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

Doctoral Defense

David Lundquist[1], Daniel Eck[2] (advisor)

December 18th, 2024

[1] davidl11@ilinois.edu
[2] 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, e.g. a so-called “cold start" (Fatemi et al. 2023)

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

Began with particular models; then built upwards, establishing an abstract, general framework for correcting forecasts that is widely applicable across 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 offer is something others can 'plug-in' to and add to.
- Software developed

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?

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

Coronavirus Health Executive Order

Public Safety

MARCH 7, 2020 | Albany, NY

At Novel Coronavirus Briefing, Governor Cuomo Declares State of Emergency to Contain Spread of Virus

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

Punchline

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

Outline

Premise: There is breaking news at some fractional lag t-ϵ

- Market (i.e. crowd-sourced) mechanisms for evaluating the news are either offline, too thin, or otherwise unavailable. They could also be simply unreliable.
- 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 Setup

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

$$\mathcal{M}_1 : \begin{aligned} y_{i,t} &= F(\mathcal{F}_{i,t-1}) + \alpha_{i,t} + \epsilon_{i,t} \\ \alpha_{i,t} &= \mathbf{x}_{i,t}^T \lambda_{i,t} \\ \mathbf{x}_{i,t}^T &= (1, x_{i,t}^1, \ldots, x_{i,t}^p) \text{ (observable and deterministic w.r.t to } \mathcal{F}_{i,t-1}) \\ \lambda_{i,t}^T &= (u_{i,t}, \lambda_t^1, \ldots, \lambda_t^p) \text{ (unobservable and potentially random)}, \end{aligned} \tag{1}$$

$$\lambda_{i,t} \sim \mathcal{F}_\lambda \text{ with } \mathbb{E}_{\mathcal{F}_\lambda}(\lambda) = \mu_{\lambda_t}, \mathrm{Var}_{\mathcal{F}_\lambda}(\lambda) = \boldsymbol{\Sigma}_{\lambda_t},$$

and time-invariant error structure

$$\epsilon_{i,t} \overset{iid}{\sim} \mathcal{F}_\epsilon \text{ with } \mathbb{E}_{\mathcal{F}_\epsilon}(\epsilon) = 0, \mathrm{Var}_{\mathcal{F}_\epsilon}(\epsilon) = \sigma_\epsilon^2,$$

Model Details

Note that $\lambda_{i,t}^{T} = (u_{i,t}, \lambda_t^1, ..., \lambda_t^p)$

- includes the term $u_{i,t}$ that is not shared among donors.
- is time-varying allows us to capture that most time points are without news shocks (in which case $\alpha_{i,t}$ is of negligible effect), but conditional upon information arriving between T_1^* and $T_1^* + 1$, we may know with near certainty that λ_{1,T_1^*+h} will be nonzero in norm for some $h > 0$.

Significance of the Covariates $\mathbf{x}_{i,t}$

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 time series under study?

Forecasting

We now present two one-step-ahead forecasts. First is the unadjusted forecast. The second is the adjusted forecast, which differs by the predicted correction term:

$$\text{Forecast 1: } \hat{y}_{unadjusted,T_1^*+1} = \hat{\mathbb{E}}[y_{1,T_1^*+1} | \mathcal{F}_{T_1^*}]$$

$$\text{Forecast 2: } \hat{y}_{adjusted,T_1^*+1} = \hat{\mathbb{E}}[y_{1,T_1^*+1} | \mathcal{F}_{T_1^*}] + \hat{\alpha}_{T_1^*+1} \ .$$

Distance-based Weighting in Action

(For ease of exposition, we omit time indices)

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

Distance-based Weighting in Action

(For ease of exposition, we omit time indices)

- Observe the pair $(\{\hat{\alpha}_i\}_{i=2}^{n+1}, \{\mathbf{x}_i\}_{i=2}^{n+1})$.
- Goal: recover weights $\{\pi_i\}_{i=2}^{n+1} \in \Delta^{n-1}$ and compute

$$\hat{\alpha}_1 := \sum_{i=2}^{n+1} \pi_i \hat{\alpha}_i,$$

our predicted correction term.

Distance-based Weighting in Action

(For ease of exposition, we omit time indices)

- Observe the pair $(\{\hat{\alpha}_i\}_{i=2}^{n+1}, \{\mathbf{x}_i\}_{i=2}^{n+1})$.
- Goal: recover weights $\{\pi_i\}_{i=2}^{n+1} \in \Delta^{n-1}$ and compute

$$\hat{\alpha}_1 := \sum_{i=2}^{n+1} \pi_i \hat{\alpha}_i,$$

  our predicted correction term.
- Following Abadie, Diamond, and Hainmueller 2010; Abadie and Gardeazabal 2003, let $\|\cdot\|_{\mathbf{S}}$ denote any semi-norm on $\mathbb{R}^p$, and define

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

Visuals That Tell The Story

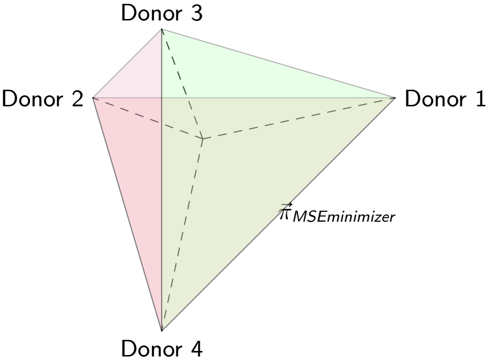
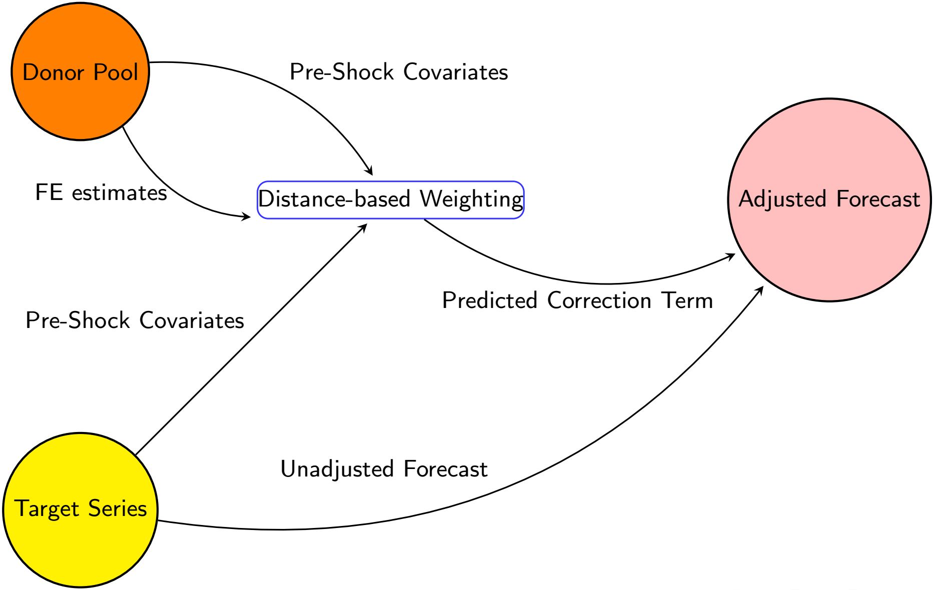


Figure: The 3-Simplex, Δ^3, where hypothetical minimizer is a convex combination of Donors 1 and 4.

Visuals That Tell The Story



Visuals That Tell The Story

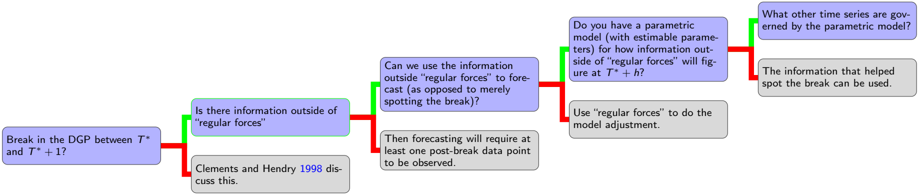


Figure: Forecast Model Adjustment: A Decision Tree

Two slides about theory

Table: A 2x2 Schema of Forecast Information, With Examples

	Conventional Econometric Models	Outside Conventional Econometric Models
internal	Lags of the series itself; past shocks	Polynomial expansion of the feature space and other transformations without solid theoretical motivation
external	Macro variables like interest rates, commodity prices; weather-related variables	Google Trends, high-frequency data like prediction markets; past shocks under similar conditions

Two slides about theory

Table: A 2x2 Schema of Forecast Information, With Examples

	Conventional Econometric Models	Outside Conventional Econometric Models
internal	Lags of the series itself; past shocks	Polynomial expansion of the feature space and other transformations without solid theoretical motivation
external	Macro variables like interest rates, commodity prices; weather-related variables	Google Trends, high-frequency data like prediction markets; past shocks under similar conditions

Three slides about theory

Table: A 2x2 Schema of Forecast Information, With Examples

	Conventional Econometric Models	Outside Conventional Econometric Models
internal	Lags of the series itself; past shocks	Polynomial expansion of the feature space and other transformations without solid theoretical motivation
external	Macro variables like interest rates, commodity prices; weather-related variables	Google Trends, high-frequency data like prediction markets; past shocks under similar conditions

Two slides about theory

Table: A 2x2 Schema of Forecast Information, With Examples

	Conventional Econometric Models	Outside Conventional Econometric Models
internal	Lags of the series itself; past shocks	Polynomial expansion of the feature space and other transformations without solid theoretical motivation
external	Macro variables like interest rates, commodity prices; weather-related variables	Google Trends, high-frequency data like prediction markets; past shocks under similar conditions

Two slides about theory

An overloaded term in applied math: similarity.
For our purposes, how should we define and determine similarity? Is it...

- Quantitative or qualitative?
- Approximation-based?
- Symmetric or asymmetric?
- Binary or n-ary?

Two slides about theory

An oft-abused term in applied math: similarity

- Quantitative or qualitative?
- Approximation-based?
- Symmetric or asymmetric?
- Binary or *n*-ary?

We call our approach distance-based weighting, which is a special case of an approximation-based *n*-ary similarity metric.

Global Overview

Abstracting away from particular models, what does our method require?

1. **Object-to-predict** Random object (indexed over time and possibly space, as well) obeying specification with additive errors

Global Overview

Abstracting away from particular models, what does our method require?

1. **Object-to-predict** Random object (indexed over time and possibly space, as well) obeying specification with additive errors
2. **Common Model Family on the Shocks** Requires that residuals be governed by a model that is shared across all units.

Global Overview

Abstracting away from particular models, what does our method require?

1. **Object-to-predict** Random object (indexed over time and possibly space, as well) obeying specification with additive errors
2. **Common Model Family on the Shocks** Requires that residuals be governed by a model that is shared across all units.
3. **Reliable and Shared Model-Fitting Procedure** There must exist a reliable model-fitting procedure for the $n + 1$ units.

Global Overview

Abstracting away from particular models, what does our method require?

1. **Object-to-predict** Random object (indexed over time and possibly space, as well) obeying specification with additive errors
2. **Common Model Family on the Shocks** Requires that residuals be governed by a model that is shared across all units.
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**

Global Overview

Abstracting away from particular models, what does our method require?

1. **Object-to-predict** Random object (indexed over time and possibly space, as well) obeying specification with additive errors
2. **Common Model Family on the Shocks** Requires that residuals be governed by a model that is shared across all units.
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 similarity.

Do we have theoretical guarantees?

We do!

Our formal results fall into the following buckets:

1. convergence in distribution results
2. consistency results
3. asymptotic loss

Here we share only the first two, and for both GARCH-X and AR(p) models.

The GARCH-X we work with

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

$$\mathcal{M}_1: \begin{aligned} \sigma^2_{i,t} &= \omega_i + \textstyle\sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} + \omega^*_{i,t}, \\ a_{i,t} &= \sigma_{i,t}((1 - D^{return}_{i,t})\epsilon_{i,t} + D^{return}_{i,t}\epsilon^*_i), \\ \omega^*_{i,t} &= D^{vol}_{i,t}[\mu_{\omega^*} + \delta^T \mathbf{x}_{i,T^*_i+1} + u_{i,t}], \end{aligned}$$

with time-invariant error structure

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

GARCH - Formal Results

Proposition

Assume

1. *For each* i, $1 \leq i \leq n$, $\{a_{i,t}\}_{t=0,\ldots,T_i}$ *obeys a GARCH-X*(m,s), *with volatility shocks found in* $\mathcal{M}_1$, *where* T_i *is the length of the* i*th series.*
2. *For each* i, $\{\omega^*_{i,t}\}_{t=0,\ldots,T_i}$ *is potentially non-zero at* $\{T_i^* + 1, ..., T_i^* + L_i^{vol}\}$, $\omega^*_{i,T_i^*+1} \equiv ... \equiv \omega^*_{i,T_i^*+L_i^{vol}}$, *and zero otherwise, where the arrival of* T_i^* *is governed by a time-invariant distribution on* $\{a_{i,t}\}_{t=0,\ldots,T_i-1}$, *and both the arrival and conclusion of the shock is observable by the researcher.*
3. *The conditions in Assumption 0 of* **han2014asymptotic** *hold.*

Then for any $i, 1 \leq i \leq n+1$, *and for any* $r, 1 \leq r \leq L_i^{vol}$, $\hat{\omega}^*_{i,T_i^*+r} \xrightarrow{p} \omega^*_{i,T_i^*+r}$ *as* $t \to \infty$. *Additionally,* $\hat{\omega}^*_{i,*} \xrightarrow{d} \omega^*_{i,T_i^*+r}$ *as* $t \to \infty$, *and if for all* $i, 1 \leq i \leq n+1$, $u_{i,t} \equiv 0$ *on* $\{T_i^* + 1, ..., T_i^* + L_i^{vol}\}$, *then* $\hat{\omega}^*_{i,T_i^*+r} \xrightarrow{p} \omega^*_{i,T_i^*+r}$.

In plain language, this result gives us distributional results about our correction terms as we estimate over arbitrarily long horizons.

GARCH - Formal Results

Having established the consistency of the estimators $\hat{\omega}^*_{i,T_i^*+r}$, we extend that result to prove asymptotic properties of the conditional forecast function itself.

Proposition

Assume

1. *All conditions listed in the preceding proposition.*
2. *There exist weights* $\{\pi_i\}_{i=2}^{n+1}$ *such that* $\mathbf{v}_{1,T_{\mathbf{1}}^*} = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T_i^*}$.

Then for any r, $1 \leq r \leq L_1^{vol}$, $\hat{\sigma}^2_{adjusted,T_{\mathbf{1}}^*+r} \xrightarrow{d} \sigma^2_{1,T_{\mathbf{1}}^*+r}$ *as* $t \to \infty$ *in the donor pool, and if for all* $i, 1 \leq i \leq n+1$, $u_{i,t} \equiv 0$ *on* $\{T_i^*+1, ..., T_i^*+L_i^{vol}\}$, *then* $\hat{\sigma}^2_{adjusted,T_{\mathbf{1}}^*+r} \xrightarrow{p} \sigma^2_{1,T_{\mathbf{1}}^*+r}$.

In plain language, our volatility prediction has a known distribution, by virtue of our correction terms have a known distribution.

AR(p) - Formal Results

We now present a similar AR(p) result.

Proposition

Assume

1. *For each* i, $1 \leq i \leq n+1$, *let* $\{y_t\}_{i=1}^{T_i}$ *follow an AR(p)-X.*
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* α_{i,T_i^*+1}

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

AR(p) - Formal Results

frame

Proposition

Assume

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

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

In plain terms: by 'cloning' the time series under study, we get a consistent predictor.

AR(p) - Formal Results

And now, as a consequence of the previous theorem, we can get distributional result for the prediction

Proposition

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

An inventory of models

- GARCH (Lundquist and Eck 2024 is under review at the *International Journal of Forecasting*)

For each of these, we would like to offer a slightly different taste of what is going on.

An inventory of models

- GARCH (Lundquist and Eck 2024 is under review at the *International Journal of Forecasting*)
- AR(p) (a canonical model helpful for illustrative purposes), which is covered in Lin and Eck 2021

For each of these, we would like to offer a slightly different taste of what is going on.

An inventory of models

- GARCH (Lundquist and Eck 2024 is under review at the *International Journal of Forecasting*)
- AR(p) (a canonical model helpful for illustrative purposes), which is covered in Lin and Eck 2021
- HAR (part of working paper)

For each of these, we would like to offer a slightly different taste of what is going on.

An inventory of models

- GARCH (Lundquist and Eck 2024 is under review at the *International Journal of Forecasting*)
- AR(p) (a canonical model helpful for illustrative purposes), which is covered in Lin and Eck 2021
- HAR (part of working paper)
- Exponential Shocks (part of working paper)

For each of these, we would like to offer a slightly different taste of what is going on.

GARCH Volatility Forecasts

The least you need to know here: GARCH is the premier volatility modeling family for daily data.

We demonstrate our method's usefulness with an easy-to-approach example.

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, as well as three Brexit-related events in 2016:
   - Prime Minister David Cameron's declaration of a Brexit referendum
   - A surprising Brexit poll result in June 2016
   - Brexit itself
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

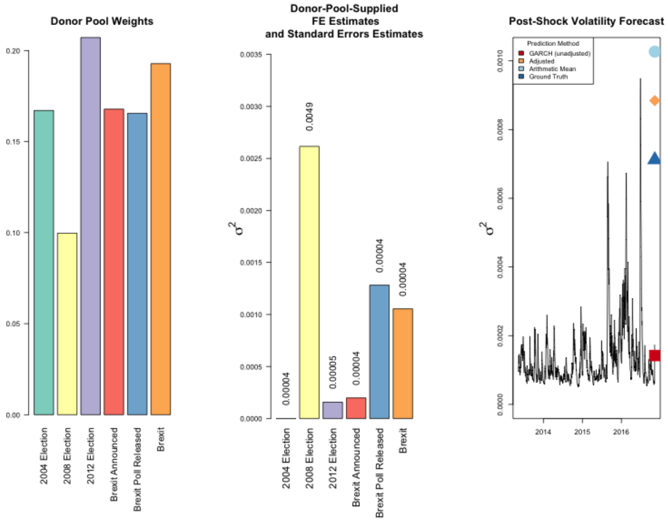


Figure: The volatility induced by the 2016 US election

AR(p)

We want to perform a simple bias-variance decomposition for an parameter-corrected AR(p) prediction. As usual, we omit time indices for easy of viewing:

$$
\begin{aligned}
\mathbb{E}[(y_t - \hat{y}_t)^2] &= \mathbb{E}[(y_t - \hat{\mu} + \sum_{k=1}^{p} \hat{\rho}_k y_{t-k} + \hat{\alpha}_t)^2] && (2) \\
&= \mathbb{E}[(\mu + \sum_{k=1}^{p} \rho_k y_{t-k} + \alpha_t + \epsilon_t - (\hat{\mu} + \sum_{k=1}^{p} \hat{\rho}_k y_{t-k} + \hat{\alpha}_t))^2] && (3) \\
&= \mathbb{E}[((\mu - \hat{\mu}) + \sum_{k=1}^{p} (\rho_i - \hat{\rho}_k) y_{t-k} + (\alpha_t - \hat{\alpha}_t) + \epsilon_t)^2] && (4)
\end{aligned}
$$

admits of the typical bias-variance decomposition. If it is known that $\mu = 0$, the last line becomes...

AR(p)

$$
\begin{aligned}
&\mathbb{E}[(\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t) + \epsilon_t)^2] \\
=&\text{Bias}^2[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \sigma^2_\epsilon \\
&&(\epsilon_t \text{ idiosyncratic}) \\
=&\Big[\mathbb{E}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)]\Big]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \sigma^2_\epsilon \\
&&(\text{Definition of Bias}) \\
=&\Big[\sum_{k=1}^{p}(\rho_k - \mathbb{E}[\hat{\rho}_k])y_{t-k} + (\alpha_t - \mathbb{E}[\hat{\alpha}_t])\Big]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k}] + \text{Var}[\hat{\alpha}_t] + \sigma^2_\epsilon \\
&&(\text{if donor shocks are uncorrelated with time series under study}) \\
=&\Big[\sum_{k=1}^{p}(\rho_k - \mathbb{E}[\hat{\rho}_k])y_{t-k} + (\alpha_t - \mathbb{E}[\hat{\alpha}_t])\Big]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k}]\sum_{i=2}^{n+1}\pi_i^2\text{Var}[\hat{\alpha}_{i,t}] \\
&+ 2\sum_{i\neq j}\pi_i\pi_j\text{Cov}[\hat{\alpha}_{i,t}, \hat{\alpha}_{j,t}] + \sigma^2_\epsilon
\end{aligned}
$$

Heterogeneous Autoregression (HAR)

The least you need to know: HAR is more recent volatility modeling approach that use high-frequency data and estimates the following predictive regression:

$$RV_t = \beta_0 + \beta_{\text{1-day}} RV_{t-1} + \beta_{\text{5-day}} \overline{\{RV_{t-1}, ..., RV_{t-5}\}} + \beta_{\text{22-day}} \overline{\{RV_{t-1}, ..., RV_{t-22}\}} + \epsilon_t$$

Heterogeneous Autoregression (HAR)

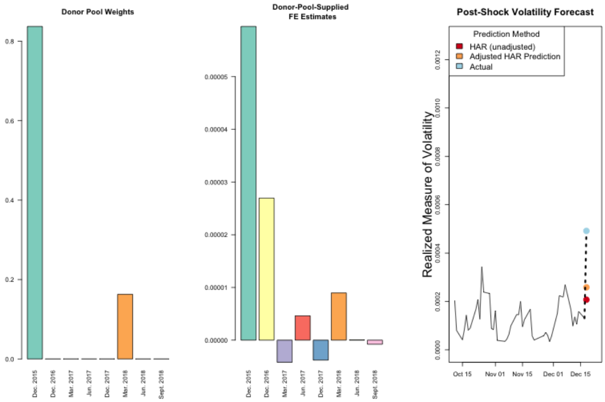


Figure: Click the image above to see a video of the R package's functionality.

Simulations: Parameter Correction Using An Aggregated Decay Parameter from Donors

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.

Now consider, for $i = 1, 2, ..., n+1$, the family of models

$$\mathcal{M}_{exp}: \quad \begin{aligned} y_{i,t} &= \nu_i + \eta[1 - e^{-\psi_i[t-T_i^*]}]\mathbf{1}_{t \geq T_i^*+1} + \epsilon_{i,t}, \epsilon_{i,t} \\ \psi_i &= \mathbf{x}_{i,t}^T \lambda_{i,t} \\ \mathbf{x}_{i,t}^T &= (1, x_{i,t}^1, ..., x_{i,t}^p) \\ \lambda_{i,t}^T &= (u_{i,t}, \lambda_t^1, ..., \lambda_t^p), \end{aligned} \tag{5}$$

with ν_i a fixed nuisance parameter, and with time-varying and observable covariate vector $\mathbf{x}_{i,t}^T$, time-varying, unobservable, and potentially random vector

$$\lambda_{i,t} \sim \mathcal{F}_\lambda \text{ with } \mathbb{E}_{\mathcal{F}_\lambda}(\lambda) = \mu_{\lambda_t}, \mathrm{Var}_{\mathcal{F}_\lambda}(\lambda) = \mathbf{\Sigma}_{\lambda_t},$$

and time-invariant error structure just as in $\mathcal{M}_1$.

Simulations: Parameter Correction Using An Aggregated Decay Parameter from Donors

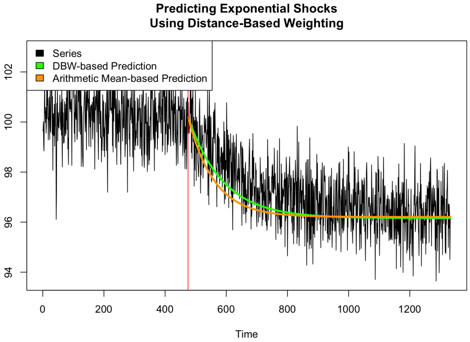


Figure: Simulated random walk centered at $\mu = 100$, subject to a shock at approximately time index 440. Shock of size -4 is not realized over a single index. Instead, shock is governed by a decay exponential decay parameter ψ_1, as are exponential shocks in $n = 10$ donor series. We estimate ψ_1 using a convex combination of the estimated decay parameters $\{\psi_i\}_{i=2}^{n+1}$, resulting in a distance-based weighting prediction of the shock. We also illustrate the arithmetic-mean-based prediction of the shock using the color orange. This prediction is based on an overestimate of ψ_1 and hence results in residuals that undershoot the time series under study.

Software

R package

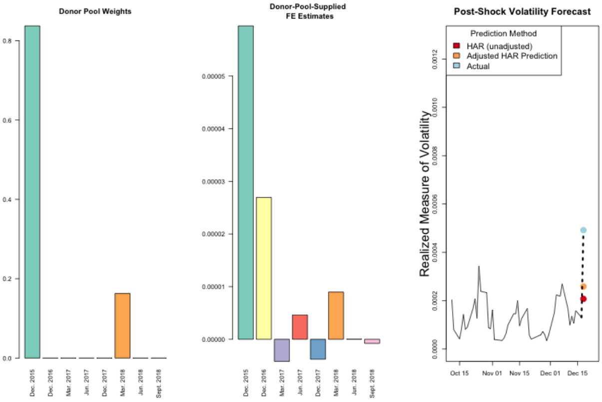


Figure: Click the image above to see a video of the R package's functionality.

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.

Message ChatGPT...

Robustness of the Approach

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

Some ideas here:

1. multiverse
2. permute donors

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^*
- Heterogeneity of DGPs
- Forecast evaluation
- We could only truly explore a subset of the vast parameter space.
- The failure of distance-based weighting to extrapolate

New Frontiers in Distance-based Weighting

- Forecast conditional upon the shock *prior* to the shock's arrival
- 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 $T_i, i = 1, 2, ...n + 1$.
- 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 $\widehat{IRF}_i(r, j)$, $i = 2, 3, ..., n + 1$?
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.

Chinco, 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*.