

# Forecast Adjustment Under Shocks: A Unification

David P. Lundquist\*, Daniel J. Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

July 25, 2024

## Abstract

Structural shocks to time series may give an observer reason to doubt the credibility of the default forecasting function. This work systematizes and unifies the rich landscape of model adjustment and model correction methods, with a special focus on forecast adjustment under the presence of shocks. We demonstrate the usefulness of similarity-based methods in forecasting and present several specific models that can benefit, along with formal results for some of those special cases.

## 1 Introduction

### Plan

What the paper should do

- Introduce an existing diffuse set of approaches to adjusting forecasts (1)
- Explain when/how similarity can help us forecast. 1.2
- distinguish the method from various tools that inspired it (1)
- show a few special cases, both examples and formal results (5)
- discuss limitations of the method (7.2)
- propose extensions (7.1)

What the paper should NOT do

- Wade too deeply into distance-based weighting details. We can refer to other work?
- wade too deeply into any of the special cases
- Tackle non-scalar random quantities (density forecasts, etc)
- Forecast combination

---

\*davidl11@illinois.edu

†dje13@illinois.edu

For various modeling and prediction tasks in time series and panel data, the salient challenge is not predicting when an event will occur but what its key properties will be. In the familiar case of scalar time series, those properties can include its post-event direction, moments, sign, magnitude, duration, and correlation structure, all over an arbitrary horizon or perhaps multiple horizons. This is not to say that predicting the arrival of an event is easy. In some cases, it may be difficult or impossible, and therefore preparing for an anticipated shock is the best one can hope for.

This work focuses on model adjustment amid structural shocks that undermine the reliability of the model at hand. Forecasting under anticipated shocks raises unavoidable questions: should the forecast model be abandoned in favor of a discretionary or ad-hoc or one-off adjustment? Does the discretion of a forecaster rule out a quantitative method for making the adjustment? What is the ultimate purpose of the adjustment, and how it is to be used? For how long is the adjustment necessary or reliable?

Herein we systematize and unify a range of conceptual approaches and tools that have developed across the broad ecosystem of the econometric and forecasting literatures. Additionally, we delineate a specific type of forecasting task called post-shock forecasting, which we broadly define as forecasting under a known shock.

Forecast model adjustment, known most widely perhaps by the term intercept-correction, has received the most attention in several articles and book chapters [Hendry and Clements, 1994, Clements and Hendry, 1996, 1998]. Of special importance is the distinction between discretionary and automated intercepts corrections. Hendry and Clements [1994] define scalar intercept corrections to be automated when they follow the simple rule of adding an estimation residual  $e_t$  to subsequent (possibly but not necessarily all) forecasts  $\hat{f}_{t+1}, \hat{f}_{t+2}, \dots$ . This procedure can colloquially referred to as setting the model back on track [Hendry and Clements, 1994]. In Hendry and Clements [1994], after recounting the bipartite division of interception corrections in discretionary and automated varieties, the authors present a six-way taxonomy of information that a modeler possesses at the time a forecast is made. The authors also consider structural change in the data-generating process during the forecast period, including as early as the first point in the forecast period (specifically in the autoregressive structure), as well as what is for them the more interesting case: structural change between  $T^* - 1$  and  $T^*$ . This current work finds both cases interesting. What if we could predict well the intercept shift that occurs between  $T^*$  and  $T^* + 1$ ?

Post-shock forecasting procedures have been explored in Lin and Eck [2021], Lundquist and Eck [2024], where the AR(1) and GARCH( $m, s$ ) cases, respectively, are treated. Both works target additive parameters in scalar time series, predicting those random effects by aggregating information from other time series. The authors leave several stones unturned, including a more general, dare say comprehensive treatment of how to forecast under any sort of shock.

A practitioner interested in forecast adjustment can choose between procedures that discretionary or automated, a variety of choices for the collection of data assembled to perform the correction, whether the data is internal (i.e. from the time series itself) or external, the parametric term to be corrected (e.g. intercept, coefficients), if any, as well as the correction function (i.e. the mapping from the data to the corrective term), including the weighting applied to the assembled data (e.g. Nearest-Neighbor, arithmetic mean, kernel methods).

The procedure presented herein is a discretionary procedure for intercept correction that integrates data internal or external to the time series under study in a systematic manner. The correction function uses an optimization step inspired by the causal inference literature. Outside of Lin and Eck [2021], Lundquist and Eck [2024], we are not aware of prior work that both introduces a parametric specification for nonzero errors and introduces a procedure for weighting appropriately the nonzero errors of similar shocks occurring outside the time series under study. We are also not familiar with any prior work that attempts to account for anticipated nonzero errors using an a parametric adjustment, i.e., a “correction function”.

## 1.1 Motivation for Intercept Correction

### 1.1.1 Setting levels back on track

#### 1.1.2 Mismeasured data

Mismeasured data is discussed in [Hendry and Clements, 1994, p. 166] as a motivator for intercept correction. Could similarity-based correction help? Here is an idea: if we believe that our most recent measurement of the series is noisy, we can disregard the point itself and instead take a convex combination of that point and the Lin and Eck [2021]-style prediction based on aggregation.

Quinton-Guerrera and Zhong Guerrón-Quintana and Zhong [2017] - concerned with correcting  $\beta$  using similarity

#### 1.1.3 Motivation from model evaluation

Part of the reason for correcting a model springs from the way we evaluate model performance. Evaluating a Model by Forecast Performance Clements and Hendry [2005]

1. unconditional versus conditional, models;
2. internal versus external standards;
3. checking constancy versus adventitious significance;
4. ex ante versus ex post evaluation (skip this one?);
5. 1-step versus multi-horizon forecasts – this is a relevant question to ask in the context of post-shocking forecasting: should we correct the earliest forecast and then allow the shock to propagate, or should we just correct each term in the horizon,  $h=1, 2, \dots, H$ ?
6. in-sample fixed coefficients versus continuous updating

## 1.2 The Meaning and Use of Similarity

Hume?

The notion of similarity appears in various statistical contexts, including matching, synthetic control, nearest-neighbor methods, not to mention the massive area of approximation theory.

Similar, in the strong sense, could mean that a shared DGP exists.

Quantitative ways of determining similarity

Distance function

Asymmetric distance functions are useful when we want to explore differences between donors and weight different contributors to that function differently.

What about qualitative ways of determining similarity – Lundquist and Eck [2024] donors are identified qualitatively

## 2 Setting

In order to motivate our procedure, we provide a visual illustration. In Figure [decide on which figure to include] we show how the aggregation of estimated excess volatilities from donors in the donor pool works when the correction function is a specially-chosen convex combination of fixed effects from the donor pool. Our method assumes a credible, parsimonious parameterization in which the shock is an affine transformation of several key covariates. The key intuition behind this shock parameterization is that as the strength of the linear signal increases relative to the idiosyncratic error, the GARCH estimation of these effects increases in accuracy. From this, it follows that the aggregated shock estimate increases in accuracy.

### 2.1 Model setup

Assume a researcher has multivariate time series data  $\mathbf{y}_{i,t} = (r_{i,t}, \mathbf{x}_{i,t})$ ,  $t = 1, \dots, T_i$ ,  $i = 1, \dots, n+1$ , and  $\mathbf{x}_{i,t}$  is a vector of covariates such that  $\mathbf{x}_{i,t} | \mathcal{F}_{i,t-1}$  is deterministic. Suppose that the analyst is interested in forecasting the volatility of  $r_{1,t}$ , the first time series in the collection, which we will denote the time series under study. We require that each time series  $\mathbf{y}_{i,t}$  is subject to a news event following  $T_i^* \leq T_i + 1$  and before witnessing  $T_i^* + 1$ .

$$y_{i,t} = F(\mathcal{F}_{t-1}) + \epsilon_t$$

Let  $I(\cdot)$  be an indicator function. Let  $T_i$  denote the time length of the time series  $i$  for  $i = 1, \dots, n+1$ , and let  $T_i^*$  denote the largest time index prior to the arrival of the news shock, with  $T_i^* < T_i$ . Let  $\delta, \mathbf{x}_{i,t} \in \mathbb{R}^p$ . For  $t = 1, \dots, T_i$  and  $i = 1, \dots, n+1$ , the model  $\mathcal{M}_1$  is defined as

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

with error structure

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

Note that  $\mathcal{M}_0$  assumes that  $\omega_i^*$  have no dependence on the covariates and are i.i.d. with  $\mathbb{E}[\omega_i^*] = \mu_{\omega^*}$ .

### 2.2 Shock Profile of a Time Series

We now introduce prediction via distance-based weighting by constructing a profile of a time series' shocks, the analogue of a covariate matrix in a synthetic control framework. The shock profile for a given GARCH-X process is nothing more than a vector of observable covariates that parameterize a shock. A collection of  $n$  such vectors yields a matrix of  $n$  columns. What distinguishes a shock profile, principally, is that the time-varying parameters that correspond to the volatility profile are non-zero at only the shock times.

Suppose that for each of the  $n$  donors, we have available  $p$  distinct covariates in the functional form of the shock. The shock profile could take the form of the  $p \times n$  matrix

$$\mathbf{V}_t = \begin{pmatrix} RV_{2,t} & RV_{3,t} & \cdots & RV_{n+1,t} \\ RV_{2,t-1} & RV_{3,t-1} & \cdots & RV_{n+1,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ IV_{2,t} & IV_{3,t} & \cdots & IV_{n+1,t} \\ IV_{2,t-1} & IV_{3,t-1} & \cdots & IV_{n+1,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ |a_{2,t}| & |a_{3,t}| & \cdots & |a_{n+1,t}| \\ |a_{2,t-1}| & |a_{3,t-1}| & \cdots & |a_{n+1,t-1}| \\ \vdots & \vdots & \ddots & \vdots \\ Volume_{2,t} & Volume_{3,t} & \cdots & Volume_{n+1,t} \\ Volume_{2,t-1} & Volume_{3,t-1} & \cdots & Volume_{n+1,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta RV_{2,t} & \Delta RV_{3,t} & \cdots & \Delta RV_{n+1,t} \\ \Delta RV_{2,t-1} & \Delta RV_{3,t-1} & \cdots & \Delta RV_{n+1,t-1} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix},$$

Covariates chosen for inclusion in a given shock profile may be levels, log differences in levels, percentage changes in levels, or absolute values thereof, among many choices.

As shown,  $\mathbf{V}_t$  displays ‘balance’ in that  $p$  covariates exist for each of the  $n$  donors. In practice, missing values, corrupted values, or unacceptably extreme or noisy estimates may necessitate some sort of matrix completion, a problem that we do not tackle in this work. We now turn to the next section, where  $\mathbf{V}_t$  is employed in a procedure to arrive at a forecast adjustment.

### 3 Model Adjustment Using Similarity-Based Parameter Correction: A Global Overview

In this section, we introduce and discuss a particular approach to model adjustment that is motivated by the circumstances laid out in Section 2.

1. a random object to forecast that depends on a linear specification or a specification that can be linearized
2. a parametric model family shared by donors<sup>1</sup>
3. a correction term for the model family shared by donors
4. a parametric specification for the correction term

This is important because with a parametric specification, we cannot get key scalars to take a convex combination of.

Does it draw on data internal or external to the time series under study, or does it draw upon both?

---

<sup>1</sup>Emphasize that there is a non-parametric version of the above: for example, one can use LSTM to predict each of the donor shocks, and then the residuals (or transformations of those residuals, e.g. to percentages) of those  $n$  models can be weighted to arrive at a correction term.

5. a reliable estimation procedure for the shared model.

When we use fixed effect estimation (under ordinary assumptions), we can construct confidence intervals for the fixed effect estimates, and then assuming independence, we can get confidence intervals for convex combinations of fixed effect estimates.

- (a) This should be straightforward

6. a reliable estimation procedure for the correction term

- (a) This might not be straightforward. Some models like GARCH, for example, might deliver very noisy estimates for indicator variables that occur just once.
- (b) When will it be as simple as a fixed effect?
- (c) When will it be something besides a fixed effect?

7. a correction function, based on the correction term, that aggregates (i.e. maps) donor correction terms based on some notion of similarity

## 4 Methodology for Similarity-based Parameter Correction

### 4.1 Forecasting

So far we have presented a setting and a data generating process for which we intend to furnish a statistical method. Now, we present two forecasts for the time series, the latter of which benefits from the proposed method. In particular, we present the GARCH forecast, also called the *unadjusted* forecast, as well as the *adjusted* forecast, which differs by an additive term:

$$\begin{aligned} \text{Forecast 1: } \hat{\sigma}_{unadjusted}^2 &= \hat{\mathbb{E}}[\sigma_{1,T_1^*+1}^2 | \mathcal{F}_{T^*}] &= \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \sigma_{i,t-j}^2 + \hat{\gamma}_i^T \mathbf{x}_{i,t} \\ \text{Forecast 2: } \hat{\sigma}_{adjusted}^2 &= \hat{\mathbb{E}}[\sigma_{1,T_1^*+1}^2 | \mathcal{F}_{T^*}] + \hat{\omega}^* &= \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \sigma_{i,t-j}^2 + \hat{\gamma}_i^T \mathbf{x}_{i,t} + \hat{\omega}^* . \end{aligned}$$

### 4.2 Shock Estimators

The problem of aggregating estimated donor shocks begins with the data constraints. Taking the estimated shocks as a given, we essentially observe the pair  $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ . We wish to recover weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$  leading to favorable forecasting properties. These weights are used to compute  $\hat{\omega}^* := \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ , our forecast adjustment term. Since the weights  $\{\pi_i\}_{i=2}^{n+1}$  are computed using  $\mathcal{F}_{T_i^*}$ , the set  $\{\pi_i\}_{i=2}^{n+1}$  is deterministic, *modulo* any stochastic ingredient in the numerical methods employed to approximate  $\mathbf{x}_{1,T^*}$  using a convex combination of donor covariates. We say more about the properties of  $\omega_i^*$  in section ??.

Following [Abadie and Gardeazabal \[2003\]](#), [Abadie et al. \[2010\]](#), let  $\|\cdot\|_{\mathbf{S}}$  denote any semi-norm on  $\mathbb{R}^p$ , and define

$$\{\pi\}_{i=2}^{n+1} = \arg \min_{\pi} \|\mathbf{v}_1 - \mathbf{V}_t \pi\|_{\mathbf{S}} .$$

In the forecast combination literature, it is discussed whether the weights employed to aggregate forecasts strive toward and meet various optimality criteria [[Timmermann, 2006](#), [Wang et al., 2023](#)]. In this work,

there are at least two senses of optimal weights that one might be interested in. First, we can think of optimal weights as a set  $\{\pi_i\}_{i=2}^{n+1}$  such that  $\omega_1 = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i$ , i.e.,  $\omega_1$  is recovered perfectly, as it belongs to convex hull of the estimated shocks. However,  $\omega_1$  is never revealed to the practitioner, and hence there is no way of verifying the extent to which this condition is satisfied.

A more promising aim is finding weights such that  $\mathbf{v}_1 = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_i$ , meaning that the shock profile of the time series under study lies within the convex hull of the donor shock profile. This condition underwrites asymptotic results in [Abadie et al. \[2010\]](#), and the intuition there extends to this work: if the shock is parameterized by an affine function of covariates, then finding a linear combination that recreates the shock should serve us well. Because the method proposed uses a point in  $\Delta^n$ , it is important to head-off possible confusion. What we are proposing is not a forecast combination method. What we are aggregating and weighting (not combining) are subcomponents of forecasts, not forecasts themselves. Moreover, from a broader perspective, forecast combination is an inapt term for what is being proposed here. First, the donor time series do not provide forecasts, nor would forecasts be needed for random variables that have already been realized. Second and more fundamentally, the theoretical underpinnings of forecast combination, while diverse [\[Wang et al., 2023\]](#), are distinct from the setting presumed in this work, where the model family is not in doubt but the parameter values and how they prevail must be learned.

Supposing that we can define optimal weights, are they necessarily unique? [Lin and Eck \[2021\]](#) discuss sufficient conditions for uniqueness as well as the implications of non-uniqueness. [Abadie and Vives-i Bastida \[2022\]](#) invoke the Carathéodory Theorem to argue for the sparseness of the weight vector. We make additional comments as well.  $(\mathbb{R}^n, \|\cdot\|)$  is a Chebyshev space, and hence for any element  $x$  and any convex set  $C \subset \mathbb{R}^n$ , there exists a unique element  $y \in C$  that minimizes  $\|x - y\|$ . However, the pre-image of  $y$  with respect to a particular operator and constraint set might not be unique. Let  $p', n'$  denote the number of linearly independent rows of  $\mathbf{V}_t$  and linearly independent columns of  $\mathbf{V}_t$ , respectively. Let  $\text{col}(\cdot)$  denote the column space of a matrix, and let  $\text{Conv}(\cdot)$  denote the convex hull of a set of column vectors.

	$\mathbf{v}_1 \in \text{Conv}(\text{col}(\mathbf{V}_t))$	$\mathbf{v}_1 \notin \text{Conv}(\text{col}(\mathbf{V}_t))$
$p' \geq n'$	Perfect fit; fit unique	Fit not perfect; fit unique
$p' < n'$	Perfect fit, not necessarily unique, Carathéodory Theorem applies	Fit not perfect, not necessarily unique, Carathéodory Theorem applies

Comparison with least-squares estimation is illustrative. Consider least-squares for the  $n$ -vector of estimated volatility shocks  $\hat{\omega}^*$ :

$$\vec{w}_{OLS} = \arg \min_w \|\hat{\omega}^* - w^T \mathbf{V}_t\|_2$$

One immediately visible problem is that this optimization problem is an optimization problem over  $p$ -vectors  $\vec{w}$  — i.e. over linear combinations of the covariates, whereas what we seek is an  $n$ -vector — a linear combination of donors. Additionally, there is no guarantee that  $\vec{w}_{OLS}$  would perform poorly as a tool for producing  $\hat{\omega}^*$ , but given the small number of donors supposed in our setting, it is risky.

### 4.3 How is the proposed method distinct from existing tools?

### 4.4 Relaxation of Assumptions

#### 4.4.1 How important is a shared DGP?

### 4.5 We can use the method on latent time series

See [Lundquist and Eck \[2024\]](#)

## 5 Formal Properties and Model-Specific Considerations

### 5.1 What are the downsides of aggregating residuals as opposed to aggregating fixed effect estimates?

Cases:

1. Fixed effect is estimated with accuracy
2. Fixed effect is estimated with considerable noise. Hence, the idiosyncratic error is estimated with considerable noise.
3. idiosyncratic error large
4. idiosyncratic error small

### 5.2 ARIMA

[Lin and Eck \[2021\]](#)

### 5.3 GARCH

We briefly highlight the findings of [\[Lundquist and Eck, 2024\]](#).

### 5.4 HAR

### 5.5 VAR

Many time series, especially macroeconomic time series, naturally arise as constituents of groups of dependent variables that interact across time.

### 5.6 LSTM/GRU

Here we demonstrate the tremendous capaciousness of our method by applying it to predictions generated by a pair of non-parametric forecasting functions, Long Short Term Memory and GRU.

- [a](#)
- [b](#)
- [c](#)
- [d](#)
- [LSTM in Python](#)

## 6 Real Data Examples

This section might not be necessary. g



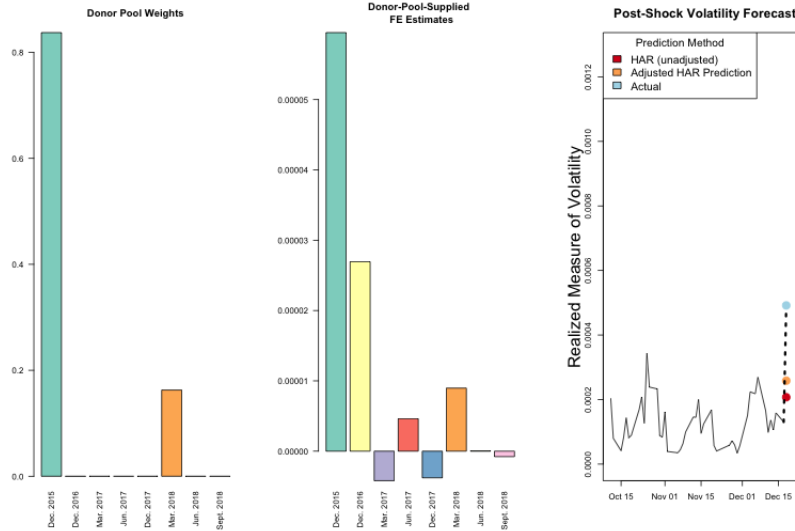


Figure 1: Volatility series of six i.i.d. GARCH processes, each of which experiences a volatility shock, indicated with a red vertical line, at a uniformly distributed point in the set  $\{756, \dots, 2520\}$  of trading days, corresponding to between 3 and 10 years of daily trading data.

## 7 Discussion

Are there any situations where we would want to intercept-correct using the post-shock term from [Lin and Eck \[2021\]](#) rather than the actual residual?

The forecast horizon — does it matter? If so, how so? [Lin and Eck \[2021\]](#) has a one-period horizon. [Clements and Hendry \[1998\]](#) p. 203 discuss how long to keep the forecast adjustment in place. For a corrected “slope parameter”, the effect of  $h$  is not so clear.

- Binary Outcome Forecasts
- Density Forecasts
- Quantile Forecasts

### 7.1 Extensions

bias-variance tradeoff and MSE decomposition

### 7.2 Limitations

### 7.3 Forecast Combination

what we are talking about here is not forecast combination, but there may be, nevertheless, a role for forecast combination: combining the forecasts generated by small differences in covariate and/or donor choice, as is done in [Lundquist and Eck \[2024\]](#).

## References

- Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, 2003.
- Alberto Abadie and Jaume Vives-i Bastida. Synthetic controls in action. *arXiv preprint arXiv:2203.06279*, 2022.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.
- Michael Clements and David F Hendry. *Forecasting economic time series*. Cambridge University Press, 1998.
- Michael P Clements and David F Hendry. Intercept corrections and structural change. *Journal of Applied Econometrics*, 11(5):475–494, 1996.
- Michael P Clements and David F Hendry. Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, 67:931–956, 2005.
- Pablo Guerrón-Quintana and Molin Zhong. Macroeconomic forecasting in times of crises. 2017.
- D Hendry and M Clements. On a theory of intercept corrections in macroeconometric forecasting. 1994.
- Jilei Lin and Daniel J Eck. Minimizing post-shock forecasting error through aggregation of outside information. *International Journal of Forecasting*, 2021.
- David Lundquist and Daniel Eck. Volatility forecasting using similarity-based parameter correction and aggregated shock information. *arXiv preprint arXiv:2406.08738*, 2024.
- Allan Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.
- Xiaoqian Wang, Rob J Hyndman, Feng Li, and Yanfei Kang. Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547, 2023.