

Forecast Adjustment Under Shocks: A Unification

David P. Lundquist*, Daniel J. Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

October 15, 2024

Abstract

When should one’s default forecasting model be adjusted, augmented, or abandoned completely? Should it be done algorithmically or via human discretion, and if the latter, how much discretion? Should an adjustment be carried out using external information, and what exactly should be adjusted? 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 news shocks, when unanticipated events may give an observer reason to doubt the credibility of the default forecasting function. We demonstrate the usefulness of similarity-based methods in forecasting and present a general framework dubbed Similarity-based Parameter Correction (SPC). We highlight several specific time series models that can benefit from SPC, along with formal results for some of those special cases.

1 Introduction

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 the time series’ post-event direction, magnitude of change, moments, duration of the event, 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, or the event might not be conceivable to many before it happens, as was the case with the COVID-19 pandemic. There, reacting sensibly to a shock or structural change may best one can hope for.

Herein we attempt to unify a range of conceptual approaches to forecasting amid shocks that have developed across the broad ecosystem of the econometric and forecasting literatures. We attempt to locate our main contribution, Similarity-based Parameter Correction, at the intersection of several important strands of thought. Additionally, we delineate a specific type of SPC called post-shock forecasting, which is induced by nuanced and interesting choices that fall under the general SPC framework. In particular, this work focuses on model adjustment amid news shocks that undermine the reliability of the model at hand. Forecasting under 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?

Forecast model adjustment, known most widely perhaps by the term intercept-correction, has received the most attention in several articles and book chapters [???]. Of special importance is the distinction between discretionary and automated intercepts corrections. ? define scalar intercept corrections to

*davidl11@illinois.edu

†dje13@illinois.edu

be automated when they follow the simple rule of adding an estimation or prediction residual e_t to subsequent (possibly but not necessarily all) forecasts $\hat{f}_{t+1}, \hat{f}_{t+2}, [\dots]$. “[S]etting the model back on track” is the informal label they provide for that procedure. In ?, 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. The term intercept correction itself can be both a bit deceptive in limiting, in that it need not pertain to intercepts. ? themselves note the possibility of correcting a coefficient in a forecast model specification, and ? developed a similarity-based procedure for correcting β using past information of the time series itself. For a review of more recent work in similarity-based forecasting, we refer the reader to ?.

What if the break in the DGP is brought about by a very particular kind of event, a news shock? What if we could predict well the intercept shift that occurs between T^* and $T^* + 1$? In ?, the authors list six necessary conditions for reacting to what we in this current paper call a news shock. Of greatest interest to us are two of the six: “the forecasting model already embodies that source of information” and “there is an operational method for selecting an appropriate model”. For such cases, forecasting procedures have been explored in ??, 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.

As the literature has developed, a current practitioner of forecast adjustment can choose between (i) procedures that are discretionary or automated, (ii) whether the data is internal (i.e. from the time series itself) or external, (iii) the term to be corrected (e.g. intercept, coefficients), if any, (iv) as well as the correction function (i.e. the mapping from the donor unit data to the correction term in the time series under study), including the weighting applied to the assembled data (e.g. Nearest-Neighbor, arithmetic mean, kernel methods). We discuss this in depth in Section 4.

The procedure presented, post-shock forecasting, 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 the notion of similarity between breaks in different time series, it and borrows from the causal inference literature to aggregate information from those similar circumstances. Beyond of ??, we are not aware of prior work uses a parametric specification for nonzero errors as well as a method for aggregating similar shocks occurring outside the time series under study.

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 History of and Motivation for Intercept Correction

See Clements and Hendry - Setting *levels* back on track. It may make less sense for flows.

Intercept correction is the correction of a forecast by adding the residual at time t to the forecasts at some subset of $\{t + 1, t + 2, \dots\}$. We are doing something very close to that which is adding the predicted correction function output at time t to the forecasts at some subset of $\{T_1^* + 1, T_1^* + 2, \dots\}$, where T_1^* is the shock time of the time series under study. [?, p. 177] note that a break in the DGP (and hence the misspecification of the default forecasting method) is the “most obvious and best understood rationale” for intercept correction. To be sure, this manuscript will make its contributions largely in this

area. However, before we get there, let's follow Clements, Hendry and other authors in asking whether we would ever want to intercept-correct an unconditional forecast as opposed to the conditional. This is what Clements and Hendry call "Exploiting the information in unconditional forecasts". [?, p. 170-1] recall how unconditional forecasts can have lower forecast variance than conditional forecasts. There seems to be two different questions that we should distinguish: we forecast would minimize the conditional expectation of MSE, and what forecast would have the lowest variance.

Mismeasured data is discussed in [?, 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 \hat{y}_t -style prediction based on aggregation.

Yet another potential motivation springs from the way we evaluate model performance. Evaluating a Model by Forecast Performance ?

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 one 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. **Here is how that matters for this paper: if you have a family of models estimated using a Gaussian loss function, and the post-shock residuals are not conditionally Gaussian (e.g. they could be conditionally Cauchy), then why even start with the default forecasting function? Why not throw it out entirely? If only the distribution of the innovations changes, then would that even matter for a forecast that is purely concerned with the conditional mean?**

1.2 Forecasting Under Shocks

This section will absolutely need to draw on ? for the following reasons:

1. Necessary conditions for forecasting a break - do I agree with them?
2. Five kinds of available information - section 4 Available information
3. Distinction between regular forces J and its complement, K
4. "forecasting facing a shift in the intercept of a model of a variable y induced by another observable variable z ;" — essentially, this is what we do in ? except z is not quantitative.

In ? (cited in ?), the authors note that when forecasts fail, it often has to do with location (i.e. intercept shifts), and therefore using additional information is not likely to help unless it can address potential location shifts. ? says that breaks in a model's parameters are not necessary or sufficient to generate time variation in forecasting performance.

? discuss that intercept correction can improve forecasts under structural breaks as well as under "deterministic misspecification", by which they mean misspecification not caused by stochastic events like location shifts. The same two justifications exist for SPC: we can use SPC when we know there to be a structural break, and we can also use it when we know our model is wrong and want to correct a parameter for some critical forecasting task.

1.3 Forecasting Using Outside Information

When is more information better? That is a central question with regard to forecasting. For example, amidst a shock, should we avail ourselves of additional internal information or external information?

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

Are there cases where it is ambiguous whether information should count as internal or external to the time series under study?

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.

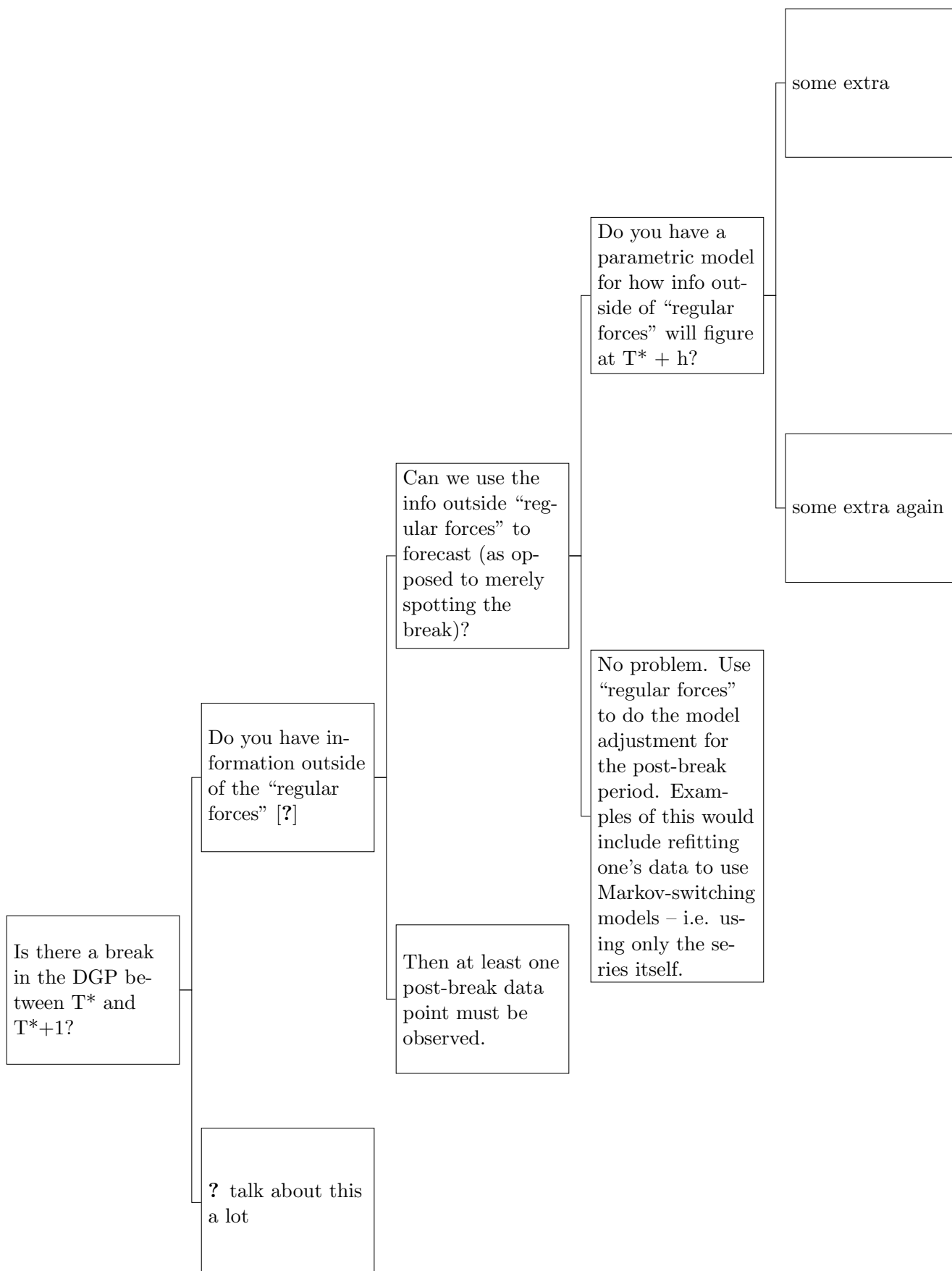
The draft here transitions to similarity, since the manuscript under development is at the intersection of post-break forecasting and similarity-based forecasting. In particular, it uses the former to motivate the latter. Now we discuss The Meaning and Use of Similarity Hume? Cite Israeli authors, if 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. Consider an asymmetric Jaccard Index $\delta_{\mathcal{J}}(\cdot, \cdot)$ that maps from any two finite-cardinality sets to the interval $[0, 1]$ [?]. So far, we have discussed only binary measures of similarity. The distance-based weighting scheme found in ? and repurposed in ? can be thought of as an n -ary similarity measure, where the n weights means little in isolation but can tell a greater story when viewed alongside the approximation they produce.

What about qualitative ways of determining similarity? In ?, 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

In order to introduce the general framework for adjusting forecasts under shocks, we provide an illustration. In Figure 1, we borrow from ? to 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.

In this work, we will be focused on shocks that occur strictly between two discrete time points, or informational available at a “fractional lag” [?]. This raises an important terminological point. Errors are a staple of stochastic modeling, and certainly no proper inventory of the wide range of uses of “noise”, “shocks”, and so can be accomplished here. Whereas in the psychometric and social-scientific fields, errors may represent unmeasured or latent capacities of a unit, in other fields like the natural sciences, errors may be used to account for variability in the instruments used to collect data. “Innovations” or “shocks”, in econometric literature, are generally understood to represent the arrival of previously unknown (but not necessarily unanticipated) information. Shocks are called structural when they can be linked back to some key feature of a scientific (usually economic) theory; otherwise, they are idiosyncratic. However, there exist other subcategories of shocks that crosscut the distinction just made. For example, a news shock is a shock in which the information material to the market is delivered via newsmedia and moves markets no earlier than the news. Shocks can also be permanent or transitory, supply-related or demand-related, and so forth. The taxonomy is vast. What matters for this work is the news shocks we posit occur between two discrete time points, so that a practitioner is faced with the question of what action is most advisable for the following time point, given the news shock.

2.1 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

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

$$\omega_{i,t}^* = D_{i,t}^{vol} [\mu_{\omega^*} + \delta^T \mathbf{x}_{i,T_i^*+1} + u_{i,t}], \quad (2)$$

$$\mathcal{M}_1: y_{i,t} = F(\mathcal{F}_{i,t-1}) + \mathbf{x}_{i,t}^T \lambda_t + \epsilon_{i,t}$$

with time-invariant error structure

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

where F maps objects from the past into future objects of $y_{i,t}$, and $\epsilon_{i,t}$ 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

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

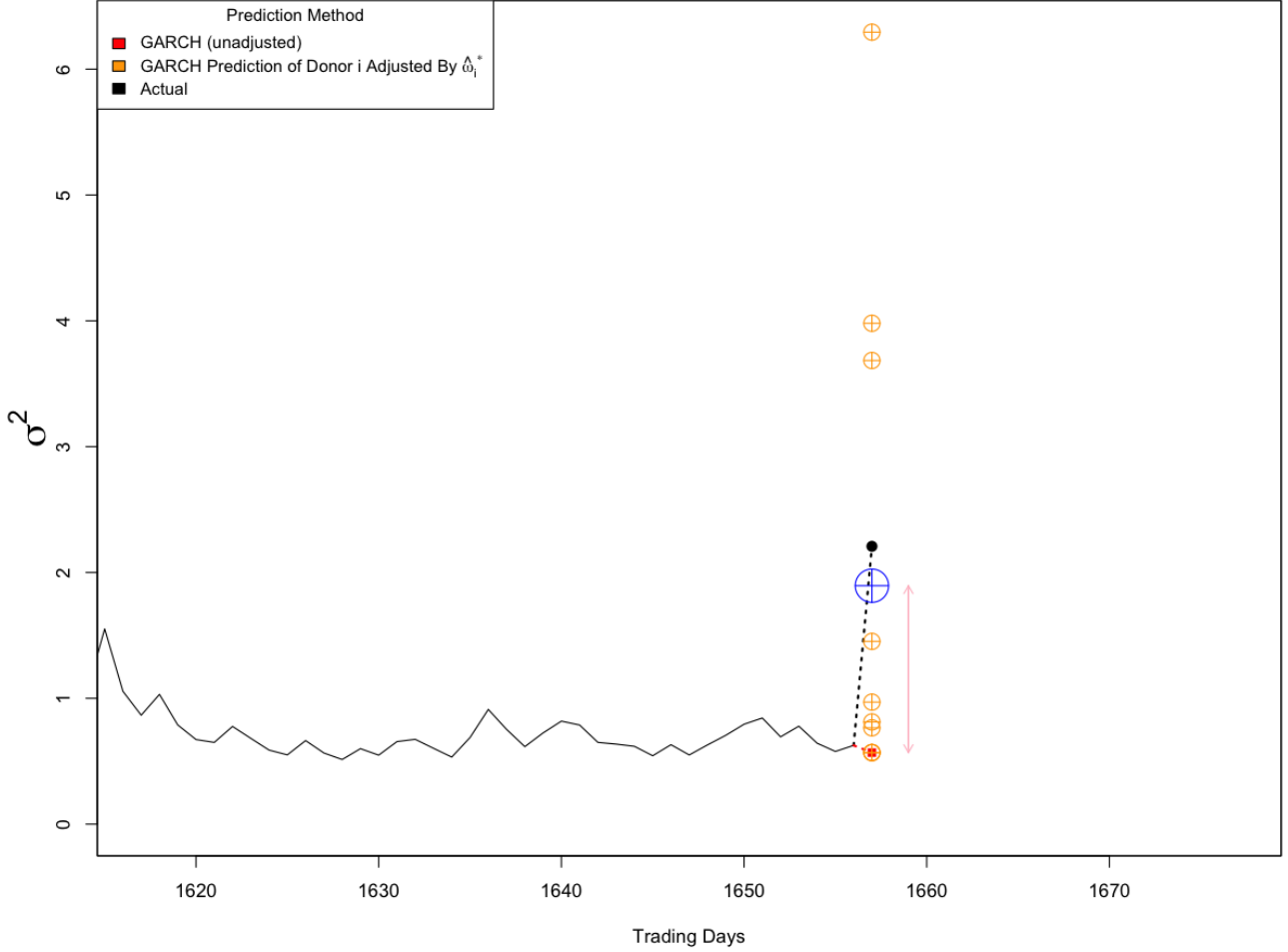


Figure 1: The simulated time series experiences a volatility shock between trading days 1,656 and 1,657. The GARCH prediction, in red, fails even to approach the volatility spike at $T^* + 1$, as do several adjusted predictions, which are orange. In contrast, the GARCH forecast adjusted by $\hat{\omega}^* = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$, 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.

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

$$\xi: \mathcal{F}_{2,T_2^*} \times \dots \times \mathcal{F}_{n+1,T_{n+1}^*} \rightarrow \mathcal{F}_{1,T_1^*+h}$$

that estimates or predicts $\mathbf{x}_{1,T^*+h}^T \lambda_{T^*+h}$ 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 \mathcal{F}_{t-1} , 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 Section 2.

3 Methodology for Similarity-based Parameter Correction

3.1 Forecasting and Correction Functions

We now present two one-step-ahead forecasts for the time series under study. The first is the unadjusted forecast. The second is the adjusted forecast, which differs by the predicted correction term. These forecasts are:

$$\begin{aligned} \text{Forecast 1: } \hat{y}_{unadjusted, T_1^*+1} &= \hat{\mathbb{E}}[\mathcal{F}_{T_1^*}] \\ \text{Forecast 2: } \hat{y}_{adjusted, T_1^*+1} &= \hat{\mathbb{E}}[\mathcal{F}_{T_1^*}] + \hat{\omega}^* . \end{aligned}$$

The problem of aggregating shocks or residuals begins with the data constraints. Let us first introduce useful notation. Let $\hat{\omega}_{i,*}^*$ denote estimated correction term for donor i . Taking the estimated correction terms as a given, we observe the tuple of information $(\{\hat{\omega}_{i,*}^*\}_{i=2}^{n+1}, \mathcal{F}_{2,T_2^*} \times \dots \times \mathcal{F}_{n+1,T_{n+1}^*})$ that is the extracted residuals as well as information important to aggregation.

One such idea comes from ?, where weights belonging to the simplex, $\{\pi_i\}_{i=2}^{n+1} \in \Delta^{n-1}$, are chosen via a distance-based weighting procedure, resulting in the correction function

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

which is just one way among many to build a correction function. Another interesting correction function can be found in ?. For yet another example, consider a correction function that aggregates survey panel of forecasts by the Survey of Professional Forecasts [?]. An obvious advantage of human forecasters is that they are able to integrate information revealed arbitrarily close to the scheduled release of the forecast, even if human forecaster’s lack of systematicity or closed-form forecasts, for example, leads to non-optimal forecast performance in general. By calculating forecast differences between the low-frequency information forecast, $\tau_{i,t} = \hat{y}_{unadjusted,t} - \hat{y}_{i,t}$, we prepare ourselves for late-breaking events – e.g. shocks – for which we might want a correction function

$$\xi: \tau \rightarrow \mathcal{F}_{1,T_1^*}$$

that weights the i -th forecaster according to that forecaster’s past performance or according to the similarity of the forecaster’s stated information set to the information available for the default forecasting model.

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

In this section, we introduce and discuss a general framework for model adjustment that generalizes and is motivated by the circumstances laid out in Section 2 by boiling down SPC to its essential five elements. The essential elements of similarity-based parameter correction are

1. **Object-to-predict** 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 multidimensional objects 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 $e_t = y_t \odot \hat{y}_t$, where \odot is subtraction in the simple case of mean squared-error. The residuals (or transformations of those residuals) of those n models can be weighted as part of a correction function. This fact is especially useful when the forecast function f is nonparametric, black-box, or stochastic with respect to \mathcal{F}_t .

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 ??.
3. **Reliable and Shared Model-Fitting Procedure** 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.
4. **Reliable Correction Term Estimation** Fourth, there must exist a reliable procedure for modeling and estimating 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. Alternatively, the correction term could be an inner product that depends upon external covariates, as is found in ??.

This might not always be straightforward. Some models like GARCH, for example, might deliver very noisy estimates for indicator variables that occur just once.

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 GDP. Similarly, if there exists a posited DGP for the correction term, it may depend upon data internal to the time series under study, or it may depend on external data, e.g. outside data could be aggregated (e.g. Nearest-Neighbor, arithmetic mean, kernel methods) to estimate the correction function.

5 Formal Properties and Model-Specific Considerations

In this section discuss particular models in which our approach has been implemented as well as other in which it has yet to be implemented. For those for which an implementation exists, we portray the model in a more general light, while also commenting on the model-specific considerations.

Recall the specification 1, which models a rare shock as an affine function of covariates:

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

5.1 ARIMA

We begin our recounting of model-specific cases by recalling ?, in which it was established that for a family of AR(1)-distributed scalar time series with a common shock distribution, forecast risk can be reduced via a similarity-based adjustment procedure. AR(p) is an easy model to work with, in part because it can be consistently estimated via OLS, and an AR(p) can approximate an ARMA(p,q) with arbitrarily-small error through via a truncation of an AR(∞) representation. Setting aside the estimation method, consider a one-step-ahead for an AR(p) at time t with an adjustment estimator, $\hat{\alpha}_t$:

$$\hat{y}_t = f(\mathcal{F}_{t-1}) = \hat{\mu} + \sum_{i=1}^p \hat{\rho}_i y_{t-i} + \hat{\alpha}_t \quad (3)$$

which can be understood as an AR(p) with time-varying coefficients. The mean squared-error of the forecast above, i.e.

$$\mathbb{E}[(y_t - \hat{y}_t)^2] = \mathbb{E}[(y_t - \hat{\mu} + \sum_{i=1}^p \hat{\rho}_i y_{t-i} + \hat{\alpha}_t)^2] \quad (4)$$

$$= \mathbb{E}[(\mu + \sum_{i=1}^p \rho_i y_{t-i} + \alpha_t - \hat{\mu} + \sum_{i=1}^p \hat{\rho}_i y_{t-i} + \hat{\alpha}_t)^2] \quad (5)$$

admits of the typical bias-variance decomposition, with the special proviso that with more terms in the prediction function, more covariance terms arise. Similarly, the covariance structure of the random effects in $\alpha = \mathbf{x}_t^T \lambda_t$ will contribute to the variance. In the real-data examples ?, the entries in λ_t are proxies of general market risk, which one would commonly take to be correlated not only with each other but also with a financial time series y_t . All of this is to say that control of the MSE may require tight control of the variance in the prediction function, which in turn will gesture in the direction of a sparse $\hat{\alpha}_t$.

When forecasting financial time series using the method above, the question of prices versus returns (or in the language of economics or engineering, levels versus flows) is relevant. Consider, for example, the persistence of a shock. A shock to an asset's return series at time t implies a persistent shock to the asset's price series at time t and beyond, as one can verify by inverting the discrete-differencing operation on a time series. However, if one forecasts the price series of the asset directly, that persistence beyond the first instantiation of the shock must be modeled explicitly. This suggests interesting opportunities in the analysis and forecasting of shock persistence for situations in which level-forecasting is appropriate, e.g. for stationary time series.

We now turn to the limitations of linear time series model class like ARIMA. The first is that linearity is a nontrivial restriction, and for the purpose of modeling financial time series, applications of SPC to TAR and SETAR would be welcome. Additionally, the efficient market hypotheses implies that market prices reflect all publicly-available information, which in turn implies that ARIMA models should be of no predictive value for market returns. This critique, in a technical sense, alleges that market returns

are random walks centered at zero, and for that reason, the coefficients of an ARMA approximation to a financial return series are uniformly zero, and that any incipient deviation from that equilibrium would not go unnoticed by market participants, who would exploit those deviations and push the coefficients back to zero, leaving only idiosyncratic noise. Such concerns about forecasting in markets with heavy feedback loops motivates our next section.

5.2 GARCH

The methods of ? are adapted in ? to forecast volatility. This is in response to concerns we raised above about the usefulness of ARIMA models for forecasting financial returns as well as concerns about heteroskedasticity. Financial time series are known to have both time-varying volatility but also volatility that *clusters*. A GARCH model, under weak conditions, is an ARMA model on the squared residuals of a time series, and empirical evidence strongly suggests that those ARMA (i.e. GARCH) coefficients are not zero, in general. The predictability of either those squared returns or σ_t^2 (in this context, the two tasks reduce to one) is bounded, however, by a function of the kurtosis of a GARCH process [?]. That the GARCH model accommodates excess kurtosis in both the unconditional and conditional returns of a financial asset is a feature of the model, and perhaps for that reason, it has been noted that GARCH models do a much better job at modeling the evolution of volatility (a descriptive or inferential objective) as opposed to forecasting volatility.

In this way, ? bridged a gap between the descriptive and predictive roles of GARCH. Fixed effects (correction terms) are estimated in the donor pool, and those fixed effects are aggregated via corrected function. A forecast adjustment is then made in the time series under study.

5.3 HAR

The Heterogeneous Autoregressive (HAR) model for realized volatility [?] uses ordinary least-squares (OLS), WLS, or even a logarithmic specification to estimate an autoregressive model. It explores influence of previous volatility on current volatility, of course, but uses a specification motivated by theory: that a heterogeneous collection of market participants interacts at different frequencies. Therefore, the volatility can be well-modeled as a linear combination of the average volatility over the previous day, previous week, and previous month (usually taken to be 22 days). Formally, HAR assumes that the log price of an asset is governed by a diffusion process (a model beyond our scope herein), and HAR estimates

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

5.3.1 A Real Data Application of Shock-Adjusted HAR Forecasting

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 SPC 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 ?.

We borrow code from ?

- [LSTM in Python](#)

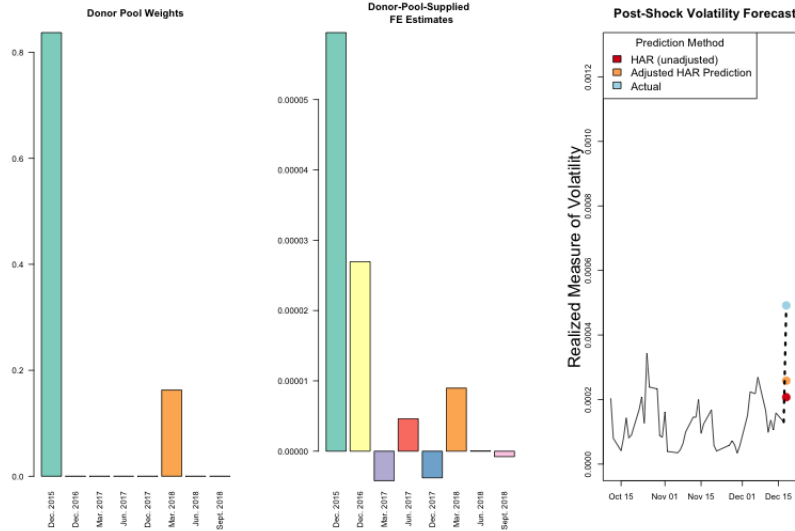


Figure 2: Over the course of the 2010s, the US FOMC began a rate-hiking cycle. Left pane: weights of the donor pool of FOMC meetings; right-pane: fixed effect estimates; right pane: the unadjusted forecast, adjusted forecast, and actual forecast

5.6 The Method of Multiple Shocks, Dual Shocks, and Shock-Partitioning

Sometimes, we might have reason to believe that a forecast requires adjustment due to not one but multiple shocks. ? 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 Discussion

The forecast horizon — does it matter? If so, how so? ? has a one-period horizon. ?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.

6.1 Shock propagation and estimation accuracy

Consider ?, 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.

6.2 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

7 Extensions

bias-variance tradeoff and MSE decomposition

- Forecast for series besides the mean or conditional mean **If you have a family of models estimated used a Gaussian loss function, and the post-shock residuals are not conditionally Gaussian (e.g. they could be conditionally Cauchy), then why even start with the default forecasting function? Why not throw it out entirely? If only the distribution of the innovations changes, then would that even matter for a forecast that is purely concerned with the conditional mean?**
- Function forecasts
- Binary Outcome Forecasts
- Density Forecasts
- Quantile Forecasts

7.1 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 ?.

Consider the leave-one-out method used on pairs of donors and covariates in ?. By averaging over $(n + 1) \cdot (p + 1)$ predictions, we average over $n + 1 \cdot p + 1$ convex combinations (where, when a donor is omitted, its weight can be imputed as zero).

7.2 Limitations

7.2.1 How important is a shared DGP?

In particular, can we relax this assumption and proceed with weaker assumptions on the donors and time series under study?

7.2.2 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.

In Scope

- Introduce an existing diffuse set of approaches to adjusting forecasts (1)
- Explain when/how similarity can help us forecast. 1.3
- 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)

Beyond Scope (but mentioned briefly)

- Wade too deeply into distance-based weighting details. State that it's simply one way to do it. Allude to ??.
- wade too deeply into any of the special cases
- Tackle non-scalar random quantities (density forecasts, etc)
- Forecast combination