

Forecast Adjustment Under Shocks: A Unification

David P. Lundquist*, Daniel J. Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

September 26, 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. 1.4
- 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. State that it's simply one way to do it. Allude to [Lin and Eck \[2021\]](#), [Lundquist and Eck \[2024\]](#).
- 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 term to be corrected (e.g. intercept, coefficients), if any, as well as the correction function (i.e. the mapping from the donor unit data to the correction 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”.

The structure of this manuscript is as follows. We first provide a review of several disparate time series literatures, including intercept correction, model-evaluation, and similarity-based forecasting. We

then provide a canonical setting in which we attempt to focus our work. This setting will be presented at a level of generality that showcases the broad applicability of our method. We then introduce the method from a global perspective, abstracting away from familiar applications. We then show several specific applications, including novel applications that cannot be found elsewhere in the literature. We close with a discussion, including possible extensions.

1.1 Here is what we are not doing

1.1.1 Intercept Correction

1.1.2 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 \[2024\]](#).

1.1.3 Tackle non-scalar random quantities (density forecasts, etc)

- Function forecasts
- Binary Outcome Forecasts
- Density Forecasts
- Quantile Forecasts

1.2 History of and Motivation for Intercept Correction

See Clements and Hendry

1.2.1 Setting levels back on track

We discussed this above in Section [1.1.1](#).

1.3 Incorporation of Outside Information is the Norm, Not the Exception

ARMA models and other models that model the future as a function of only the past values and past innovations of the series itself are well-known to be unrealistic as a data-generating processes. They are also well-known to be poor predictors of the future. What we are doing here is therefore not out of the ordinary but distinct in at least two ways: (1) we are incorporating outside information faster, and (2) we are proposing a principled, systematic way to do it that defies that curse of dimensionality. If one wanted to forecast using linear models and high-dimensional information set, then some kind of method for reducing model complexity would be required. We avoid that.

1.3.1 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 β using similarity

1.3.2 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

Here is the argument: if we are going to judge a model by forecast performance, we need to entertain the possibility that the model may breakdown and that our pre-breakdown-based criteria for model performance will be invalidated.

1.4 The Meaning and Use of Similarity

Hume? Cite Israeli authors, if so.

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. The uses of similarity are too sprawling to recount in detail. Below we talk about how it could matter in this work.

Similar, in the strong sense, could mean that a shared DGP exists. It could mean similar in the covariates that parameterize the shocks. It matters where we look for similarity and how we estimate similarity.

Quantitative ways of determining similarity include matching a target object to other objects along one or more variables, where exact matches may be likely or may be statistically impossible. In metric spaces, the notion of a metric that quantifies nearness is crucial to fundamental concepts of proximity. This suggests an optimization problem such that, for any target object Q , the search for similarity means a search for an approximator A s.t. $\delta(Q, A)$ is smallest, where δ is an appropriate metric.

Asymmetric distance functions (which are of course not metrics) 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? In [Lundquist and Eck \[2024\]](#), donors are identified by matching the qualitative aspects of the shock in the time series under study to shocks previously witnessed in other series. Only after the donor pool has been identified based on qualitative facts about the shocks is there a role for quantitative tools.

2 Setting

It's unclear how much I can show the reader, given that I want this to be general.

In order to introduce the general framework for adjusting forecasts under shocks, we provide an 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 estimation of these effects increases in accuracy. From this, it follows that the aggregated shock estimate increases in accuracy.

The preceding information is correct. However, it comes from the previous paper, so it needs to be changed to suit the style of the current paper.

2.1 A shock between discrete time points

In this work, we will be focused on shocks that occur strictly between two discrete time points. This raises an important terminological point. “Shocks”, in econometric literature, can be structural or idiosyncratic, permanent or transitory, supply-related or demand-related, and so forth. The taxonomy is vast.

2.2 Formal Modeling Setup

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t} = (r_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \dots, T_i$, $i = 1, \dots, n + 1$, where $r_{i,t}$ is scalar and $\mathbf{x}_{i,t}$ is a vector of covariates such that $\mathbf{x}_{i,t} | \mathcal{F}_{i,t-1}$ is deterministic and observed. Suppose that the analyst is interested in forecasting $r_{1,t}$, the first time series in the collection, which we will denote *the time series under study*. We assume the following general data-generating process

$$y_{i,t} = F(\mathcal{F}_{i,t-1}) + \mathbf{x}_{i,t}^T \lambda_t + \epsilon_{i,t}$$

where $\epsilon_{i,t}$ is mean-zero and uncorrelated across time and donors. Note that this specification above is capacious enough to include data-generating processes of the form

$$y_{i,t} = \log Y_{i,t} = \log a_{i,t} y_{i,t-k} e^{\mathbf{x}_{i,t}^T \lambda_t + \epsilon_{i,t}} = \log a_{i,t} + \log y_{i,t-k} + \mathbf{x}_{i,t}^T \lambda_t + \epsilon_{i,t}$$

with $k \geq 1$ and $a_{i,t}, y_{i,t}$ supported on \mathbb{R}^+ . The term $\mathbf{x}_{i,t}^T \lambda_t$ will be deemed the *correction term*, and the function

$$\xi: \mathcal{F}_t \times \mathcal{F}_t \rightarrow \mathbb{C}$$

will be deemed the *correction function*. The correction function maps observable, deterministic events and unobservable events into a space so that something might be learned about λ_t . If $\mathbf{x}_{i,t}$ were not observable or not deterministic with respect to $\sqcup - \infty$, then the correction term would represent an additional error term (not necessarily idiosyncratic.)

We require that each time series $\mathbf{y}_{i,t}$ is subject to an observed news event following $T_i^* \leq T_i + 1$ and before witnessing $T_i^* + 1$. The reasons for this represent the heart and soul of the framework proposed. We are not proposing a method to correct for unanticipated shocks that have been reflected in the observed data. Instead, we are proposing a method to correct for shocks for which the qualitative kernel of information is known and the observed quantitative instantiation of that which is not yet known.

So far we have reviewed a long and rich literature about model adjustment, and we have also introduced a data-generating process upon which the rest of this manuscript will be based. We now proceed to introduce a particular framework of solutions to the circumstances assumed in 2.



Figure 1: The simulated time series experiences a volatility shock between trading days 1,656 and 1,657. The GARCH prediction, in red, fails even to approach the volatility spike at $T^* + 1$, as do several adjusted predictions, which are orange. In contrast, the GARCH forecast adjusted by $\hat{\omega}^* = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$, a convex combination of the estimated shocks in the donor pool, achieves directional correctness as well as a smaller absolute loss in its prediction. The pink vertical line serves to indicate the adjustment of size $\hat{\omega}^*$ that allows the blue bullseye to approach more closely the ground truth.

3 Methodology for Similarity-based Parameter Correction

3.1 Forecasting

We now turn to the next section, where \mathbf{V}_t is employed in a procedure to arrive at a forecast adjustment. For illustration, we present two one-step-ahead forecasts for the time series under study. The first is the unadjusted GARCH forecast. The second is the adjusted forecast, which differs by the additive term $\hat{\omega}^*$ that is computed via our distance-based weighting procedures. These forecasts are:

These should be kept, with appropriate things changed.

$$\hat{\sigma}_{unadjusted, T_1^*+1}^2 = \hat{\mathbb{E}}[|\mathcal{F}_{T_1^*}|] = ,$$

$$\hat{\sigma}_{adjusted, T_1^*+1}^2 = \hat{\mathbb{E}}[\mathcal{F}_{T_1^*}] + \hat{\omega}^* = .$$

3.2 Shock Estimators

The problem of aggregating estimated donor shocks begins with the data constraints. Let us first introduce useful notation. Let $\hat{\omega}_{i,*}^*$ denote the shock estimate for donor i that is obtain via an indicator variable over time points $T_i^* + 1, \dots, T_i^* + L_{i,vol}$. That estimation procedure is justified by the assumption that at each nonzero shock point, the shocks will differ but will be equal in distribution.

Taking the estimated shocks as a given, we observe the pair $(\{\hat{\omega}_{i,*}^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$. Let $\Delta^n = \{\pi \in \mathbb{R}^n : \sum_{i=1}^n \pi_i = 1, \pi_i \geq 0 \mid i = 1, \dots, n\}$. We wish to recover weights $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$ leading to favorable forecasting properties. These weights are used to define and compute an aggregate shock estimate

$$\hat{\omega}^* = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_{i,*}^*, \quad (1)$$

which will be taken to be 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 and observed, *modulo* any stochastic ingredient in the numerical methods employed to approximate \mathbf{x}_{1,T^*+1} using a convex combination of donor covariates. We say more about the properties of the shocks $\omega_{i,T_i^*+1}^*, \dots, \omega_{i,T_i^*+L_{i,vol}}^*$ in section ??.

Following [Abadie and Gardeazabal \[2003\]](#), [Abadie et al. \[2010\]](#), [Lin and Eck \[2021\]](#), let $\|\cdot\|_{\mathbf{S}}$ denote any semi-norm on \mathbb{R}^p , and define

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

In the forecast combination literature, it is of interest whether the weights employed to aggregate forecasts strive toward and meet various optimality criteria [[Timmermann, 2006](#), [Wang et al., 2023](#)]. In our endeavor, 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., ω_1 is recovered perfectly, as it belongs to convex hull of the estimated shocks. However, ω_1 is never revealed to the practitioner, and hence there is no way of verifying the extent to which this condition is satisfied.

3.3 Shock propagation and estimation accuracy

Consider [Lundquist and Eck \[2024\]](#), in which a real world example shows how volatility can be forecasted following surprising election outcomes. There, the estimation of fixed effects is done around elections in a donor pool. That means that for each past election or poll, etcetera, the method must accurately estimate the surprise-induced volatility boost. That means picking a window around such an effect.

proposals

1. For a given data-generating process, simulate the accuracy and variability of shock estimates using post-shock measurement periods of difference lengths.
2. Develop formal results to quantify (estimate, bound, etc) the risk of using different post-estimation measurement lengths. This matters because it might be that a donor has experienced a shock in the last m days, and if m is too small, using that donor may be risky.

3.4 How is the proposed method distinct from existing tools?

3.5 Relaxation of Assumptions

3.5.1 How important is a shared DGP?

The assumption of a shared family can be relaxed in certain circumstances. For example, if a certain donor is not like the time series under study with respect to the covariates, then the lack of a shared family will not matter much.

3.6 We can use the method on latent time series

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

4 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. It may not yet be clear what is essential and what is not. Here we make that clear.

4.1 The five essential elements of similarity-based parameter correction

Should I just use OLS as an example throughout all six?

Most fundamentally, the method requires a random object (indexed over time, of course) that obeys a specification with additive errors, or, at the very least, a specification that can be transformed to have additive errors. This requirement is suitably weak, so as to include models that are not linear or not linear in each of their parameters. It also includes objects that multidimensional as well as objects from non-Euclidean probability spaces, like some function spaces. Notice that we did not say *parametric* specification. The reason for this is that, given any forecast function f from an information set \mathcal{F}_t , and given any appropriate loss function L , we can define a residual $e_t = y \odot \hat{y}$. \odot is subtraction in the simple case of MSE. The residuals (or transformations of those residuals) of those n models can be weighted to arrive at a correction term. This fact is especially useful when the forecast function f is nonparametric, black-box, or stochastic with respect to \mathcal{F}_t .

Second, the method requires that residuals be governed by a model that is shared across all units. This ensures that in the estimation of structural 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 \[2024\]](#).

Third, there must exist a reliable model-fitting procedure for the $n + 1$ units, one that will allow the assumption of a shared family for the residuals to be reflected in the residuals. Here, reliable means several things. It means that the estimation procedure must produce a credible description of each data-generating process or a good prediction function, so as to aid in estimating residual in each unit.

‘Reliable’ may mean the estimators have low variance, and it may also mean that the estimation procedure is robust, to some degree, to misspecification bias. However, in theory, the lack of these properties is not harmful to the method unless the lack of these properties harms the correction term’s estimation. 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.

Fourth, there must exist a reliable procedure for modeling the correction term in each unit. Again, here we care about low variance as well as robustness to misspecification. The very simplest correction

term is the residual itself. The correction term could be an inner product that depends upon external covariates, as is found in [Lin and Eck \[2021\]](#), [Lundquist and Eck \[2024\]](#).

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

Fifth and finally, 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. It could depend upon external covariates.

5 Formal Properties and Model-Specific Considerations

5.1 ARIMA

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

5.2 GARCH

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

5.3 HAR

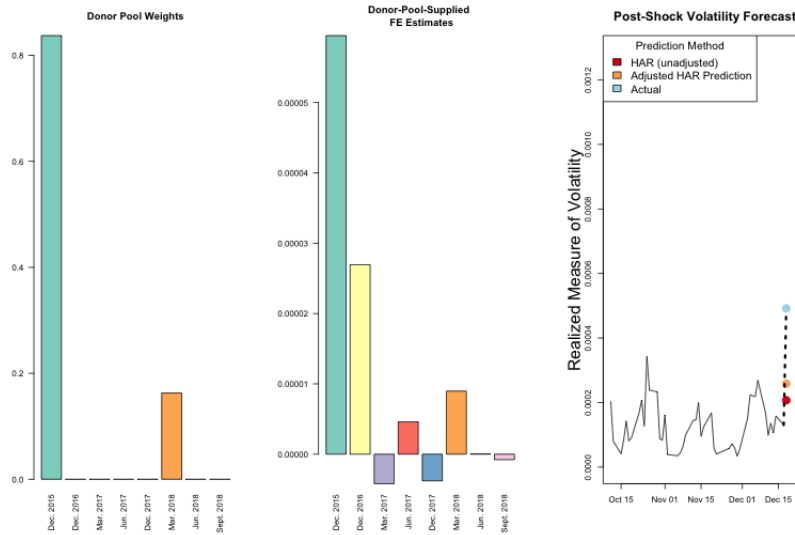


Figure 2: Volatility series of six i.i.d. GARCH processes, each of which experiences a volatility shock, indicated with a red vertical line, at a uniformly distributed point in the set $\{756, \dots, 2520\}$ of trading days, corresponding to between 3 and 10 years of daily trading data.

5.4 VAR

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

5.5 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. We will re-examine an application found in [Lin and Eck \[2021\]](#).

We borrow code from [Brownlee \[2022\]](#)

- [LSTM in Python](#)

5.5.1 Dual-shocks and the method of shock-partitioning

5.6 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

5.7 The Method of Multiple Shocks

Sometimes, we might have reason to believe that a forecast requires adjustment due to not one but multiple shocks. [Lin and Eck \[2021\]](#) explored a “dual” shock in their analysis of oil company equity price shocks due to the March 2020 COVID lockdowns.

Method 1: for each source of the k shocks, assemble n_j donors and p_j covariates, with $j = 2, \dots, n_j + 1$, compute weights $\{\pi_{i,j}\}_{i=2}^{n_j+1}$. With these k weight sets, a set of k aggregated shock estimators can be constructed and averaged (or aggregated in some other way).

Method 2: the k shocks can be represented using one application of distance-based weighting, with the vector δ and the vector \mathbf{x}_{i,T_i^*+1} can encode a partition of shocks by using zeros in appropriate entries.

6 Real Data Examples

This section might not be necessary.

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.

7.1 Extensions

bias-variance tradeoff and MSE decomposition

- Function forecasts
- Binary Outcome Forecasts
- Density Forecasts
- Quantile Forecasts

7.2 Limitations

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, 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.
- Jason Brownlee. Time series prediction with lstm recurrent neural networks in python with keras, Aug 2022. URL <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>.
- 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.