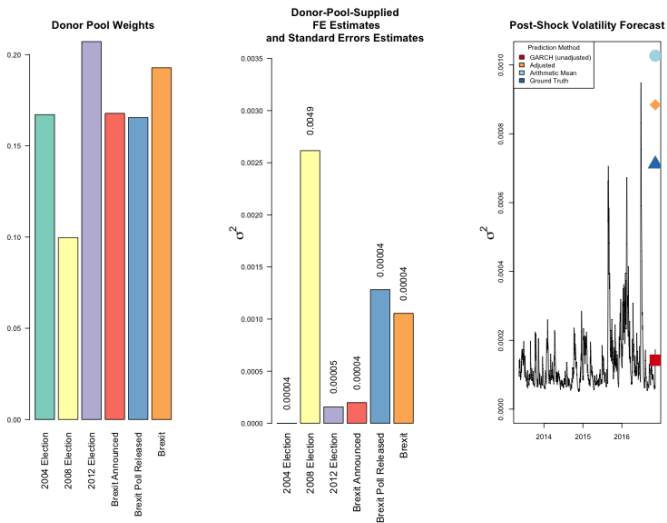


Figure: The volatility induced by the 2016 US election

# The Post-prelim Direction of my research

- Extending similarity-based parameter correction to the state-of-the-art HAR model
- Extending similarity-based parameter correction to non-linear shock models
- Building out a general framework for parameter correction

# What we will discuss in this section

- 1 Role of outside information
- 2 The Meaning and Use of Similarity

# Simulations: Parameter Correction Using An Aggregated Decay Parameter from Donors

Here we will include the exponential shock simulations.

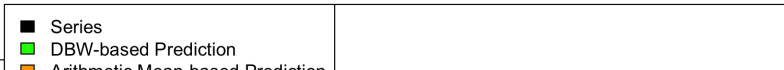
Most elementary simulation regime tests Hypothesis 1 and 2 by varying  $\delta$  and  $u_{i,t}$ .

Recall an  $\mathcal{M}_1$  model on the volatility, which is characterized by an exogenous shock to the volatility equation generated by an affine function of the covariates:

$$\begin{aligned}\sigma_{i,t}^2 &= \omega_i + \omega_{i,t}^* + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \\ a_{i,t} &= \sigma_{i,t}((1 - D_{i,t}^{\text{return}})\epsilon_{i,t} + D_{i,t}^{\text{return}}\epsilon_i^*) \\ \mathcal{M}_1: \quad \omega_{i,t}^* &= D_{i,t}^{\text{vol}}[\mu_{\omega^*} + \delta' \mathbf{v}_{i,t} + u_{i,t}] \\ D_{i,t}^{\text{return}} &\equiv 0\end{aligned}$$

We now investigate the exponential decay model introduced just above. We model the decay parameter  $\psi$  as a dot product of donor-specific and donor-invariant quantities.

## Predicting Exponential Shocks Using Distance-Based Weighting



# Software

R package

## LLMs

## Use NLP to identify donors.

## You

What are some events that are similar to Chinese government's ban on iPhones for government officials in September 2023?

## ChatGPT

The Chinese government's ban on iPhones for government officials in September 2023 is a notable example of government-led restrictions or bans on certain technology products or services due to security concerns or political reasons. Here are some events or actions that are somewhat similar:

1. **Huawei Ban in the United States:** The United States government banned Huawei, a Chinese telecommunications company, from selling its equipment and infrastructure in the U.S. due to concerns about national security and espionage. This ban was initiated in 2019 and has had significant repercussions for Huawei's business globally.
2. **Russian Ban on Telegram:** In 2018, the Russian government banned Telegram, a popular messaging app, due to the company's refusal to provide encryption keys to Russian security agencies. The ban resulted in the disruption of Telegram services within Russia and sparked debates about privacy and government surveillance.
3. **India's Ban on Chinese Apps:** In 2020, the Indian government banned several Chinese mobile apps, including TikTok and WeChat, citing concerns about data privacy and national security after a border clash between Indian and Chinese troops. The ban affected millions of users and led to diplomatic tensions between India and China.
4. **Turkey's Wikipedia Ban:** In 2017, the Turkish government banned access to Wikipedia, alleging that the website was involved in a smear campaign against Turkey. The ban stemmed from Wikipedia's refusal to remove content that the Turkish government deemed offensive. Despite international criticism, the ban remained in place for over two years before being lifted in 2020.
5. **Iran's Ban on Messaging Apps:** The Iranian government has intermittently banned messaging apps such as Telegram and WhatsApp over concerns about the spread of misinformation, political dissent, and national security threats. These bans have been part of broader efforts to control online communication and social media platforms within Iran.

These examples demonstrate how governments may impose bans or restrictions on technology products, services, or platforms for various reasons, including national security, political control, and regulatory compliance.





# Robustness of the Approach

What assurances do we have that the method will be good?

We shall group the extensions into five buckets:

- How much can we automate?
- Alternatives for fixed effect estimation
- Alternative estimators and estimands
- What can you do with a volatility forecast?
- Where else is distanced-based weighting useful?
- Can we extend the results of (Bodilsen and Lunde [2023](#))

# How much can we automate?

What if the covariates are difficult to specify?

Proposed solution:

Use shrinkage estimation to detect fleeting signals in the cross section of  $a_t^2$  (Chinco, Clark-Joseph, and Ye [2019](#)).

# Limitations of what we're currently doing

- Our real data examples cannot be scaled up due to the need to for human involvement in donor and covariate curation
- We assume that the exogenous variables (distinct from the covariates) are known at time  $T_i^*$

# New Frontiers in Distance-based Weighting

- Integrate lessons from literature on under/over reactions to information shocks (Jiang and Zhu [2017](#))
- Distance-based Weighting of Impulse Response Functions

## Distance-based Weighting of Impulse Response Functions

Suppose

- We have a collection of  $p$ -variate time series of lengths  $T_i, i = 1, 2, \dots, n + 1$ .
- We are interested in the response of variable  $r$  to shocks in variable  $j$ ,  $1 \leq r \leq j \leq p$ .

There many ways to estimate  $IRF_1(r, j)$ .

Can we somehow aggregate the estimates  $\widehat{IRF}_i(r, j), i = 2, 3, \dots, n + 1$ ?

Additional research questions:

- What DGP would best motivate/justify such a method?
- Which method of IRF estimation would perform best?

## References I

-  Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505.
-  Abadie, Alberto and Javier Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country". In: *American Economic Review* 93.1, pp. 113–132.
-  Bodilsen, Simon Tranberg and Asger Lunde (2023). "Exploiting news analytics for volatility forecasting". In: *Available at SSRN* 4401032.
-  Castle, Jennifer L, Michael P Clements, and David F Hendry (2013). "Forecasting by factors, by variables, by both or neither?" In: *Journal of Econometrics* 177.2, pp. 305–319.
-  Chincó, Alex, Adam D Clark-Joseph, and Mao Ye (2019). "Sparse signals in the cross-section of returns". In: *The Journal of Finance* 74.1, pp. 449–492.
-  Clements, Michael and David F Hendry (1998). *Forecasting economic time series*. Cambridge University Press.

## References II

-  Clements, Michael P and David F Hendry (2005). "Guest Editors' introduction: Information in economic forecasting". In: *Oxford Bulletin of Economics & Statistics* 67.S1, pp. 713–753.
-  Fatemi, Zahra et al. (2023). "Mitigating cold-start forecasting using cold causal demand forecasting model". In: *arXiv preprint arXiv:2306.09261*.
-  Jiang, George J and Kevin X Zhu (2017). "Information shocks and short-term market underreaction". In: *Journal of Financial Economics* 124.1, pp. 43–64.
-  Lin, Jilei and Daniel J Eck (2021). "Minimizing post-shock forecasting error through aggregation of outside information". In: *International Journal of Forecasting*.
-  Lundquist, David and Daniel Eck (2024). "Volatility Forecasting Using Similarity-based Parameter Correction and Aggregated Shock Information". In: *arXiv preprint arXiv:2406.08738*.