Forecast Adjustment Under Shocks: A Unification

David P. Lundquist,[*] Daniel J. Eck[†]

Department of Statistics, University of Illinois at Urbana-Champaign

November 25, 2024

Abstract

Situations arise in which one's default times series forecasting model must be adjusted, augmented, or abandoned completely, generating second-order questions about algorithmic design and the role of human judgment, as well as the information leveraged for forecast adjustment. This work further 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

In time series and panel data, sometimes the salient challenge is not predicting when an event will occur or begin 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, or 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. The event might not even be conceivable to many before it happens, as was the case with the COVID-19 pandemic. On the other end of the spectrum are events whose arrival need not be predicted because they are scheduled in advance, e.g. scheduled news, like releases of economic data or election results. In those cases, conditional forecasts allow a statistician to proceed as if the event has arrived. Consider the analysis of Mike Wilson, Morgan Stanley's CIO and Chief US Equity Strategist, in October 2024 regarding the upcoming 2024 US Presidential US [Wilson, 2024]:

> While some argue a Trump win would be a headwind for growth and equity markets, due to tariff risks and slowing immigration, we think there's an additional element from the 2016 experience that is worth considering—rising animal spirits. More specifically, in 2016 Trump's pro-business approach led to the largest 3-month positive impact on small business confidence in the past 40 years. It also translated into a spike in individual investor sentiment. It appears to me that markets may be trying to front-run a repeat of this outcome as Trump's win in 2016 came as a surprise to pundits and markets alike.

Wilson here assesses how market participants are incorporating current information as well as past forecast failures, including the potential overweighting of more recent failures. His assessment implies that the

---

[*]davidl11@ilinois.edu

[†]dje13@illinois.edu

shock conditional on a Trump victory may disappoint observers, as it has already been pulled forward into current forecasts. In short, this time around may be different. Of course, it may be that participants have merely pulled *too much* forward by correctly assessing the conditional impact of a Trump victory but overestimating its likelihood, or vice-versa. Regardless, what Wilson is engaging in is what Lundquist and Eck [2024] identify as a crucial reaction to uncertain territory: what, if anything, does it resemble from the past?

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. We in fact 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 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 received its earliest attention under the auspices of "intercept correction" [Hendry and Clements, 1994, Clements and Hendry, 1996, 1998, 1999]. Of fundamental 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 or prediction residual e_t to subsequent (possibly but not necessarily all) forecasts $\hat{f}_{t+1}, \hat{f}_{t+2}, [...]$. "[S]etting the model back on track" is the informal label they provide for that procedure.

The term intercept correction can be construed more broadly to encompass correction of terms besides the intercept, and justification for it can be found beyond the original goal of setting forecasts back on track. In later work, Clements and Hendry [1999] present seven additional interpretations for intercept correction, and Hendry and Clements [1994], Clements and Hendry [1999] themselves note the possibility of correcting a coefficient in a forecast model specification. Guerrón-Quintana and Zhong [2017] develop a similarity-based procedure for correcting the parameter 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 Lundquist and Eck [2024] and references therein.

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 Castle et al. [2016], 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 intercept shifts that occurs between T^* and $T^* + 1$, forecasting procedures have been explored in Lin and Eck [2021], Lundquist and Eck [2024], where the AR(1)-X and GARCH(m, s) cases, respectively, are treated. Both works target additive parameters in scalar time series, predicting the additive parameter in the time series under study by aggregating information from other time series. The authors leave several stones unturned, including a more general, dare say comprehensive treatment of forecasting under shocks.

As the literature has developed, a current practitioner of forecast adjustment can now choose (i) procedures that are discretionary or automated, (ii) adjusting using internal data (i.e. from the time series itself) or external data, (iii) the term to be corrected (e.g. intercept, coefficients), if any, (iv) as well as the specification of the correction function (i.e. the mapping from the donor unit data to the correction term in the time series under study), including the aggregation mechanism applied to the assembled data (e.g. Nearest-Neighbor, arithmetic mean, kernel methods). We discuss this in depth in Section 5.

Figure 1: Forecast Model Adjustment: A Decision Tree

If intercept correction is the correction of a forecast by adding the residual at time $T^* + 1$ to the forecasts at some subset of $\{T^*+2, T^*+3, ...\}$, in this manuscript, we are doing something rather close to that by adding the predicted correction term to each forecast in some some subset of $\{T_1^*+1, T_1^*+2, ...\}$, where $T_1^* + 1$ is first integer index for the time series under study following the shock. The particular 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. It is a discretionary procedure in the sense that substantive choices must be made at each instance of use. The correction function uses the notion of similarity at multiple levels: first, to identify similarity between breaks in different time series, and second, in borrowing from the causal inference literature to aggregate information from those similar circumstances. Parametric specifications for location shifts in the DGP can be found in Castle et al. [2011], where locations shifts are modeled as induced by additional observable variables, as well as in Lundquist and Eck [2024], where the arrival of the news shock is observable, even if its quantitative instantiation is not. Beyond Lin and Eck [2021], Lundquist and Eck [2024], we are not aware of prior work that proposes a method for aggregating similar shocks occurring outside the time series under study in order to correct the time series under study.

The structure of this manuscript is as follows. We provide a deeper and more critical review of several 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.

2   Forecasting Amid Shocks

[Hendry and Clements, 1994, p. 177] characterize breaks in the DGP (and hence the misspecification of the default forecasting method) as the "most obvious and best understood rationale" for intercept correction. In this section, we wish to locate the task of forecasting amid shocks within the broader literature on forecasting under breaks in the DGP.

Some terminological notes are in order before we proceed, however. Errors are a staple of stochastic modeling. They go by the synonyms "noise", "shocks", or "innovations". The meaning of these terms may not differ as much in their mathematical representations in symbols but in how they are interpreted. Whereas in the psychometric and social-scientific fields, errors may represent unmeasured or latent ca-

pacities 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 uncontemplated) information. Shocks are called structural when they are can be linked back to some key feature of an scientific (usually economic) theory; otherwise, they are idiosyncratic. Additionally, 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 release itself. Shocks can also be permanent or transitory, supply-related or demand-related, and so forth. The taxonomy is vast. We use the term "shock" as a special case of a break in the DGP, even though, strictly speaking, some shocks will neither result from a break in the DGP nor herald the emergence of a new regime. The news shocks examined in this manuscript, however, should be considered breaks in the DGP, no matter whether they are transient or permanent. Additionally, in this work, we will be focused on news shocks that occur strictly between two discrete time points, or information available at a "fractional lag" [Castle et al., 2011], 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 The Role of Outside Information

When is more information better? That is a central question with regard to forecasting. In Clements and Hendry [2005b] (cited in Castle et al. [2013]), the authors note that when forecasts fail, it often has to do with location (i.e. intercept shifts). Therefore, using additional information is not likely to help unless it can address potential location shifts. We should clarify what we mean by "additional information". Do we mean any information not included in the default forecast function? Could that include something as straightforward as additional lags of the time series being forecast? Perhaps a more useful distinction to make is that between information internal to the time series being forecast and information external to the time series being forecast. Of course, the usefulness of this distinction will hinge, in part, on whether we can cleanly partition information into internal and external buckets. A cross-cutting distinction originates in Castle et al. [2011], which discusses necessary conditions for forecasting under breaks as well as a taxonomy of information available for doing so. The authors make a fundamental distinction between "regular forces", which includes the typical standbys of economic theory, and information potentially relevant to location shifts.

We wish now to sharply distinguish between essential and inessential concepts. For example, whether information is relevant to location shifts or not is a worthy and complex question that depends upon numerous factors, including but not necessarily limited to the DGP. We prefer not to conflate that distinction with the internality-externality distinction, or whether the information belongs to the traditional sort of information used in econometric modeling, though certain tendencies and general patterns may emerge. We make clear what is essential in Table 1.

Table 1: A 2x2 Schema of Forecast Information, With Examples

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

2.2 The Meaning and Use of Similarity

This manuscript aims to contribute at the intersection of post-break forecasting and similarity-based forecasting. In particular, we use the former to motivate the latter. 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. Similarity-based reasoning can also be given a philosophical treatment when viewed as a special case of analogical reasoning [Bartha, 2013], from which we can glean a useful general framework in which reasoning by analogy involves a source domain and a target domain. Our target domain, simply put, contains the thing we'd like to know, while the source domain is our reservoir of information from which to draw.

Our objective as statisticians is to make forecast corrections by exploiting information from previous time series that have undergone shocks under similar circumstances. As a provisional definition, we say that a break in the DGP of a time series is similar to a break in another just in case the causes of the breaks share key, material qualities. Note here that we have not said anything about the time series themselves (the details of their DGPs or their moments) being similar, nor the entities that they represent (e.g. the firms, the financial instruments, the governments, etcetera) being similar. Why that is so will become clear in Section 3.

Having provided only a provision definition, we would like to introduce useful distinctions regard the concept of similarity. First, quantitative ways of determining similarity naturally tend toward the idea that similarity is reducible to determining what serves as a best approximation. Those approaches include matching a target object to other objects along one or more variables, where unique matches or exact matches may be likely or may be statistically certain, likely, unlikely, or impossible, subject to the setting. In metric spaces, the notion of a metric that quantifies nearness is crucial to fundamental concepts of proximity. This suggests a set of optimization problems such that for any target object Q, the search for similarity means a search for an approximator or collections of approximators A such that $\delta(A, Q)$ is smallest, where δ is an appropriate metric.

The notion of approximating a target does not necessarily mean that the target need be similar to the approximator. This suggests a role for asymmetric distance functions δ (which fail to meet the methematical requirements of a metric) for exploring differences between donors and weight their contributions accordingly. Consider an asymmetric Jaccard Index $\delta_{\mathcal{J}}(\cdot, \cdot)$ that maps from any two finite-cardinality sets A, B to the interval $[0, 1]$ [Garg et al., 2015], penalizing A more for its incongruence with B than B is penalized for its incongruence with A. Asymmetry also plays an important role in weighting schemes far beyond traditional forecasting, including Transformer mechanisms inside the field of deep learning. The asymmetry inherent in query, key, and value matrices allow hierarchical relations in language to be represented mathematically [Bishop and Bishop, 2023, p. 364].

So far, we have discussed only binary measures of similarity. The distance-based weighting scheme found in Lin and Eck [2021] and repurposed in Lundquist and Eck [2024] 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. However, an n-ary similarity measure will not always reduce to the task of constructing an approximator to the target using n original objects. We might instead insist on a notion of similarity such that n objects are said to be similar, as a collection, to the target if and only if each of the n matches the target perfectly.

What about qualitative ways of determining similarity? In Lin and Eck [2021], 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 methods. The role of the covariate vector in their models is two-fold: the covariates appear in the shock distribution, and the covariates also help determine which donors are most important. It is to a generalization of this procedure that we now turn.
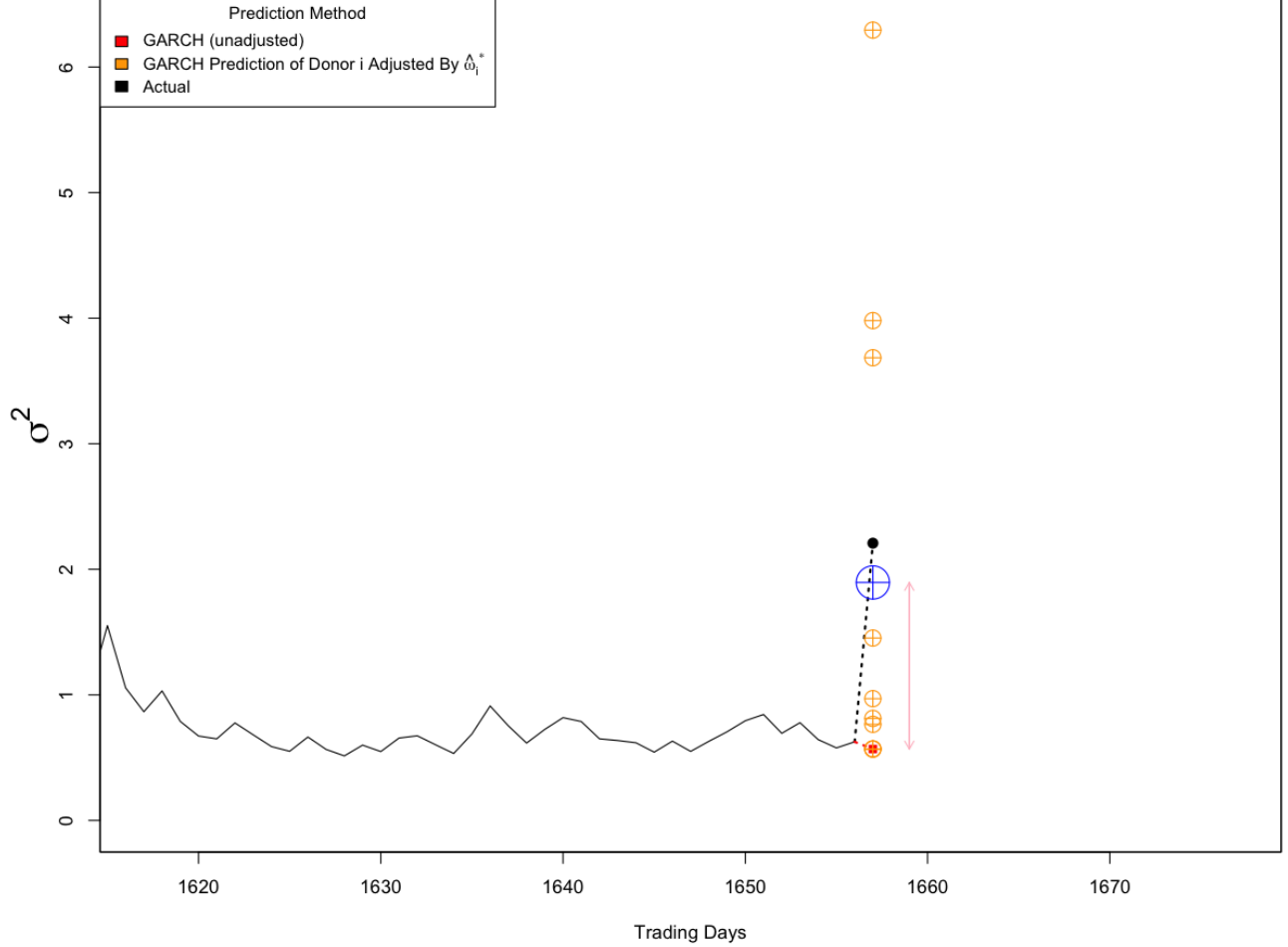
3 Setting

Figure 2: 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.

We will suppose that a researcher has a multivariate time series data $\mathbf{y}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n+1$, where $y_{i,t}$ 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 $y_{1,t}$, 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 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 which is not yet known.

We assume the following general data-generating process

$$\mathcal{M}_1: \begin{aligned} y_{i,t} &= F(\mathcal{F}_{i,t-1}) + \alpha_{i,t} + \epsilon_{i,t} \\ \alpha_{i,t} &= \mathbf{x}_{i,t}^T \lambda_{i,t} \\ \mathbf{x}_{i,t}^T &= (1, x_{i,t}^1, ..., x_{i,t}^p) \\ \lambda_{i,t}^T &= (u_{i,t}, \lambda_t^1, ..., \lambda_t^p), \end{aligned} \tag{1}$$

with time-varying and observable covariate vector $\mathbf{x}_{i,t}^T$, time-varying and unknown parameter vector

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

and time-invariant error structure

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

where F maps objects from the past into future objects of $y_{i,t}$, and $\epsilon_{i,t}$ uncorrelated across time and donors. Note that the dot product $\alpha_{i,t} = \mathbf{x}_{i,t}^T \lambda_{i,t}$ includes the term $u_{i,t}$ that is not shared among donors. It encodes the known possibility that the shocks will vary across donors both in ways that can predicted based on pre-shock covariates as well as in ways that cannot be predicted. That the parameter vector is time-varying allows us to capture that most time points are without news shocks (in which case $\alpha_{i,t}$ is of neglible effect), but conditional upon information arriving between T_1^* and $T_1^* + 1$, we may know with near certainty that λ_{1,T_1^*+h} will be nonzero in norm for some $h > 0$. Note that the specification (1) above is sufficiently capacious as to include data-generating processes of the form

$$y_{i,t} = \log Y_{i,t} = \log a_{i,t} y_{i,t-k} e^{\alpha_{i,t}+\epsilon_{i,t}} = \log a_{i,t} + \log y_{i,t-k} + \alpha_{i,t} + \epsilon_{i,t}$$

with $k \geq 1$ and $a_{i,t}, y_{i,t}$ supported on $\mathbb{R}^+$.

The term $\mathbf{x}_{i,t}^T \lambda_{i,t}$ will be deemed the *correction term* for donor i at time t, and the function

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

that estimates or predicts α_{1,T_1^*+h} will be deemed the *correction function*. The correction function maps donor information — be it observables or inferences — to the prediction function for the time series under study.

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 3. What we are doing is distinct in at least two ways. First, we are incorporating outside information faster, and second, 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.

4 SPC Forecasting Methodology 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. The forecasts are:

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

The problem of aggregating estimated shocks or residuals begins with the data constraints. Let us first introduce useful notation. Let $\hat{\alpha}_{i,*}$ denote estimated correction term for donor i. Taking the estimated shocks or residuals as a given, we essentially observe the tuple of information $(\{\hat{\alpha}_{i,*}\}_{i=2}^{n+1}, \mathcal{F}_{2,T_2^*} \times \ldots \times \mathcal{F}_{n+1,T_{n+1}^*})$. One approach to aggregation comes from Lin and Eck [2021], 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

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

which is just one way among many to build a correction function. Another interesting correction function can be found in Foroni et al. [2022], where percentage residuals of past forecasts are used. For yet another example, consider a correction function that aggregates the survey panel of forecasts by the Survey of Professional Forecasts [Croushore, 1993]. 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

$$\xi \colon \boldsymbol{\tau} \to \mathcal{F}_{1,T_1^*}$$

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.

4.1 Diagnostics for Similarity-Based Parameter Correction

The assumptions we make on the DGP for shocks to both donors and the time series under study allow us to evaluate our prediction tool even before we predict. In this subsection, we will elaborate on a permutation method proposed in Lin and Eck [2021]. The idea is simply to use each of the n donors as an object to predict, while the remaining $n-1$ serve as actual donors. Permutation tests for inference were proposed as early as Abadie et al. [2010] in the context of synthetic control. Although we will not embark upon a formal exploration of the permutation method for predictive model evaluation, we will note briefly its motivation, chiefly that the donors share a DGP and the model particulars are invariant to permutation. To the extent that those assumptions hold in practice, model performance can be gauged by methods as simple as evaluating the losses of the 'pseudo-time-series-under-study' to the more sophisticated, like assessing the correlation between the goodness of approximation and the losses. One can also assess the plausibility of the $\{\hat{\alpha}_i\}$ being uncorrelated, which is central to MSE-minimization, as we will see in Section 6.1.

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

In this section, we introduce and discuss a general framework for model adjustment that both generalizes and is motivated by the circumstances laid out in Section 3 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 and possibly space, as well) that obeys a specification with additive errors, or, at the very least, a specification that can be transformed to have additive errors. 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 Lin and Eck [2021], Lundquist and Eck [2024].

3. **Reliable and Shared Model-Fitting Procedure** There must exist a reliable model-fitting procedure for the $n + 1$ units, one that will allow the assumption of a shared family for the residuals to be reflected in the residuals. Here, reliable could mean any number of 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 Lin and Eck [2021], Lundquist and Eck [2024].

   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 DGP. 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.

5.1 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 $j = 2, ..., n_j+1$, and 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: Each of the k shocks can be represented using an 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.

An assumption of independence or at least uncorrelatedness among the k donor pools is helpful here.

6 Formal Properties and Model-Specific Considerations

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

6.1 ARIMA

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

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

We begin our recounting of model-specific cases by recalling Lin and Eck [2021], in which it was established that for a family of AR(1)-X-distributed scalar time series with a common shock distribution, forecast risk can be reduced via a similarity-based adjustment procedure. An 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, $\hat{\alpha}_t$:

$$\hat{y}_t = f(\mathcal{F}_{t-1}) = \hat{\mu} + \sum_{k=1}^{p} \hat{\rho}_k y_{t-k} + \hat{\alpha}_t \tag{2}$$

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

$$\mathbb{E}[(y_t - \hat{y}_t)^2] = \mathbb{E}[(y_t - \hat{\mu} + \sum_{k=1}^{p} \hat{\rho}_k y_{t-k} + \hat{\alpha}_t)^2] \tag{3}$$

$$= \mathbb{E}[(\mu + \sum_{k=1}^{p} \rho_k y_{t-k} + \alpha_t + \epsilon_t - (\hat{\mu} + \sum_{k=1}^{p} \hat{\rho}_k y_{t-k} + \hat{\alpha}_t))^2] \tag{4}$$

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

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

admits of the typical bias-variance decomposition. If it is known that $\mu = 0$, (5) becomes

$$\mathbb{E}[(\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t) + \epsilon_t)^2]$$

$$= \text{Bias}^2[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \sigma_\epsilon^2 \qquad (\epsilon_t \text{ idiosyncratic})$$

$$= \left[\mathbb{E}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)]\right]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + (\alpha_t - \hat{\alpha}_t)] + \sigma_\epsilon^2 \qquad (\text{Definition of Bias})$$

$$= \left[\sum_{k=1}^{p}(\rho_k - \mathbb{E}[\hat{\rho}_k])y_{t-k} + (\alpha_t - \mathbb{E}[\hat{\alpha}_t])\right]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k}] + \text{Var}[\hat{\alpha}_t] + \sigma_\epsilon^2$$

$$\text{(if donor shocks are uncorrelated with time series under study)}$$

$$= \left[\sum_{k=1}^{p}(\rho_k - \mathbb{E}[\hat{\rho}_k])y_{t-k} + (\alpha_t - \mathbb{E}[\hat{\alpha}_t])\right]^2 + \text{Var}[\sum_{k=1}^{p}(\rho_k - \hat{\rho}_k)y_{t-k} + \sum_{i=2}^{n+1}\pi_i^2\text{Var}[\hat{\alpha}_{i,t}] + 2\sum_{i \neq j}\pi_i\pi_j\text{Cov}[\hat{\alpha}_{i,t}, \hat{\alpha}_{j,t}] + \sigma_\epsilon^2$$

$$\tag{6}$$

Assume we have no control over the accuracy of the estimates $\{\hat{\rho}\}_{k=1}^{p}$ nor any control over the estimates $\{\hat{\alpha}_{i,t}\}_{i=2}^{n+1}$. Then our only control over the MSE comes comes from the weight vector $\vec{\pi}$, so long as it is not unique. If we insist that $\hat{\alpha}_t$ be an unbiased predictor, then the MSE will tend to have a sparse vector $\vec{\pi}$ as its minimizer, so as to zero-out the covariance terms. 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 $\hat{\alpha}_t$ and sparse estimation of the autoregressive structure.

How does the MSE of the adjusted forecast compare with the unadjusted forecast? Since the unadjusted forecast MSE would lack the final two sums in expression 6, the adjusted model therefore has lower MSE if and only if the reduction in the squared bias is greater than $\sum_{i=2}^{n+1}\pi_i^2\text{Var}[\hat{\alpha}_{i,t}] + 2\sum_{i \neq j}\pi_i\pi_j\text{Cov}[\hat{\alpha}_{i,t}, \hat{\alpha}_{j,t}]$.

Proposition 1. *Let $2 \leq i \leq n+1$. The tuple of estimators $(\hat{\rho}_1, ..., \hat{\rho}_p, \hat{\alpha}_{i,t})$ is consistent as $t \to \infty$.*

Proposition 2. *Let $\{\hat{y}_{1,T_1^*+r}\}_{r=1}^{h}$ denote the vector of adjusted predictions (adjusted through h steps ahead) in the time series under study. $\{\hat{y}_{1,T_1^*+r}\}_{r=1}^{h} \xrightarrow{d} \{y_{1,T_1^*+r}\}_{r=1}^{h}$.*

11

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.

It must be noted here that adding a past residual to future forecasts assumes implicitly that whatever caused the residual has some persistence. Hence, intercept correction might be thought more appropriate for time series of levels, or at least times series of rates that are not necessarily mean-reverting, like inflation. Of course, if one has a strong grasp of how long the intercept correction will be necessary, then all of that is moot, but in practice, knowledge of that persistence may be harder to come by than foreknowledge of the need for intercept correction.

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.

6.2 GARCH

The methods of Lin and Eck [2021] are adapted in Lundquist and Eck [2024] 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 as well as volatility that *clusters*, i.e. with volatile trading periods that bunch together. 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 the predictability of 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 [Francq and Zakoian, 2019]. 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, Lundquist and Eck [2024] 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.

Consider how GARCH fits into our Specification (1):

$$\sigma_{i,t}^2 = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \ .$$

Let $I(\cdot)$ be an indicator function. Let T_i denote the time length of the time series i for $i = 1, \ldots, n+1$, and let T_i^* denote the largest time index prior to the arrival of the news shock, with $T_i^* < T_i$, to ensure that

there is at least one post-shock realization for each series i. Let $\delta, \mathbf{x}_{i,t} \in \mathbb{R}^p$. Let $\mathcal{F}_i$ with a single-variable subscript denote a univariate, time-invariant σ-algebra, and $\mathcal{F}_{i,t}$ denote the canonical product filtration for donor i at time t. Let $D_{i,t}^{return} = I(t \in \{T_i^*+1, ..., T_i^*+L_{i,return}\})$ and $D_{i,t}^{vol} = I(t \in \{T_i^*+1, ..., T_i^*+L_{i,vol}\})$, and let $L_{i,return}, L_{i,vol}$ denote the lengths of the level and volatility shocks, respectively. For $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$, the model $\mathcal{M}_1$ is defined as

$$\sigma_{i,t}^2 = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} + \omega_{i,t}^*,$$

$$\mathcal{M}_1: \quad a_{i,t} = \sigma_{i,t}((1 - D_{i,t}^{return})\epsilon_{i,t} + D_{i,t}^{return}\epsilon_i^*),$$

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

with time-invariant error structure

$$\epsilon_{i,t} \overset{iid}{\sim} \mathcal{F}_\epsilon \text{ with } \mathrm{E}_{\mathcal{F}_\epsilon}(\epsilon) = 0, \mathrm{Var}_{\mathcal{F}_\epsilon}(\epsilon) = 1,$$

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

$$u_{i,t} \overset{iid}{\sim} \mathcal{F}_u \text{ with } \mathrm{Var}_{\mathcal{F}_u}(u) = \sigma_u^2,$$

$$\epsilon_{i,t} \perp\!\!\!\perp \epsilon_{i,t}^* \perp\!\!\!\perp u_{i,t}.$$

Let $\mathcal{M}_0$ denote the subclass of $\mathcal{M}_1$ models such that $\delta \equiv 0$. Note that $\mathcal{M}_0$ assumes that nonzero shocks have no dependence on the covariates and are i.i.d. with $\mathbb{E}[\omega_{i,t}^*] = \mu_{\omega^*}$, where the lack of indices i or t on μ_{ω^*} indicates that it is shared across donors and is time-invariant. Models $\mathcal{M}_1$ and $\mathcal{M}_0$ parameterize changing dynamics after a shock is observed (when $t \geq T_i^* + 1$ for each series i). It is important to note that our modeling setup will additionally suppose that $\mathbf{x}_{i,t}$ will be observed before $\sigma_{i,t}^2$ is observed.

Note also the divergences from Lin and Eck [2021]. As already established, the model presented here is intended to account for ARCH effects. Additionally, here we also diverge by modeling the shocks $\omega_{i,t}^*$ as time-varying quantities obeying a mixture distribution. In particular, prior to and after the shock, the $\omega_{i,t}^*$ are uniformly zero. During the shock, i.e. from $T_i^* + 1$ through $T_i^* + L_i^{vol}$, the $\omega_{i,t}^*$ are distributed i.i.d, with a core signal component that depends solely on the unknown parameter δ and the observable vector $\mathbf{x}_{i,T_i^*+1}$. Wherever necessary, to distinguish the two distributions that make up the shock distribution, we use the term *nonzero shocks* to refer to the shocks that are not uniformly zero.

The nonzero shocks each include an idiosyncratic noise term that varies across time and donors. This feature allows the nonzero shocks to exhibit some variation as the shock progresses, even within a single donor. The noise term could be modeled with additional complexity, including serial correlation, or could exhibit deterministic or stochastic decay throughout the shock period, proportional to a deterministic or stochastic function of time, or could peak somewhere in the middle of the shock period. Each of these ideas might correspond to some particular, realistic feature of the world.

The model $\mathcal{M}_1$ is defined by a volatility equation and mean equation, as is any GARCH model. The decision to model the volatility shock $\omega_{i,t}^*$ as an additive random effect is consistent with our setup. However, the decision to model the level effect $\epsilon_{i,t}^*$ as a temporary rupture in the otherwise i.i.d. sequence of innovations $\epsilon_{i,t}$ may not be intuitive. One way of arguing for this choice is that, in a discrete time series model, if we assume the arrival of news in the time between T_i^* and $T_i^* + 1$, we do not have an easy way to express a conditional distribution of the innovation ϵ_{i,T_i^*+1} given the arrival of information. Using $\epsilon_{i,t}^*$ thus breaks this impasse. This justification also explains why we do not parameterize the level shock at $T_i^* + 1$ as a sum of two shocks, ϵ_{i,T_i^*+1} and $\epsilon_{i,T_i^*+1}^*$, which would represent the level shock as generated by two independent sources of stochasticity. While we want to model the level shocks at T_i^*+1 as potentially large in absolute value, we also want to retain the property of a unitary source of noise.

6.3 HAR

The Heterogeneous Autoregressive (HAR) model for realized volatility [Corsi, 2009] uses ordinary least-squares (OLS), WLS, or some other appropriate linear model to fit a predictive regression model. It

explores the 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 a governed by a diffusion process (a model beyond our scope herein), and HAR estimates

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

Recently, Bodilsen and Lunde [2023] find that overnight news improves the predictive ability of HAR regressions with a logarithmic transformation on the realized volatility estimate. The authors distinguish between earnings-related and non-earnings-related news occurring after hours, finding that predictive performance is improved for one-day-ahead forecasts but no more. Consider now how HAR fits into Specification (1):

$$\mathcal{M}_1: \begin{aligned} &RV_{i,t} = \beta_0 + \beta_{1\text{-day}}RV_{i,t-1} + \beta_{5\text{-day}}\overline{\{RV_{i,t-1}, ..., RV_{i,t-5}\}} + \beta_{22\text{-day}}\overline{\{RV_{i,t-1}, ..., RV_{i,t-22}\}} + \alpha_{i,t} + \epsilon_t \\ &\alpha_{i,t} = \mathbf{x}_{i,t}^T \lambda_{i,t} \\ &\mathbf{x}_{i,t}^T = (1, x_{i,t}^1, ..., x_{i,t}^p) \\ &\lambda_{i,t}^T = (u_{i,t}, \lambda_t^1, ..., \lambda_t^p), \end{aligned}$$

$$(7)$$

We should note that the regressions under discussion here are predictive regressions, with no suggestion that the model is fundamentally driven by the linear specification that is estimated. However, if it were, we would especially desire answers to the following

1. Are the relevance and novelty filters used by Bodilsen and Lunde [2023] appropriate, and does it represent a qualitative decision? Does it represent an instance of similarity-based matching?

2. Could we relax the requirement on the relevance score and instead let a numerical algorithm determine which stories and relevant and how relevant?

3. Could we relax the requirement on the novelty score?

4. Bias-variance tradeoff: are we able to reduce the bias or the variance or both via the method of Bodilsen and Lunde [2023]?

5. What is the 'frontier' that separates using a similarity-based approach like SPC as opposed to an approach using count variables?

6. What are advantages and disadvantages of using counts versus indicator variables versus alternatives? Can we confirm the theory that count variable approaches reduce variance while, as is well-established, SPC reduces bias?

7. Does SPC do better when the shock is a nonlinear function of the covariates compared to a method that uses regression terms for each covariate? Does SPC do better under those circumstances compared to count variable-based approaches?

8. What DGP on overnight firm-specific news would most justify the approach in Bodilsen and Lunde [2023]?

9. Is earning-related and non-earnings-related the appropriate disaggregation to use?

10. Is it appropriate to use counts at each time point, or would indicator variables be superior, or rather, under what conditions would indicator variables be superior?

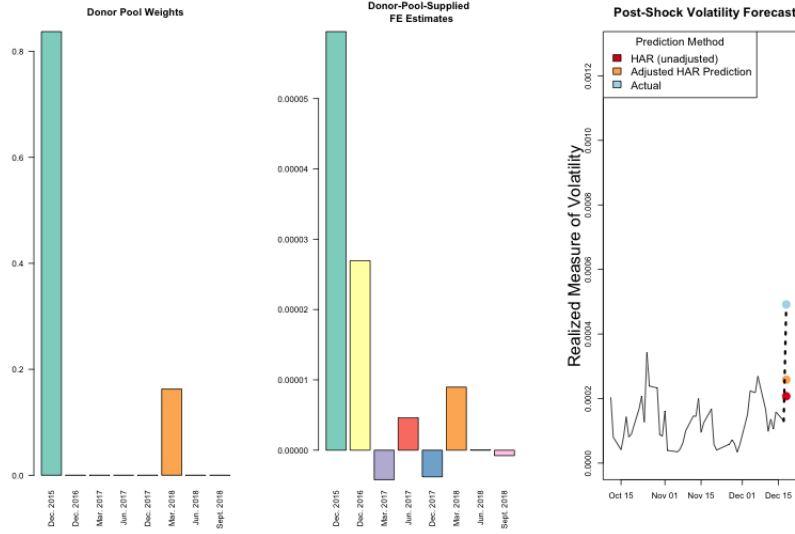### 6.3.1 A Real Data Application of Shock-Adjusted HAR Forecasting

Figure 4: 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

6.4 "Exponential Breaks"

In Castle et al. [2011], the authors discuss intercept shifts that decay deterministically. Those shifts are parameterized by a scalar η and a decay rate ψ. We adapt their model for our own purposes:

$$y_{i,t} = \alpha_i + \eta[1 - e^{-\psi[t-T_i^*]}]\mathbf{1}_{t \geq T_i^*+1} + \epsilon_{i,t}, \epsilon_{i,t} \overset{iid}{\sim} [0, \sigma^2] \ . \tag{8}$$

Such a specification captures the dynamics of a shock that is realized over an infinite horizon but which can be practically approached as a permanent shock. This poses a novel case for our method, in that the model is parametric and yet without estimability of all shock parameters, aggregation may not be straightforward or easy. Consider, for any i, $2 \leq i \leq n+1$, the estimates $\hat{\alpha}_{i,OLS}$ obtained via regression on the first $T_i^* + 1$ points as well as $\hat{\eta}_{OLS}$ obtained by estimating a fixed effect at $t = T_i$. Consider also a subsequent estimator, an estimator for ψ, that is derivable from inverting the model specification and plugging in the estimates $\hat{\alpha}_{i,OLS}, \hat{\eta}_{OLS}$:

$$\hat{\psi}_{i,plugin} = \frac{1}{T_i - T_i^*} \sum_{t \geq T_i^*+1}^{T_i} -log\big[\big(1 - \frac{y_{i,t} - \hat{\alpha}_i}{\hat{\eta}}\big)\big]/[t - T_i^*] \ . \tag{9}$$

(9) is, to put it lightly, a complicated estimator with potentially dubious properties. To begin, (9) is defined only when the term inside the logarithm is positive for each $t \geq T_i^* + 2$. We could get around this by dropping any summands for which that term is nonpositive and instead calculate an average over only the remaining subset of summands. Additionally, variance estimates for $\hat{\psi}_{i,plugin}$ may require resampling

methods or the delta method. Estimation error from $\hat{\alpha}$ and $\hat{\eta}$ would naturally propagate to $\hat{\psi}_{i,plugin}$, and the expected estimation error would depend upon the size of the pre-shock sample.

Castle et al. [2011] report that MSFE is lower by using intercept correction instead of non-linear methods to estimate ψ.

Now consider a practitioner in need of a conditional forecast for $T_1^* + h, h \geq 1$. Could aggregating the prediction residuals $r_{i,T_i^*+h} = y_{T_i^*+h} - \hat{y}_{i,T_i^*+h}, h \geq 1$ lead us to a superior predictions in the time series under study?

Consider also these complicating scenarios

1. We can estimate λ using a cross-sectional regression, and that may be better estimator than doing it one-by-one in each donor.

2. Here we have taken λ to be fixed and shared across donors. If λ is a random, however, then estimating it becomes harder.
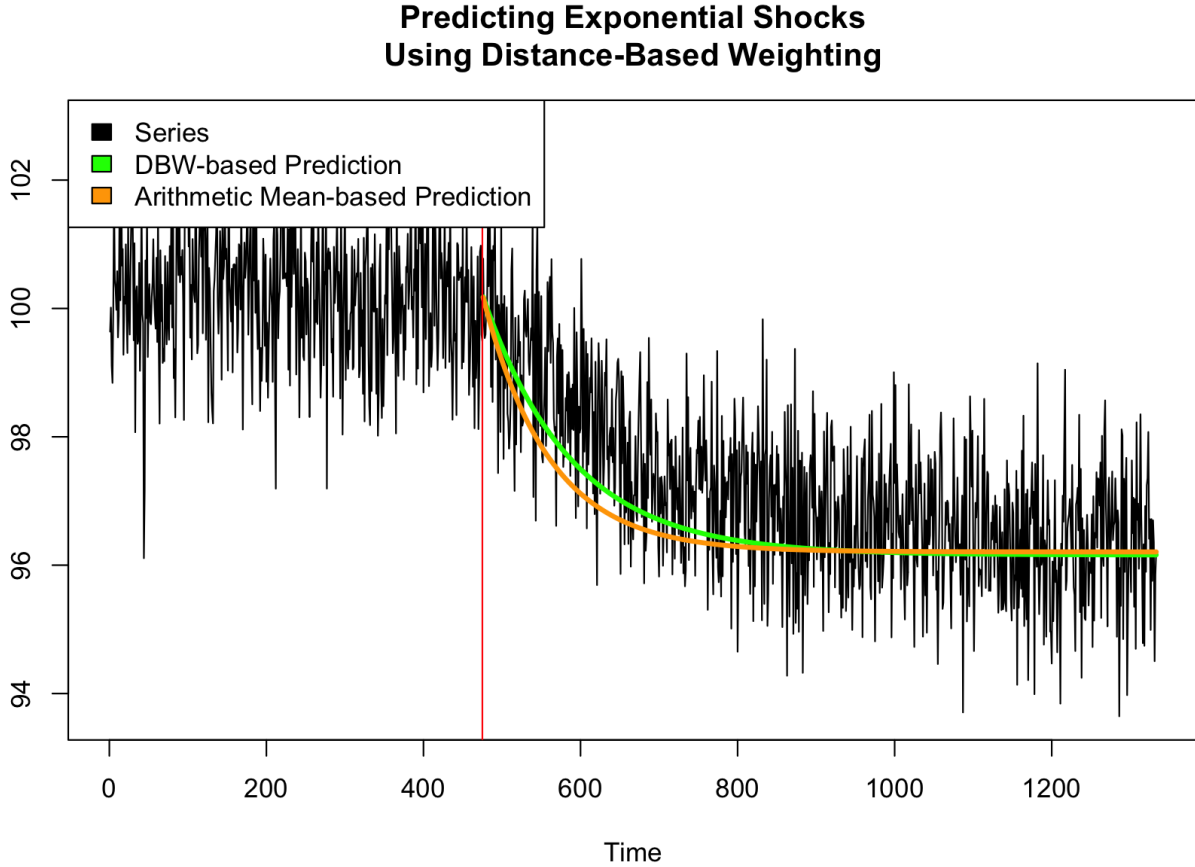
6.4.1 Simulation Study

**Predicting Exponential Shocks
Using Distance-Based Weighting**

Figure 5: The simulated time series experiences a shock...

What is the Model Confidence Set (MCS)?

7   Discussion

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   Asking questions about the covariate vector and conditional shock distribution

We pose two questions.

What if a given entry of the covariate vector was helpful at only one of those roles? For example, what if an entry was particularly useful in yielding good weights but was not important for the shock distribution — perhaps because the corresponding shock parameter was zero.

Conversely, suppose a given entry was known to be part of the shock distribution, but the inclusion of that covariate in distance-based weighting was often not worth it, perhaps because it was hard to approximate with using donors, and the corresponding shock parameter was small in magnitude with high probability.

7.2   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.

7.3   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

8   Extensions

- 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? Yet another potential motivation springs from the way we evaluate model performance. In Clements and Hendry [2005a], the authors discuss six dichotomies regarding forecast evaluation as a way of advising caution about simplistic notions of performance evaluation. These ideas matter for the current manuscript for at least one reason. We must entertain the possibility that breaks in the DGP will invalidate the standard forecast evaluation framework under which we had been working. If, for example, we 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?

- Bias-variance decomposition for ARIMA, as introduced in (6.1)

- Shrinkage estimation of autoregressive structure on all donors, then use our method

- TAR

- SETAR

- Function forecasts

- Binary Outcome Forecasts

- Density Forecasts

- Quantile Forecasts

8.1 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 Lundquist and Eck [2024].

Consider the leave-one-out method used on pairs of donors and covariates in Lundquist and Eck [2024]. By averaging over $(n+1)\cdot(p+1)$ predictions, we average over $(n+1)\cdot(p+1)$ convex combinations (where, when a donor is ommitted, its weight can be imputed as zero).

8.2 Limitations

8.2.1 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?

8.2.2 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. 2.1

- distinguish the method from various tools that inspired it (1)

- show a few special cases, both examples and formal results (6)

- discuss limitations of the method (8.2)

- propose extensions (8)

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 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

9 Supplement

9.1 Formal Results

Proof of Proposition 1 We claim

The result for each $\hat{\omega}_{i,*}^*$ follows from the consistency proof of the QMLE in GARCH-X models, as established by Han and Kristensen [2014]. As for a convex combination of $\{\hat{\omega}_{i,*}^*\}_{i=2}^{n+1}$, consider the following argument, where we have suppressed the time-indexation:

$$\mathbf{x}_1 = \sum_{i=2}^{n+1} \pi_i \mathbf{x}_i \qquad \text{(by an assumption of the proposition)}$$

$$\implies \delta^T \mathbf{x}_1 = \delta^T \sum_{i=2}^{n+1} \pi_i \mathbf{x}_i = \sum_{i=2}^{n+1} \pi_i \delta^T \mathbf{x}_i \qquad \text{(by left-multiplication by } \delta^T)$$

$$\implies \delta^T \mathbf{x}_1 = \sum_{i=2}^{n+1} \pi_i (\omega_i^* - \mu_{\omega^*} - u_{i,t}) \qquad \text{(by substitution and definition of } \omega_i^*)$$

$$\implies \mu_{\omega^*} + \delta^T \mathbf{x}_1 + \sum_{i=2}^{n+1} \pi_i u_{i,t} = \sum_{i=2}^{n+1} \pi_i \omega_i^* \qquad \text{(by moving terms to the left-hand side)}$$

Now, since each $\hat{\omega}_i^*$ converges in probability to ω_i^*, by Slutsky, any finite linear combination of $\hat{\omega}_i^*$ converges in probability to the corresponding finite linear combination of ω_i^*. Hence, $\sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ converges in probability to the right-hand side of the final line. By setting $u_{i,t}$ to zero for all shock points and all i, we then have $\omega_1^* = \mu_{\omega^*} + \delta^T \mathbf{x}_1$ on the left-hand side. Without taking the $u_{i,t}$ to be zero at the shock points, however, we nevertheless have convergence in distribution. The proof is as follows. For any collection of $n+1$ i.i.d variables, any convex combination of any n of them is equal in distribution to the one left out. Therefore, the left-hand side of the final line is equal in distribution to ω_1, as desired.

$\square$

Proof of Proposition 2 We claim

Without loss of generality, we prove the case for $r = 1$. The general case follows from iterative one-step-ahead forecasts. Recall the conditional expectation of the variance for the GARCH-X(m,s) model:

$$\mathbb{E}[\sigma^2_{adjusted,t+1}|\mathcal{F}_t] = \omega_i + \sum_{k=1}^{m_i} \alpha_{1,k} a^2_{1,t+1-k} + \sum_{j=1}^{s_1} \beta_{1,j} \sigma^2_{1,t+1-j} + \gamma_1^T \mathbf{x}_{1,t+1} + \omega^*_{1,*} \ .$$

By replacing parameters with their estimates, we arrive at the prediction

$$\hat{\sigma}^2_{adjusted,t+1}|\mathcal{F}_t = \hat{\omega}_1 + \sum_{k=1}^{m_1} \hat{\alpha}_{i,k} a^2_{i,t+1-k} + \sum_{j=1}^{s_1} \hat{\beta}_{i,j} \hat{\sigma}^2_{1,t+1-j} + \hat{\gamma}_1^T \mathbf{x}_{1,t+1} + \hat{\omega}^*_{1,*} \ ,$$

which converges to

$$\sigma^2_{adjusted,t+1}|\mathcal{F}_t = \omega_i + \sum_{k=1}^{m_i} \alpha_{1,k} a^2_{i,t+1-k} + \sum_{j=1}^{s_1} \beta_{i,j} \sigma^2_{1,t+1-j} + \gamma_1^T \mathbf{x}_{1,t+1} + \omega^*_{1,*}$$

in distribution or probability (depending upon how $\hat{\omega}^*_*$ converges) as $t \to \infty$ in the donor pool by a simple application of Slutsky's Theorem.

$\square$

References

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.

Paul Bartha. Analogy and analogical reasoning. 2013.

Christopher M Bishop and Hugh Bishop. *Deep learning: Foundations and concepts*. Springer Nature, 2023.

Simon Tranberg Bodilsen and Asger Lunde. Exploiting news analytics for volatility forecasting. *Available at SSRN 4401032*, 2023.

Jennifer L Castle, Nicholas WP Fawcett, and David F Hendry. Forecasting breaks and forecasting during breaks. 2011.

Jennifer L Castle, Michael P Clements, and David F Hendry. Forecasting by factors, by variables, by both or neither? *Journal of Econometrics*, 177(2):305–319, 2013.

Jennifer L Castle, Michael P Clements, and David F Hendry. An overview of forecasting facing breaks. *Journal of Business Cycle Research*, 12(1):3–23, 2016.

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. *Forecasting non-stationary economic time series*. mit Press, 1999.

Michael P Clements and David F Hendry. Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, 67:931–956, 2005a.

Michael P Clements and David F Hendry. Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics & Statistics*, 67(S1):713–753, 2005b.

Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.

Dean D Croushore. Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia*, 6:3, 1993.

Claudia Foroni, Massimiliano Marcellino, and Dalibor Stevanovic. Forecasting the covid-19 recession and recovery: Lessons from the financial crisis. *International Journal of Forecasting*, 38(2):596–612, 2022.

Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

Ankita Garg, Catherine G Enright, and Michael G Madden. On asymmetric similarity search. In *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*, pages 649–654. IEEE, 2015.

Pablo Guerrón-Quintana and Molin Zhong. Macroeconomic forecasting in times of crises. 2017.

Heejoon Han and Dennis Kristensen. Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3):416–429, 2014.

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.

Mike Wilson. What's boosting cyclical stocks?, 2024. URL https://www.morganstanley.com/ideas/thoughts-on-the-market-wilson. Published: 2024-10-21.