Volatility Forecasting Using Similarity-based Parameter Correction and Aggregated Shock Information

Abstract

We develop a procedure for forecasting the volatility of a time series immediately following a news shock. Adapting the similarity-based framework of Lin and Eck (2021), we exploit series that have experienced similar shocks. We aggregate their shock-induced excess volatilities by positing the shocks to be affine functions of exogenous covariates. The volatility shocks are modeled as random effects and estimated as fixed effects. The estimates are then aggregated into a scalar quantity and use to adjust the GARCH forecast of the time series under study by an additive term. The adjusted and unadjusted forecasts are evaluated using the unobservable but easily-estimated realized volatility (RV). Numerical simulations are provided to illustrate conditions and hyperparameters under which our method thrives. A detailed real-world application for forecasting volatility after the outcome of the 2016 United States presidential election demonstrates the applicability of the method.
Keywords: GARCH, Forecasting, Volatility, Finance, Risk, Pooling, Aggregation, Similarity

1. Introduction

Reacting to a seemingly unprecedented event might involve the question: what, if anything, does it resemble from the past? Such might be the case with event-driven investing strategies, where the identification of the event could arise via the news pages or corporate communications and hence contains a qualitative, narrative element (Kenton, 2022). Matching a current crisis to past events is a problem with unsurprising statistical angles: identification, sample size, weighting, risk, and robustness, among many others.

In the context of foreign exchange rate market structure, Dominguez and Panthaki (2006) speculate "[w]hether news is scheduled or non-scheduled its influence on exchange rates may be related to the state of the market at the time of the news arrival. News that arrives during periods of high uncertainty may have different effects on the exchange rate, than news that arrives in calmer periods." The authors also note that non-scheduled news may require more time for markets to digest, leading to greater heterogeneity (including but not limited to higher dispersion) of responses. We take inspiration from these arguments, developing a method suitable to the conditions that typically accompany news shocks. Central to our endeavor is the observation that news shocks have both a qualitative (the description of the event as well as future contingents surrounding it) and quantitative component.

In this work we focus on the second central moment of a time series, the volatility. One of the most important aspects of a positively-valued time series $(P_t)_{t\in\mathbb{N}}$, especially financial time series, is the volatility of the return series $(r_t)_{t\in\mathbb{N}}$. A financial asset's price series may exhibit behavior that makes inapplicable and uninterpretable the traditional methods of time series analysis. In contrast, the return series is scale-free (Tsay, 2005), easily-interpreted, and often at least weakly stationary. Our reasons for studying volatility are two-fold. Whereas once returns were thought to react to specific news events, now stock price movements are believed to be overwhelmingly noise-based (see Boudoukh, Feldman, Kogan, and Richardson (2019) and references therein), and even if one could construct credible models for describing and forecasting price series and return series, that would not necessarily tell us much about the variability of such forecasts nor enlighten us about the evolution of the variability of $(P_t)_{t\in\mathbb{N}}$ and $(r_t)_{t\in\mathbb{N}}$ or under rapidly changing conditions.

No matter how a time series or its transformations are modeled, forecasting in the presence of news shocks requires a methodological framework that sensibly incorporates relevant information that has yet to manifest in market price or derivative quantities like volatility. In this setting, regime-change models (see Bauwens, Preminger, and Rombouts (2006) and citations therein) are of little use because under the assumption of a known exogenous shock, there is no need to estimate a regime-change time, nor is there data following the exogenous shock event to fit a model. Asymmetric GARCH models were an early attempt to account for fact that negative returns typically beget greater volatility than positive returns (Hansen, Huang, and Shek, 2012). Asymmetry and mixed-frequency are employed in Wang, Ma, and Liu (2020) to forecast under the presence of extreme shocks. Problematically, any asymmetric model will depend upon the observation of a negative return to provide the most updated volatility forecast, but under the circumstances posited herein, no such return has been observed.

Similar problems and more exist for Realized GARCH models (Hansen et al., 2012) in our setting. These models incorporates observable measures of volatility, known as "realized measures", like implied volatility (IV). Under the assumptions herein, no post-shock data is available, and even if it were, Realized GARCH does not promise to perform well, since Black-Scholes implied volatility is a biased estimator of volatility (Mayhew, 1995; Christensen and Prabhala, 1998), with the bias increasing in times of crises, when options may be far out-of-the-money. GARCH models have been shown slow to adapt to spikes in volatility (Andersen, Bollerslev, Diebold, and Labys, 2003).

The approach herein sidesteps the modeling shortcomings of Realized GARCH and other approaches which require observations of post-shock outcomes by substituting modeling assumptions. The method herein proceeds under the assumption that similar news events occasion volatility shocks arising from a common conditional shock distribution. The procedure proposed does not require post-shock information like returns or market-implied quantities from the time series under study. Hence, we also avoid questions about what realized measure to use and when as well as questions about the usefulness of high-frequency data, although these remain intriguing avenues for future work.

The primary methodological tool presented in this work is fixed effect estimation followed by a distance-based weighting procedure for pooling those estimates. The use of fixed effect estimation for the study of structural shocks has a pedigree in macroeconomic analysis (Romer and Romer (1989) as referenced in Kilian and Lütkepohl (2017); see also discussion of determinstic exogenous events in Engle and Patton (2001)). We employ fixed effect estimation on the basis of a well-established conceptual assumption that shocks of economic time series can be modeled as mixtures, in particular, mixtures of ordinary innovations and shocks due to rare events (see Phillips (1996) and references therein). In the forecasting literature, the term "intercept correction" has come to refer to a modeling technique in which nonzero errors are explicitly permitted (Hendry and Clements, 1994; Clements and Hendry, 1998). They summarize the literature as distinguishing two families of intercept correction: so-called "discretionary" intercept corrections that attempt to account for future events, without hard-coding an ad-hoc adjustment into the model specification, and second, "automated" intercept corrections that attempt to adjust for persistent misspecification using past errors. Guerrón-Quintana and Zhong (2017) use weighted subsets of a scalar series' own past to correct forecast errors by an additive term. Dendramis, Kapetanios, and Marcellino (2020) a introduces a similarity-based forecasting procedure for time-varying coefficients of a linear model. Foroni, Marcellino, and Stevanovic (2022) employ a form of intercept correction in order to adjust forecasts to the COVID-19 shock in the spring of 2020 based on the proportional misses of the same model applied to the Great Recession.

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

Our procedure is a discretionary procedure for intercept correction that incorporates systematically data internal or external to the time series under study. The correction function, as we shall see, includes a convex combination of fixed effect estimates from a donor pool. In the causal inference literature, a donor pool is a set of units of observation that share distributional properties with the prime object of analysis. In Abadie, Diamond, and Hainmueller (2010), the authors build upon previous work in causal inference whereby a treatment effect can be estimated via comparison with a synthetic time series that represents the control. In their setting, the synthetic unit is constructed using a convex combination of the donors. The particular convex combination employed is a function of the distance between the time series under study and the donors. Donors that are closer to the time series under study are given greater weight. Conversely, donors that are unlike the time series under study receive little to no weight. Lin and Eck (2021) adapt the notion of a donor pool as well as distance-based weighting for the purpose of prediction. Their one-step-ahead forecasts use distance-based-weighting to aggregate shock estimates from the donor pool's time series according to the similarity to the time series under study. Their approach does not take into account the ARCH effects commonly observed in time series, especially financial times series, leaving unaccounted for the variability in the variability of time series forecasts. Outside of Lin and Eck (2021), we know of no prior work that both introduces a parametric specification for nonzero errors and introduces a procedure for weighting appropriately the nonzero errors of similar shocks occurring outside the time series under study. Likewise, we are not familiar with any prior work that attempts to account for anticipated nonzero errors using an explicit parametric adjustment, i.e., what we will call a "correction function".

The method proposed herein is supported by both numerical evidence as well as a real data example. The numerical evidence takes the form of simulations and analysis of the method's performance across a considerable grid of varying parameters, allowing us to evaluate reasonable hypotheses about the method but also uncover some intriguing nuances. The real data example, on the other hand, demonstrates the usefulness of the method by applying it to the United States (US) presidential election. We find that the method would have allowed a researcher to predict the volatility of a financial asset that immediately followed the 2016 US presidential. Our discussion of this example sheds light on some subtleties of the method as well as its robustness to various choices that must be made to construct and assemble donors and appropriate covariates (Steegen, Tuerlinckx, Gelman, and Vanpaemel, 2016).

2. Setting

In order to motivate our procedure, we provide a visual illustration. In Figure 1, we show how the aggregation of estimated shock-induced excess volatilities from donors in the donor pool works when the correction function is a specially-chosen convex combination of fixed effects from the donor pool. Our method assumes a credible, parsimonious parameterization in which the shock is an affine transformation of several key covariates. The key intuition behind this shock parameterization is that as the strength of the linear signal increases relative to the idiosyncratic error, the GARCH estimation of these effects increases in accuracy. From this, it follows that the aggregated shock estimate increases in accuracy.

2.1. A Primer on GARCH

We define the log return of an asset between $t-1$ and t as $r_t = \log(\frac{P_t}{P_{t-1}})$, where P_t denotes the price at time t. The class of ARIMA(p, d, q) models (Box, 2013) provides a framework for modeling the autoregressive structure of r_t. These models assume a certain dependence structure between r_t and $(r_k)_{k\leq t}$, yet their errors — often called innovations in the financial time series context due to how they represent the impact of new information — are nevertheless assumed to be i.i.d. with mean zero and constant variance. The ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) models provide

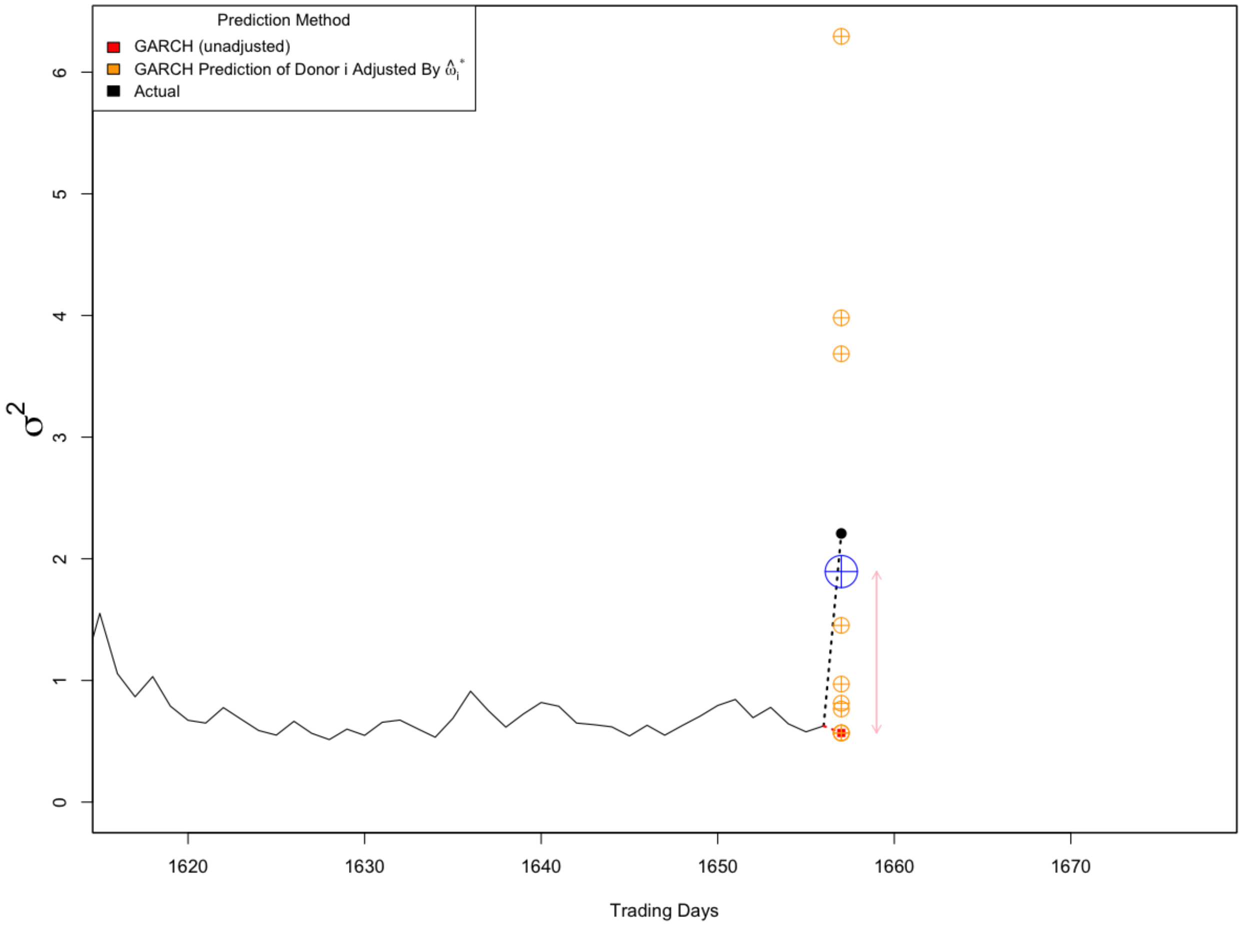


Figure 1: The simulated time series experiences a volatility shock between trading days 1,656 and 1,657. The GARCH prediction, in red, fails even to approach the volatility spike at $T^* + 1$, as do several adjusted predictions, which are orange. In contrast, the GARCH forecast adjusted by $\hat{\omega}^*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

elegant alternatives to the homoskedasticity assumption. In fact, the GARCH framework in its most basic form disregards r_t and instead turns its interest to the series r_t^2 (once properly centered, i.e. after assuming a mean-model for returns).

To that end, let $a_t = r_t - \mu_t$, where μ_t is the mean of the log return series r_t. We thus derive a mean-zero process $(a_t)_{t \in \mathbb{N}}$ with the property that $\mathbb{E}[a_t^2] = \text{Var}[a_t]$. Under the assumption of time-invariant volatility, the series a_t^2 should exhibit no autocorrelation at any lag $\ell \geq 1$. This assumption motivates tests for ARCH effects, that is, tests for the clustering of volatility. These tests explore the alternative hypothesis that σ_t^2 is not only a time-varying parameter but furthermore a function of past squared residuals of the mean model. In particular, the ARCH(m) model is an autoregressive model in which σ_t^2 is a deterministic function of the past m values of r_t^2. The GARCH(m, s) framework take this one step further by modeling σ_t^2 as a linear combination of the past m values of r_t^2 and well as the past s values of σ_t^2. In functional form, a GARCH process (sometimes called a strong GARCH process (Francq and Zakoian, 2019, p. 19)) is given by

$$\sigma_t^2 = \omega + \sum_{k=1}^{m} \alpha_k a_{t-k}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$
$$a_t = \sigma_t \epsilon_t$$
$$\epsilon_t \overset{iid}{\sim} E[\epsilon_t] = 0, Var[\epsilon_t] = 1$$
$$\forall k, j, \alpha_k, \beta_j \geq 0$$
$$\forall t, \omega, \sigma_t > 0 \ .$$

Assuming further that σ_t^2 depends on a vector of exogenous covariates $\mathbf{v}_t$, we have a GARCH-X(m, s). The volatility equation then becomes

$$\sigma_t^2 = \omega + \sum_{k=1}^{m} \alpha_k a_{t-k}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 + \gamma^T \mathbf{v}_t \ . \tag{1}$$

2.2. Model setup

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t} = (r_{i,t},\ \mathbf{v}_{i,t})$, $t = 1, \ldots, T_i$, $i = 1, \ldots, n+1$, where $r_{i,t}$ is scalar and $\mathbf{v}_{i,t}$ is a vector of covariates such that $\mathbf{v}_{i,t}|\mathcal{F}_{i,t-1}$ is deterministic. Suppose that the analyst is interested in forecasting the volatility of $r_{1,t}$, the first time series in the collection, which we will denote *the time series under study*. We require that each time series $\mathbf{y}_{i,t}$ is subject to an observed news event following $T_i^* \leq T_i + 1$ and before witnessing $T_i^* + 1$. We are implicitly leveraging the fact that financial assets are heavily traded during market hours, yet only thinly traded (if traded at all) outside market hours. In contrast, the arrival of market-moving news does not obey any such restrictions. In light of the foregoing, we can denote our collection of GARCH-X volatility equations of interest using the following notation

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

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 $\delta, \mathbf{v}_{i,t} \in \mathbb{R}^p$. 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 $D^{vol}_{i,t} = I(t \in \{T^*_i + 1, ..., T^*_i + L_{i,vol}\})$, 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

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

with time-invariant error structure

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

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 $\mathbf{v}_{i,t}$ will be observed before $\sigma^2_{i,t}$ 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 $T^*_i + L^{vol}_i$, 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 $\mathbf{v}_{i,T^*_i+1}$. 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 a 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 a the noise term that increases or decays throughout the shock, proportional to a deterministic function of time, or 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 ϵ^*_{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.

3. Methodology for Similarity-based Parameter Correction

We now introduce how a researcher can carry out predictions via distance-based weighting. The first step is to gather the p covariates that will parametrize a change in dynamics to model $\mathcal{M}_1$ and will represent the p-dimensional space in which weights are generated. This set of covariates parameterizes the shock to $\sigma^2_{i,t}$ of magnitude $\omega^*_{i,t}$ in model $\mathcal{M}_1$ above. Let $\mathbf{V}_t \in \mathbb{R}^{p\times n}$ be formed with columns $\mathbf{v}_{i,t}$, $i = 2, \ldots, n+1$. Without loss or generality, let the indices of each $\mathbf{v}_{i,t}$ be shifted and arranged such that $\mathbf{V}_{T^*+1}$ has as columns $\mathbf{v}_{2,T_2^*+1}, ..., \mathbf{v}_{n+1,T_{n+1}^*+1}$. Formally, in vector notation, the donor shocks are given by $\vec{\omega}_t = \vec{\mu}_{\omega^*} + \delta^T\mathbf{V}_t + \vec{u}_t$, where were have suppressed the $D^{vol}_{i,t}$ notation for simplicity. Ideally, $\mathbf{V}_t$ will display 'balance' in that p covariates exist for each of the n donors. In practice, missing values, corrupted values, or unacceptably extreme or noisy estimates may necessitate some sort of matrix completion, a problem that we do not tackle in this work.

3.1. Forecasting

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

$$\begin{aligned}
\hat{\sigma}^2_{unadjusted,T_1^*+1} = \hat{\mathbb{E}}[\sigma^2_{1,T_1^*+1}|\mathcal{F}_{T_1^*}] \quad &= \quad \hat{\omega}_1 + \sum_{k=1}^{m_1}\hat{\alpha}_{1,k}a^2_{1,T_1^*+1-k} + \sum_{j=1}^{s_1}\hat{\beta}_{1,j}\sigma^2_{1,T_1^*+1-j} + \hat{\gamma}_1^T\mathbf{v}_{1,T_1^*+1}, \\
\hat{\sigma}^2_{adjusted,T_1^*+1} = \hat{\mathbb{E}}[\sigma^2_{1,T_1^*+1}|\mathcal{F}_{T_1^*}] + \hat{\omega}^* \quad &= \quad \hat{\omega}_1 + \sum_{k=1}^{m_1}\hat{\alpha}_{1,k}a^2_{1,T_1^*+1-k} + \sum_{j=1}^{s_1}\hat{\beta}_{1,j}\sigma^2_{1,T_1^*+1-j} + \hat{\gamma}_1^T\mathbf{v}_{1,T_1^*+1} + \hat{\omega}^* \ .
\end{aligned}$$

Note that GARCH models can be parameterized as ARMA models on the squares of the scalar time series a_t^2 (Tsay, 2005; Francq and Zakoian, 2019, p. 18, p. 46), assuming that a_t satisfies fourth-order stationarity. This fact matters for forecasting because the h-step-ahead forecasting function for a GARCH model is, just like for an ARMA model, the conditional expectation function, $\mathbb{E}[\sigma^2_{i,T_i^*+h}|\mathcal{F}_{T_i^*}]$, or practically speaking, the estimate thereof, $\hat{\mathbb{E}}[\sigma^2_{i,T_i^*+h}|\mathcal{F}_{T_i^*}]$ (Zivot, 2009). Here we have presented one-step-ahead forecasts for a GARCH-X(m, s), without loss of generality. For $h = 2, 3, 4, ...$, the conditional expectation is computed recursively, as is standard for iterative autoregressive forecasts. Note that such h-step ahead predictions will additionally require $\mathbf{v}_{i,t}$, $t \geq T_i^* + 2$ to be known, or at least well-estimated in practice. Thus we will focus on the one-step-ahead forecast unless stated otherwise.

3.2. Excess Volatility Estimators

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

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

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

which will be taken to be our forecast adjustment term. Since the weights $\{\pi_i\}_{i=2}^{n+1}$ are computed using $\mathcal{F}_{T_i^*}$, the set $\{\pi_i\}_{i=2}^{n+1}$ is deterministic and observed, *modulo* any stochastic ingredient in the numerical methods employed to approximate $\mathbf{v}_{1,T_1^*+1}$ using a convex combination of donor covariates. We say more about the properties of the shocks $\omega_{i,T_i^*+1}^*, ..., \omega_{i,T_i^*+L_i^{vol}}^*$ in section 4.

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

$$\{\pi\}_{i=2}^{n+1} = \arg\min_{\pi} \|\mathbf{v}_{1,T_1^*+1} - \mathbf{V}_{T^*+1}\pi\|_{\mathbf{S}}.$$

In the forecast combination literature, it is of interest whether the weights employed to aggregate forecasts strive toward and meet various optimality criteria (Timmermann, 2006; Wang, Hyndman, Li, and Kang, 2023). In our endeavor, there are at least two senses of optimal weights that one might be interested in. First, we can think of optimal weights as a set $\{\pi_i\}_{i=2}^{n+1}$ such that $\omega_1 = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_{i,*}$, i.e., ω_1 is recovered perfectly, as it belongs to convex hull of the estimated shocks. However, ω_1 is never revealed to the practitioner, and hence there is no way of verifying the extent to which this condition is satisfied.

A more promising aim is finding weights such that $\mathbf{v}_1 = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T_1^*+1}$, meaning that the covariates of the time series under study lies within the convex hull of the donor covariates. This condition underwrites asymptotic results in Abadie et al. (2010), and the intuition there extends to this work: if the shock is parameterized by an affine function of covariates, then finding a linear combination that recreates the shock should serve us well. Because the method proposed uses a point in Δ^n, it is important to head-off possible confusion. What we are proposing is not a forecast combination method. What we are aggregating and weighting (not combining) are subcomponents of forecasts, not forecasts themselves. Moreover, from a broader perspective, forecast combination is an inapt term for what is being proposed here. First, the donor time series do not provide forecasts, nor would forecasts be needed for random variables that have already been realized. Second and more fundamentally, the theoretical underpinnings of forecast combination, while diverse (Wang et al., 2023), are distinct from the setting presumed in this work.

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

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

Comparison with least-squares estimation is also illustrative. Consider least-squares for the p-vector δ

$$\hat{\delta} = \arg\min_{\delta} \|\hat{\omega}^* - \delta^T \mathbf{V}_{T^*+1}\|_2,$$

where $\mathbf{V}_{T^*+1}$ could include an adjoined row $\mathbf{1}_n$, in order to estimate the locator parameter of the shock. One problem is that this optimization problem is an optimization problem over p-vectors δ — over linear combinations of the covariates, whereas what we seek is an n-vector — a linear combination of donors. Additionally, there is no guarantee that δ would perform poorly as a tool for producing $\hat{\omega}^*$, but given the small number of donors supposed in our setting, it is risky.

3.3. Ground Truth Estimators

The time-varying parameter σ_t^2 is a quantity for which even identifying an observable effect in the real world is far more challenging. In this work, we use a common estimator of the variance called realized volatility (RV), one which has the virtue of being "model-free" in the sense that it requires no modeling assumptions (Andersen and Benzoni, 2010). The realized variance itself can be decomposed into the sum of a continous component and a jump component, with the latter being less predictable and less persistent (Andersen, Bollerslev, and Diebold, 2007), cited in De Luca et al. (2006), two factors that further motivate the method employed herein.

Suppose we examine K units of of time, where each unit is divided into m intervals of length $\frac{1}{m}$. We adapt the notation of Andersen and Benzoni (2008). Let $p_t = \log P_t$, and let $\tilde{r}(t, \frac{1}{m}) = p_t - p_{t-\frac{1}{m}}$. Suppressing the i index, we estimate the variance the log return series using Realized Volatility of the K consecutive trading days that conclude with day t, denoted $RV_t^{K,m}$, using

$$RV_t^{K,m} = \frac{1}{K}\sum_{v=1}^{Km} \tilde{r}^2(v/m, 1/m),$$

where the K trading days have been chopped into Km equally-sized blocks.

Assuming that the K units $\tilde{r}(t,1) = p_t - p_{t-1}$ are such that $\tilde{r}(t,1) \overset{iid}{\sim} N(\mu, \delta^2)$, it is easily verified that $\mathbb{E}[RV^{K,m}] = \frac{\mu^2}{m} + \delta^2$, which is a biased but consistent estimator of the variance. We will proceed using $m = 77$, corresponding to the 6.5-hour trading day chopped into 5-minute blocks, with the first block omitted in order to ignore unusual trading behavior at the start of the day. Other sensible choices for m can replace our specification if desired.

3.4. Loss Functions

We are interested in point forecasts for $\sigma^2_{1,T_i^*+h}|\mathcal{F}_{T_i^*}$, $h = 1, 2, ...$, the h-step ahead conditional variance for the time series under study. Let L^h with the subscripted pair {prediction method, ground truth estimator}, denote the loss function for an h-step-ahead forecast using a given prediction function and ground truth estimator. For example, if we suppress the time index, the one-step-ahead MSE using our method and Realized Volatility as the ground truth is

$$\text{MSE}^1_{\text{aggregated estimaor, RV}} = (\hat{\sigma}^2_{aggregatedestimaor} - \hat{\sigma}^2_{RV})^2$$

Also of interest in absolute percentage error for an h-step-ahead forecast, defined as

$$\text{APE}^h_{method,groundtruth} = \frac{|\hat{\sigma}^2_{h,method} - \hat{\sigma}^2_{h,groundtruth}|}{\hat{\sigma}^2_{h,groundtruth}}$$

Finally, we introduce the QL (quasi-likelihood) Loss (Brownlees, Engle, and Kelly, 2011):

$$\text{QL}^h_{method,groundtruth} = \frac{\hat{\sigma}^2_{h,groundtruth}}{\hat{\sigma}^2_{h,method}} - \log\frac{\hat{\sigma}^2_{h,groundtruth}}{\hat{\sigma}^2_{h,method}} - 1\ .$$

What distinguishes QL Loss is that it is multiplicative rather than additive. This has benefits, both practical and theoretical. As Brownlees et al. (2011) explain, the technical properties of the QL Loss allow researchers to compare forecasts across heterogeneous time series, whereas additive loss functions like MSE unfairly penalize forecasts made under market turbulence. For this reason and others, we proceed to evaluate the method, both in simulations and real data examples, using the QL loss.

4. Properties of Volatility Shocks and Shock Estimators

In this section we will provide estimation properties for our shock estimator (2) and our adjusted forecast. These results will be with $m_i = s_i = 1$, for all $i = 1, \ldots, n+1$ in model $\mathcal{M}_1$ unless otherwise stated, i.e. a GARCH(1,1) model with additional parameterizations for shocks. Note that the dual level-volatility shock in $\mathcal{M}_1$ has a marginal effect on the conditional variance $\sigma^2_{i,t}$ that reflects the geometric decay of innovations in autoregressive models. As usual, assume $\alpha + \beta < 1$. Furthermore, assume that both the volatility shock and the level shock are of length one only, and consider a circumstance with no exogenous covariate $\mathbf{v}_{i,t}$. Assume also that $r \geq 2$, which is necessary in order to isolate the effects of the level shock $\epsilon^*_{i,t}$. Then

$$
\begin{aligned}
\sigma^2_{i,T_i^*+r+1} | \mathcal{F}_{T_i^*+r} &= \omega_i + \alpha_i a^2_{T_i^*+r} + \beta_i \sigma^2_{i,T_i^*+r} \\
&= \omega_i + \alpha_i (\sigma_{i,T_i^*+r} \epsilon_{T_i^*+r})^2 + \beta_i \sigma^2_{i,T_i^*+r} \\
&= \omega_i + \sigma^2_{i,T_i^*+r} (\alpha_i (\epsilon_{T_i^*+r})^2 + \beta_i) \ .
\end{aligned}
\tag{3}
$$

In Equation (3), observe that ω^*_{i,T_i^*+1} and ϵ^*_{i,T_i^*+1} each appear at most once, through the term $\sigma^2_{T_i^*+r}$. This might lead one to suspect geometric decay of the shocks ω^*_{i,T_i^*+1} and ϵ^*_{i,T_i^*+1}. Such a suspicion is easier to justify by examining the conditional expectation of the variance, $\mathbb{E}[\sigma^2_{i,T_i^*+r+1} | \mathcal{F}_{T_i^*+r}]$, which also happens to be the principal forecasting tool for a GARCH model (Zivot, 2009). Indeed, if we assume unit variance for all $\epsilon_{i,t}$ except, of course, $\epsilon^*_{i,t}$, then we have

$$
\begin{aligned}
\mathbb{E}[\sigma^2_{i,T_i^*+r+1} | \mathcal{F}_{T_i^*+r}] &= \mathbb{E}[\omega_i + \alpha a^2_{T_i^*+r} + \beta \sigma^2_{i,T_i^*+r} | \mathcal{F}_{T_i^*+r}] \\
&= \omega_i + \mathbb{E}[\alpha (\sigma_{i,T_i^*+r} \epsilon_{T_i^*+r})^2 | \mathcal{F}_{T_i^*+r}] + \beta \sigma^2_{i,T_i^*+r} \\
&= \omega_i + \alpha \sigma^2_{i,T_i^*+r} + \beta \sigma^2_{i,T_i^*+r} \qquad \text{(Due to the unit variance assumption)} \\
&= \omega_i + \sigma^2_{i,T_i^*+r} (\alpha + \beta) \ .
\end{aligned}
$$

By repeated substitution, in conditional expectation, the shock belongs to the equivalence class $\mathcal{O}((\alpha + \beta)^r)$. We generalize this observation in the following proposition.

Proposition 1. *Let a_t be a mean-zero time series obeying a GARCH(1,1) specification with unit-variance errors, all prior to the arrival of a volatility shock of length $L_{vol} \geq 1$ and level shock of length $L_{return} \geq 1$ at some time $T^* + 1$. Then for any r such that $r \geq max\{L_{i,vol}, L_{i,level}\} + 1$,*

$$
\mathbb{E}[\sigma^2_{i,T_i^*+r+1} | \mathcal{F}_{T_i^*+r}] = \omega_i + (\alpha + \beta) \sigma^2_{i,T_i^*+r}.
$$

In other words, for a GARCH(1,1), once two time points removed from the longest shock length, the volatility shock and level shock can be subsumed into one. However, prior to being two time points removed, there is no such guarantee. For example, one can take $r = 1$ and level shock of length at least 1 to see that

$$\begin{aligned}
\mathbb{E}[\sigma^2_{i,T_i^*+2}|\mathcal{F}_{T_i^*+1}] &= \mathbb{E}[\omega_i + \alpha a^2_{T_i^*+1} + \beta\sigma^2_{i,T_i^*+1}|\mathcal{F}_{T_i^*+1}] \\
&= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T_i^*+r}\epsilon^*_{T^*+1})^2|\mathcal{F}_{T_i^*+1}] + \beta\sigma^2_{i,T_i^*+1} \\
&= \omega_i + \alpha\sigma^2_{i,T_i^*+1}(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta\sigma^2_{i,T_i^*+1} \\
&= \omega_i + \sigma^2_{i,T_i^*+1}(\alpha(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta),
\end{aligned}$$

where $(\alpha(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta)$ may be greater than 1, permitting explosive behavior, at least in the short term. After both shocks have been exhausted, their influence disappears quickly. This short-memory effect has implications for the method being developed herein. First, there may be different risks associated with overestimating and underestimating level shock and volatility shock lengths. Therefore, estimation of effects among the donors should err on the side of underestimating, not overestimating, the length of the max shock, since overestimation of the shock length brings with it the risk of underestimating ω^*. Second, a practitioner of the method needs some idea of how long the the respective shocks in the time series under study might last. There are couple of obvious strategies: take all the donors, and over all the donor shock lengths, take the minimium. Alternatively, one could take the maximum.

4.1. Consistency of the shock estimators

The estimators $\hat{\omega}^*_{i,T_i^*+1}$ in (2) are central to our forecast adjustment strategy. Here we show that the quantities $\hat{\omega}^*_{i,T_i^*+1}$ possess an important asymptotic property that bolsters that overall reliability of the method proposed herein.

Proposition 2. *Assume*

1. *For each i, $\{a_{i,t}\}_{t=0,...,T_i}$ obeys a GARCH-X(m, s), as laid out in Equation (1), with volatility shocks found in $\mathcal{M}_1$, where T_i is the length of the ith series.*
2. *For each $i, \{D^{vol}_{i,t}\omega^*_{i,t}\}_{t=0,...,T_i}$ is potentially non-zero at $\{T_i^*+1, ..., T_i^*+L_i^{vol}\}$, $\omega^*_{i,T_i^*+1} \equiv ... \equiv \omega^*_{i,T_i^*+L_i^{vol}}$, and zero otherwise, where the arrival of T_i^* is governed by a time-invariant distribution on $\{a_{i,t}\}_{t=0,...,T_i-1}$, and both the arrival and conclusion of the shock is observable by the researcher.*
3. *The conditions in Assumption 0 of Han and Kristensen (2014) hold.*

Then for any $i, 1 \leq i \leq n+1$, and for any $r, 1 \leq r \leq k$, $\hat{\omega}^*_{i,T_i^*+r} \xrightarrow{p} \omega^*_{i,T_i^*+r}$ as $t \to \infty$. Additionally, $\hat{\omega}^*_{i,*} \xrightarrow{p} \omega^*_{i,T_i^*+r}$ as $t \to \infty$. That is, whether estimated individually or collectively, the nonzero shock estimators are consistent.

Lemma 1. *Under assumption 2, for each $i, i = 1, ..., n+1$, $\{\omega^*_{i,t}\}_{t=0,...,T_i}$ is a strictly stationary series.*

4.2. Consistency of the Conditional Forecast Function

Having established the consistency of the estimators $\hat{\omega}^*_{i,T^*_i+1}$, we extend that result to prove the consistency of the conditional forecast function itself.

Proposition 3. *Assume*

1. *All conditions listed in Proposition 2.*

2. *There exist weights $\{\pi_i\}_{i=2}^{n+1}$ such that $\boldsymbol{v}_{1,T^*_1} = \sum_{i=2}^{n+1} \pi_i \boldsymbol{v}_{i,T^*_i}$.*

Then for any r, $1 \leq r \leq k$, $\hat{\sigma}^2_{adjusted,T^*_i+r} \xrightarrow{p} \sigma^2_{1,T^*_i+r}$ as $t \to \infty$.

4.3. Asymptotic Loss

We now evaluate the loss and risk of our method under two scenarios: first, under arbitrary distribution of σ^2_{t+1}, and then second, under the assumption that the data-generating process is correctly specified. For simplicity, we suppress the index i on all quantities.

For 1-step-ahead forecast of σ^2_{t+1} where $t = T^*$, consider the difference

$$\begin{aligned}
&QL(\hat{\sigma}^2_{t+1,unadjusted}, \sigma^2_{t+1}) - QL(\hat{\sigma}^2_{t+1,adjusted}, \sigma^2_{t+1}) \\
&= (\frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,unadjusted}} - \log \frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,unadjusted}} - 1) - (\frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,adjusted}} - \log \frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,adjusted}} - 1) \\
&= \frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,unadjusted}} - \frac{\sigma^2_{t+1}}{\hat{\sigma}^2_{t+1,adjusted}} + \log \frac{\hat{\sigma}^2_{t+1,unadjusted}}{\hat{\sigma}^2_{t+1,adjusted}} \\
&= \frac{\sigma^2_{t+1}(\hat{\sigma}^2_{t+1,adjusted} - \hat{\sigma}^2_{t+1,unadjusted})}{\hat{\sigma}^2_{t+1,adjusted}\hat{\sigma}^2_{t+1,unadjusted}} + \log \frac{\hat{\sigma}^2_{t+1,unadjusted}}{\hat{\sigma}^2_{t+1,adjusted}} .
\end{aligned} \tag{4}$$

For simplicity, we work with a GARCH(1,1) that experiences a volatility shock at a single time point for which we would like to provide a point forecast. Then (4) can be expressed as

$$\frac{\sigma^2_{t+1}\hat{\omega}^*_{t+1}}{\hat{\sigma}^2_{t+1,adjusted}\hat{\sigma}^2_{t+1,unadjusted}} + \log \frac{\hat{\omega} + \hat{\alpha}a^2_t + \hat{\beta}\sigma^2_t}{\hat{\omega} + \hat{\alpha}a^2_t + \hat{\beta}\sigma^2_t + \hat{\omega}^*_{t+1}}.$$

It is easily verified that as $\hat{\omega}^*_{t+1} \to 0^+$, the difference in the losses goes to zero. On the other hand, as $\hat{\omega}^*_{t+1}$ becomes large, the difference in the losses turns negative, with the lesson being that $\hat{\omega}^*_{t+1}$ must be in appropriate proportion to σ^2_{t+1} in order for the adjusted forecast to outperform the unadjusted forecast. This explains why it is so important to avoid using a naive adjustment estimator, $\overline{\omega^*}$, the arithmetic mean of the estimated shocks. We conclude this section with a broader result.

Proposition 4. *Assume the conditions in Propositions 2 and 3. Let $1 \leq i \leq n+1$. Then as $t \to \infty$,*

$$QL(\hat{\sigma}^2_{t+1,unadjusted}, \sigma^2_{t+1}) - QL(\hat{\sigma}^2_{t+1,adjusted}, \sigma^2_{t+1}) \xrightarrow{p} \frac{\omega^*_{i,t+1}}{\sigma^2_{t+1} - \omega^*_{i,t+1}} + \log \frac{\sigma^2_{t+1} - \omega^*_{i,t+1}}{\sigma^2_{t+1}} \geq 0.$$

Hence, a correctly specified $\mathcal{M}_1$ model will outperform the unadjusted forecast asymptotically.

5. Numerical Examples

In this section, we demonstrate the effectiveness of the proposed method using Monte Carlo simulations. All simulations will use $\mathcal{M}_1$ volatility models and $\mathcal{M}_0$ models on the returns, i.e. $D_{i,t}^{return} \equiv 0$ for each i and each t. We first explain the parameters to be varied, the parameters that remain fixed, and the behavior we expect to observe. For our purposes, a parameter is defined to be fixed whenever it does not vary across any of the simulations performed, and conversely, a parameter is varying whenever it varies across at least two simulations performed. The overarching story told by the simulations is that our method performs well as the magnitude of the signal in the shock grows relative to the variance of the idiosyncratic noise term of the shock.

5.1. Fixed Parameters

Each simulation, $i = 1, ...n + 1$ is a GARCH(1,1) process of length T_i chosen randomly from the set of integers $\{756, ..., 2520\}$, corresponding to approximately 3-10 years of daily financial data. We fixed the intercept of the processes at the *garchx* package default of $\omega = .2$ (Sucarrat, 2020). We also use the values $\alpha = .1, \beta = .82$, corresponding to a GARCH process with greater influence from past values of $\sigma_{i,t}^2$ than past values of $a_{i,t}^2$.

5.2. Varying Parameters

We vary exactly five parameters: $\mu_V, \sigma_V, \mu_\delta, \mu_{\omega^*}, \sigma_u$, each of which is critical to evaluating the method's responsiveness to changing conditions in the shock distribution. μ_V, σ_V govern the elements of the covariates. μ_δ interacts with $\mathbf{V}_t$ via the dot-product operation, of course. μ_{ω^*} is a location parameter for the volatility shock, and σ_u is the standard deviation of the volatility shock.

We add an important note about the parameter μ_δ, which governs a vector δ of length p with elements that increase monotonically in proportion to $1, ..., p$, after which they are scaled by the factor $\frac{2 \cdot \mu_\delta}{p(p+1)}$, so that a randomly selected entry of the vector has mean μ_δ. The heterogeneity of the entries in δ is critical to simulating the plausible heterogeneity that will exist in the random effects structure. Absent a heterogeneous vector δ, the shock would not vary systematically with regard to each variable in the covariate vector. Such a setup would fail to approximate the realities of financial markets.

5.3. Evaluative Framework and Monte Carlo Results

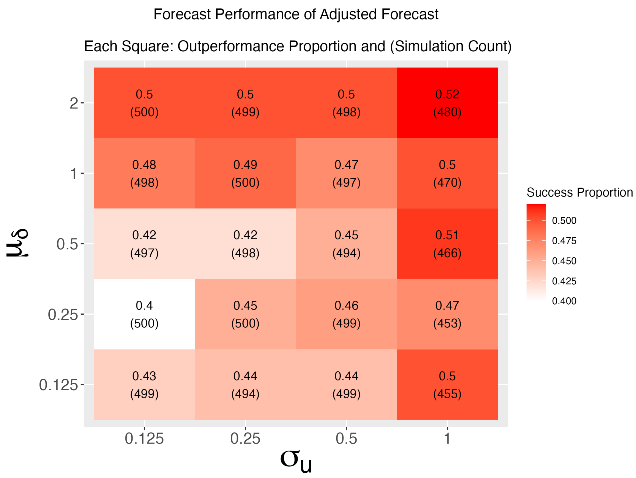
Consistent with our evaluative framework declared in 3.4, we compare adjusted and unadjusted forecasts using QL Loss, calculating the fraction of the simulations in which the adjusted forecast yields a QL Loss no smaller than that of the unadjusted forecast. The simulations fall into three broad categories, which we will denote **Signal and Noise** (displayed in Figure 2), **Interaction between Signal and Volatility Profile Means** (displayed in Figure 3), and lastly, **Interaction between Shock Intercept and Shock Noise** (displayed in Figure 4).

The first category plots two parameters of the shock equation, μ_δ and σ_u, in grids while increasing the mean of the covariate vectors $\mathbf{v}$ over three plots. Recall that δ is a parameter shared between donors, whereas each vector $\mathbf{v}_{i,t}$ in $\mathbf{V}_{i,t}$ is i.i.d. multivariate normal. In Figure 2a, we see very little variation across the grid. However, in Figures 2b and 2c, where μ_v is increased, we observe an expected phenomenon: for any fixed column of the plots, the performance of the adjusted forecast improves with increasing μ_δ. A somewhat subtler phenomenon is also observed: for a fixed row of the grids in Figures 2b and 2c, for $\mu_\delta \geq .5$, increasing σ_u leads to a decline in performance. However, this is not observed for smaller levels of μ_δ, where the signal is too weak.

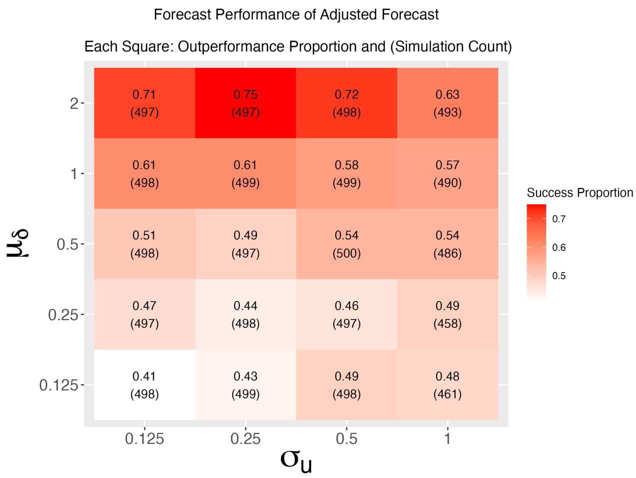
The second category of simulations examines μ_δ and μ_v, while increasing the noise σ_u. The row-wise and column-wise increases in performance that we witnessed in the first category largely hold in the second category of simulations, with one similar disclaimer: the relationship is strained under large levels of noise, like $\sigma_u = 1$.

Lastly, the third set of simulations takes a look at the role of the intercept of the conditional shock distribution, μ_{ω^*}, in particular, its interaction with σ_u. We can think of μ_{ω^*} as being the sole component of $\mathbf{v}$ that has no variance, i.e. because it's corresponding entry in the vector $\mathbf{v}$ is the scalar 1. For larger values of μ_{ω^*}, the outperformance of the adjusted forecast declines with increases in σ_u, but this effect is absent for smaller values of μ_{ω^*}

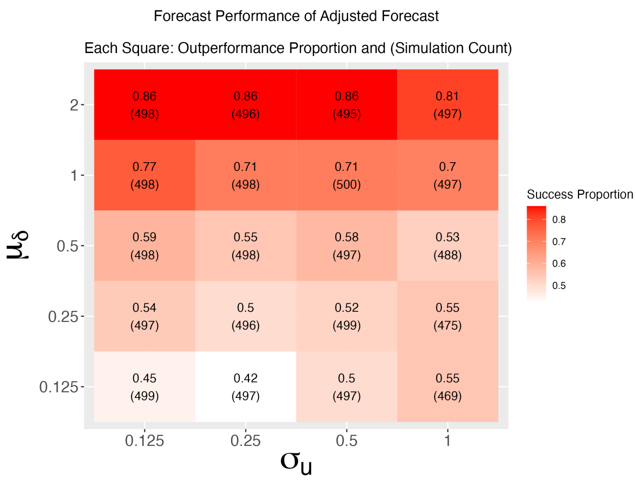
Signal and Noise



(a) Fixed values: $\mu_v = .125, \sigma_v = .125, \mu_{\omega^*} = .125$



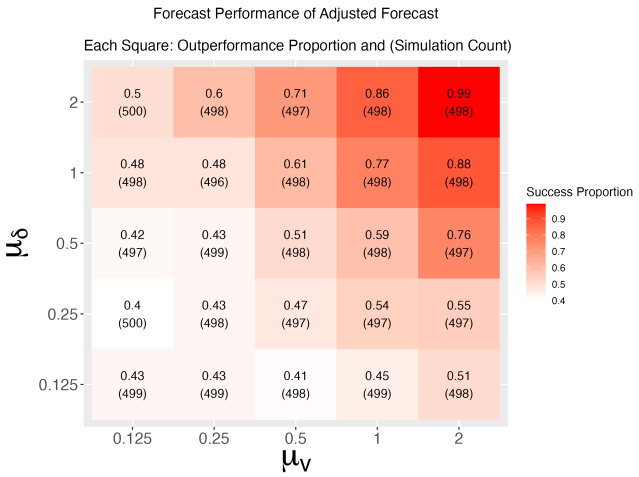
(b) Fixed values: $\mu_v = .5, \sigma_v = .125, \mu_{\omega^*} = .125$



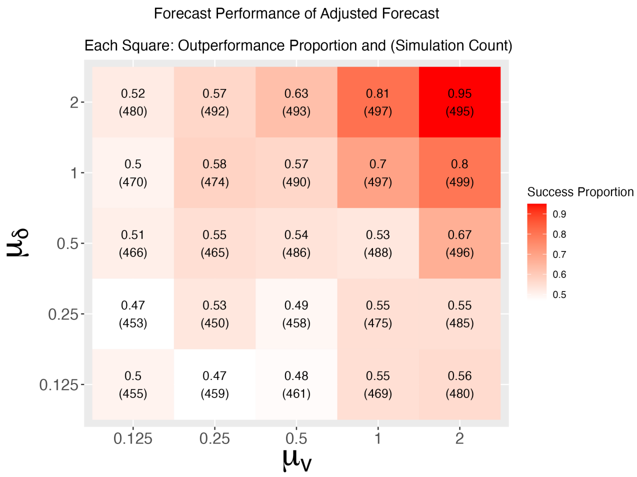
(c) Fixed values: $\mu_v = 1, \sigma_v = .125, \mu_{\omega^*} = .125$

Figure 2: In the progression from 2a to 2b to 2c, we see that increasing μ_δ leads to improved performance of the adjusted forecast, and this effect is intensified as μ_v increases. However, it is not consistently true that an increasing noise undermines the performance of the adjusted forecast. Rather, that phenomenon requires both large quantites of μ_δ and μ_v.

Interaction between Signal and Volatility Profile Means



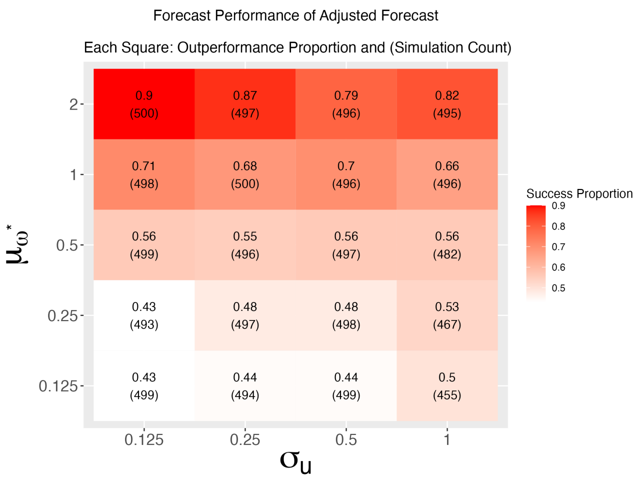
(a) Fixed values: $\sigma_v = .125, \mu_{\omega^*} = .125, \sigma_u = .125$



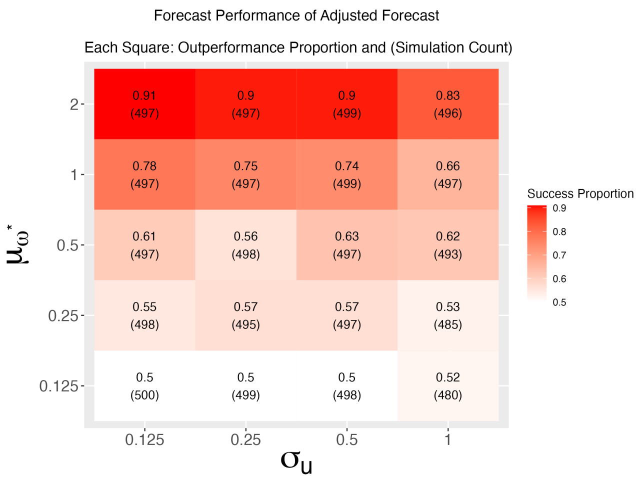
(b) Fixed values: $\sigma_v = .125, \mu_{\omega^*} = .125, \sigma_u = 1$

Figure 3: We compare the interaction of μ_δ and μ_v at two different levels of shock noise. In the low-noise regime, the peak performance of the adjusted forecast is higher, and the ascent is faster along both dimensions. However, in the high-noise regime, the adjusted forecast performs well even at low levels of μ_δ and μ_v.

Interaction between Shock Intercept and Shock Noise



(a) Fixed values: $\mu_v = .125, \sigma_v = .125, \delta = .125$



(b) Fixed values: $\mu_v = .125, \sigma_v = .125, \delta = 2$

Figure 4: The interaction between the intercept μ_{ω^*} and σ_u suggests that the parameters behave as expected. However, a larger value of δ in 4b attenuates the effect of increasing noise.

6. Forecasting after the results of the 2016 US presidential election

We demonstrate the applicability of our methodology using a real data set at the intersection of financial trading and electoral politics. The goal of this analysis will be to provide a forecast for the volatility of financial services ETF IYG[1] (an ETF composed of American financial majors JPMorgan, Bank of America, etcetera) on Wednesday November 9th, 2016 using only the information available at the close of US financial markets at 4pm ET on November 8th, 2016, either after Donald Trump won the 2016 US presidential election or in a scenario in which Trump would win the day. Our specifications for this analysis are outlined below:

1. **Model choice** We assume a GARCH(1,1) for the daily log return series of IYG in each donor. As argued in Hansen and Lunde (2005), a GARCH(1,1) is rarely dominated by more heavily-parameterized GARCH specifications. It thus provides a defensible choice when motivation or time for choosing another model is lacking. For the time series under study and the donor series alike, we fit a GARCH(1,1) on almost four years of market data prior to the shock.

2. **Covariate Choice** We choose covariates that could plausibly satisfy the model assumptions spelled out earlier, that is, risk-related and macroeconomic covariates that could plausibly be weighted and summed in a shock distribution. We thus choose the log return Crude Oil (CL.F), the VIX (VIX) and the log return of the VIX, the log returns of the three-month, five-year, ten-year, and thirty-year US Treasuries. Finally, we include the squares of the demeaned log return of IYG for the three trading days preceding the shocks, in an effort to capture the 'local' ARCH effects prior to the shocks. For each variable in the volatility profile, we compute the sample mean and sample standard deviation across the $n+1$ events, allowing us to scale the variables to have zero mean and unit variance. Hence, no single variable can dominate the distance-based weighting procedure.

3. **Donor pool construction** We choose the three most recent US presidential elections prior to the 2016 election as well as several Brexit-related events in the first half of 2016. The inclusion of the three US presidential elections is straightforward and easily justified. Those three elections are the only presidential elections since the advent of the ETF IYG, and it is not unreasonable that three elections held between 2004 and 2012 would share a conditional shock distribution with the time series under study in 2016. We exclude the midterm congressional elections in the US (i.e. those held in even years not divisible by four), which generate far lower voter turnout and feature no national races. The inclusion of Brexit-relate donors is less straightforward but perhaps more interesting for our method. We first explain the three donors and then justify

[1] It has been noted that GARCH effects are more attenuated in aggregated returns (Zivot, 2009), which suggests against using the S&P 500 or similar indices as an example.

their inclusion. The first Brexit-related donor is former UK Prime Minister David Cameron's announcement, on February 20th, 2016, that a referendum would occur to decide the UK's future in the EU. The second is a poll released on June 14th, 2016 in the UK reflecting higher support for Brexit than had been thought. Last is the Brexit vote itself, which occurred on June 23rd, 2016, with polls closing that evening and markets reacting the next morning. Even though the movement to withdraw from the EU was a long-running, complex phenomenon that occurred outside of the United States, nevertheless the Leave campaign possessed a very similar right-wing populism to that of the Trump campaign (Wilson, 2017). Furthermore, Brexit preceded the 2016 US presidential election by less than five months and was, according to several news outlets, seen as a harbinger of Trump's victory (Kay, 2016; Collins, 2016). This view was made clear by Anne Applebaum who stated: "I do realize that it's facile to talk about the impact on a U.S. election which is still many months away, that it's too simple to say "first Brexit, then Trump." But there is a way in which this election has to be seen, at the very least, as a possible harbinger of the future. This referendum campaign, as I wrote a few days ago, was not fought on the issues that are normally central to British elections. Identity politics trumped economics; arguments about "independence" and "sovereignty" defeated arguments about British influence and importance. The advice of once-trusted institutions was ignored. Elected leaders were swept aside. If that kind of transformation can take place in the U.K., then it can happen in the United States, too. We have been warned" (Applebaum, 2016). However, it is important to note that the Leave decision foretelling a Trump victory was not a consensus view (Martin and Burns, 2016; Barro, 2016). In our forecasting setting, these three Brexit donors will be an imperfect donors, as will the preceding US presidential elections which feature different candidates and campaigns.

4. **Choice of estimator for volatility** We use the sum of squared five-minute log returns of IYG on November 9th, 2016, otherwise known as the Realized Volatility (RV) estimator of volatility (Andersen and Benzoni, 2008), as our proxy. We exclude the first five minutes of the trading day, resulting in a sum of 77 squared five-minute returns generated between 9:35am and 4pm.

5. **Data Sources** All daily market data is provided via the YahooFinance API available in the quantmod package in R (Ryan, Ulrich, Thielen, Teetor, Bronder, and Ulrich, 2015). In order to calculate log equity returns, we use Adjusted Close of IYG. The realized volatility is computed using high-frequency quote data available from Wharton Research Data Services (WRDS) (Wachowicz, 2020). The spread between the AAA and BAA yields is provided by Federal Reserve Economic Data (FRED) and accessed via the quantmod package.

We now discuss results displayed in the three subplots in Figure 5 in order from left to right. On the left, we see that distanced-based weighting places the largest weights on the 2004 and 2012 elections and Brexit-relatex donors, with only modest weight on the 2008 election. Assuming an approximately correct specification of the covariates, this is interpreted to mean that even of the 2016 US election had a general climate of risk and tension less extreme than 2008 and more similar to conditions of the other donors. The three Brexit donors also have large, similar weights, which is not entirely surprising, as they were temporally proximate both to one another as well as the 2016 US presidential election. In the middle plot, we notice that the fixed effect estimates for the 2008 election is considerable, while the 2012 election and three Brexit donor estimates are modest by comparison, with only 2004 registering a near-zero fixed effect estimate. On the right, we observe in black the $\hat{\sigma}^2$ fitted by a GARCH(1,1) for the time series under study. We also observe four colored points, each listed in the legend: three predictions and the ground truth. We include the prediction derived by adjusting the GARCH(1,1) prediction by the arithmetic mean of the fixed effect estimates. As is evident, our method comes remarkably close the ground truth. The prediction is not only directionally correct, i.e. we predict a volatility spike where there is one; the prediction also outperforms the unadjusted prediction. Remarkably, the arithmetic-mean based prediction here demonstrates the inherent risk in failing to weight each donor appropriately. The 2008 election receives far more weight than is called for, as simple averaging ignores the radically different conditions on the evening of those two events.

Naturally, one might ask how sensitive this prediction is to at least two kinds of model specification: donor pool specification and covariate specification. There are two responses to these concerns. First, although the practitioner lacks a priori knowledge of the adequacy of the donors with respect to the time series under study, it is possible to gauge the diversity of the donor information by examining the singular values of the volatility profile. Second, we follow Steegen et al. (2016) in a executing a multiverse analysis. In the supplement we carry out leave-one-out analyses on both the donor set and the covariate set. Our multiverse analysis concludes that our adjustment approach beats the unadjusted forecast in every leave-one-out configuration.

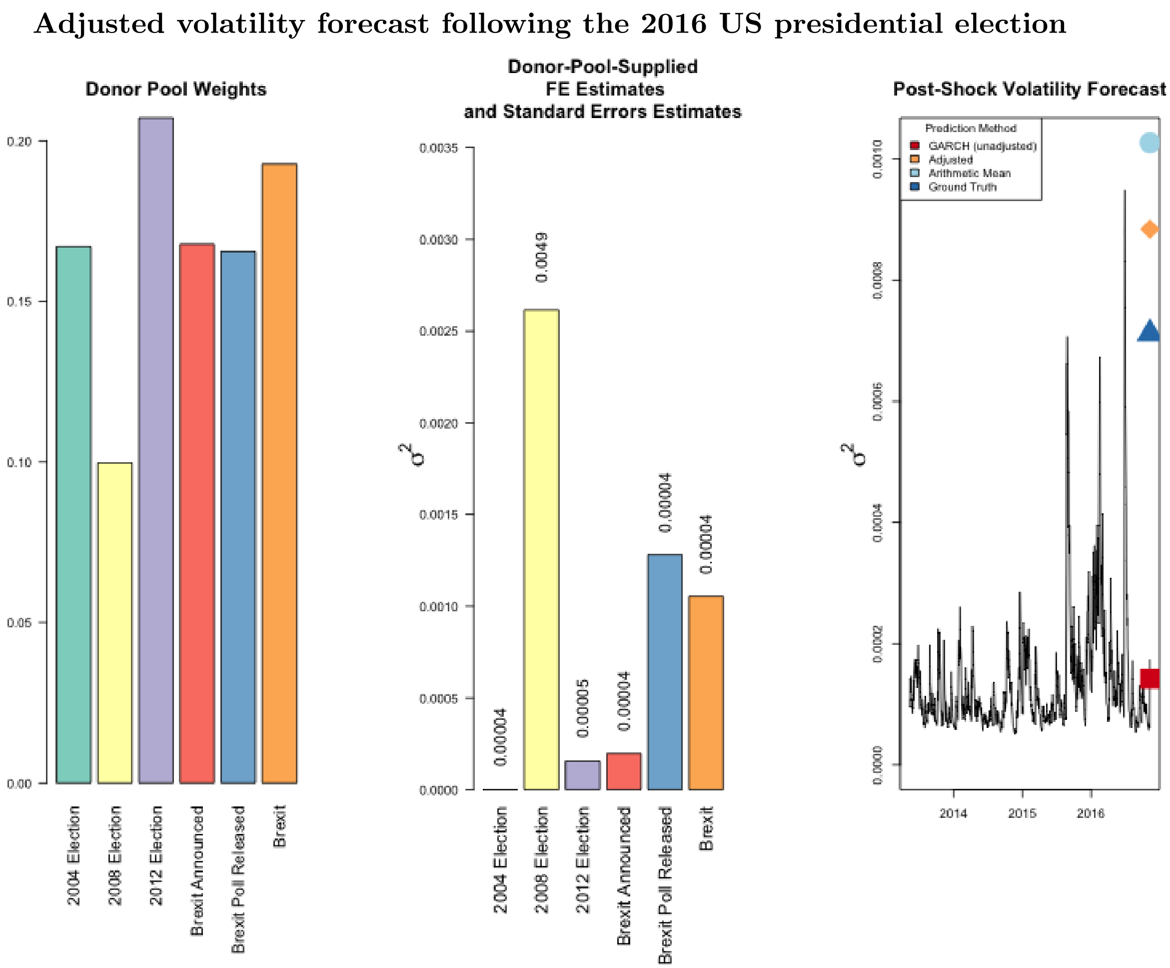


Figure 5: Donor pool weights (left panel), individual shock effects (center panel), and forecasts (right panel).

7. Discussion

This present work applies and innovates techniques in similarity-based forecasting and information aggregation in order to provide better GARCH forecasts under news shocks. It extends Lin and Eck (2021) principally by substituting a GARCH model for an AR(1), i.e. by modeling both the mean and volatility of a univariate time series. An interesting connection insight related to Lin and Eck (2021) is that the GARCH model, under weak assumptions, can be represented as an ARMA on the squared residuals a_t^2. Hence, a shock $\omega_{i,t}^*$ to the volatility at time t is an identical shock to a_t^2. However, this use-case becomes less attractive in situations where the sign of the time series under study following the shock is uncertain. Insofar as this work has made advances that accommodate heteroskedastic time series, the benefits may redound most amply to applications like Value-At-Risk (VaR) and Expected Shortfall, where $\hat{\sigma}_t^2$ is an input.

This work as well as Lin and Eck (2021) itself can be viewed as a generalization of Phillips (1996) where the rare event probability λ is assumed to be 1 and the error associated with the rare event is estimated using external series. In our setting, if the shock to the time series under study were known and yet so underdescribed, i.e. so lacking in qualitative context, then Phillips (1996) would be recommended. However, in our setting the rare event is not only rare but also contains specific qualitative features that are best accounted for via data aggregation.

The method under development does not strictly require knowledge of the length of the shocks in the donor pool, but correctly sizing up those shock lengths is helpful to proper estimation of the shocks in the donor pool. An important question remains: even if the donor pool shock lengths are assumed to be known, how do we advise the operator to forecast the time series under study? In other words, for how long is the adjustment estimator ω^* valid, applicable, and reliable? One idea suggested by the paper is obvious: why not aggregate the shock lengths from the donors as well and round that quantity or take the floor or ceiling of any non-integer value? This is worth pursuing. However, it may be that estimating the persistence of a volatility shock induced by news is an endeavor deserving of its own study, where aggregation methods might naturally arise as helpful tools.

There is also a broader discussion to be had regarding the degree of model heterogeneity permitted in fitting the donors' series.

7.1. Comparison with KNN

Clements and Hendry (1996) (cited in Guerrón-Quintana and Zhong (2017)) note intercept corrections find their theoretical basis in the occurrence of structural breaks, whereas Nearest-Neighbor methods, being nonparametric, are theoretically more adept at accounting for nonlinearity. The present work examines news shocks, which are more closely related to structural breaks. Hence, neither nearest neighbor methods nor nonlinearity figure heavily in our work. However, there are deeper observations to be made about KNN as it relates to our method.

The method presented here is unlike traditional KNN in that we are not trying to learn a function, first and foremost. We are trying to estimate a parameter. KNN runs into the problem: the curse of dimensionality. In contrast, large p is not a problem in synthetic methods, because the thing estimated is the vector w with $n-1$ degrees of freedom. For KNN, a high-dimensional space, i.e. large p, corresponding to many covariates, is a difficult space in which to work with distances (Hastie, Tibshirani, and Friedman, 2009). In contrast, large p is not a problem in and of itself for synthetic control — in fact, asymptotic results exist for p (Abadie et al., 2010).

As is pointed out in Hastie et al. (2009), KNN regression performs well when K is chosen small enough that one can simply average the points $\{y_i\}_{i=1}^{N_{train}}$ in the neighborhood around each element in $\{y_i\}_{i=1}^{N_{test}}$ to get good predictions. As we have noted above, the arithmetic mean-based estimator of ω^*, denoted $\overline{\omega^*}$, corresponds to KNN when $K = n$, the number of donors. Fundamentally, the idea that n is small enough and the donors are homogeneous enough that one could simply average the $\hat{\omega}_i$ is at odds with the assumed variation in the shocks.

In KNN regression, the hyperparameter K must be learned. In similarity-based parameter correction, the number of donors is not learned. A donor pool is curated, and then careful rules of thumb can be applied to determine whether a given donor should be included or excluded. While it would not necessarily hurt to 'learn' the appropriate number of donors to use, this information would probably not be as useful as knowing which donors and covariates provide the basis for the best forecasts. This brings us to a deeper point about the distinction between similarity-based methods in the style of Lin and Eck (2021) and KNN. In KNN, the exogenous variables are taken as a given and nearness to the object to be predicted depends on the distance function chosen. In contrast, in Lin and Eck (2021), the determination of nearness begins with a qualitative step, i.e. curating the units between which we will calculate distances and from we will ultimately derive weights.

7.2. Donor Pool Construction

Should we gather as many donors as possible and pick them quantitatively? It would be counter to the method proposed to assemble a vast number of donors, lacking careful scrutiny of the qualitative fit, and let the optimization simply pick the donors, via weighting. What makes a donor good is not merely its quantitative fit but its qualitative fit as well. Abadie and Vives-i Bastida (2022) make a similar point about large donor pools.

What matters is that the donors chosen are properly situated in the p-dimensional predictor space, so as to allow proper estimation of the weights. That being said there may be subjectivity in selecting donors, and care will be needed to make good selections. In this work we included Brexit and the previous three US presidential elections as donors used for forecasting volatility following the 2016 US presidential election. The resulting adjusted forecast with this donor pool performed very well. In our supplement we show that adjusted forecasts perform well under a variety of donor pool choices and covariate choices.

7.3. The Nature and Estimation of Volatility Shocks

Not all of the volatility of an asset return may be related to news (Boudoukh et al., 2019). This explains our inclusion of an idiosyncratic noise term in the shock specification. However, this point also gestures in the direction of possible unexplained variation in the shocks. Chinco, Clark-Joseph, and Ye (2019) find that for predicting 1-minute returns, highly transitory firm-specific news is useful. The authors conclude that news about fundamentals is predictive.

It would be a pyrrhic victory for our method if the volatility profile indeed underlies real-world shocks but the volatility profile is radically high-dimensional or the signal in the shocks is overwhelmed by the noise term. Even unbiased predictions can be unhelpful if they are high in variability, and in Section 8.4, we show how the benefits of forecast combination. Another possibility is high-frequency data and the use of linear models like HAR (Corsi, Audrino, and Renó, 2012). HAR would not only increase the sample size available for discovering the autoregressive structure of a series' realized volatility. It would also open the door to high-dimensional regression methods, shrinkage estimators, and more.

7.4. Parameter Stability

There is an important question about the stability of the GARCH parameters under the presence of a shock, and on parameter instability as it pertains to forecasting, we refer readers to Rossi (2013, 2021). There are at least two reasons that we do not explore parameter instability or methods to adjust for it. First, the marginal effect of coefficient changes at the shock time would, under the assumptions in this work, be swamped by the random effect. Second, the estimation of post-shock parameter values would require at least several — better yet, dozens — of post-shock data points, whereas this work assumes access to zero post-shock data points. However, it is possible that similarity-based estimators for the GARCH coefficients could be produced, for example, by adapting the methods of Dendramis et al. (2020).

8. Supplement

Here we provide additional useful information about the foregoing results and analyses. We begin with several sections that re-examine the results of our real data example through a multiverse perspective. We also explore forecast combination as a tool for real data analysis by evaluating the mean and median of the full set of forecasts generated by dropping one covariate and one donor at a time. Finally, we provide proofs for the numerous formal results found earlier in this work.

8.1. Leave-one-out: How we analyzed a multiverse of 63 predictions

Given the option of redoing our analysis, with eight covariates and six donors, a natural question from a place of skepticism is, how stable are the results — how contingent are they on a particular specification? To answer these questions, in Table 1 we generate all 63 predictions yielded the decision of leave out any one of the donors (or none) and any one of the covariates (or none). This analysis includes the prediction presented above in Figure 5, which can be viewed as the null model, at least in the sense that it provides and baseline for comparison.

8.2. Sensitivity to Covariates Chosen

Any covariate that appears near to the bottom of Table 1 more often is, in a greedy sense, a more important covariate for this prediction task, since by dropping it, a larger loss results. Similarly, any covariate that appears more often in forecasts below the null are covariates that lead to a worse forecast than the null. The covariate that thus stands out in positive way is the squared, demeaned log return of IYG, of which we use the three days preceeding the date T_i^*. This suggests that the behavior of the observable time series itself (and its transformation) may be a useful proxy for risk conditions, a result wholly consistent with the core consistency result for Synthetic Control found in Abadie et al. (2010).

8.3. Sensitivity to Donors Chosen

The most glaring result visible in Table 1 regarding donor selection is two-fold: on one hand, omitting the donor that is the Brexit Poll of June 13th, 2016 leads to the worst QL Loss quantities, far worse than having the full set of donors. It could very well be that the poll released that day attenuates the contribution of Brexit itself, and without accounting for the variety of Brexit-related events in 2016, a poor prediction results. Hence, the inclusion of all three Brexit donors is justified.

On the other hand, the 2004 election appears most often at the top of the table, suggesting that is an unhelpful donor. This is unsurprising if one considers that the fixed effect estimate for 2004 is nearly zero. That means that by dropping the 2004 election, other donors must account for the approximately 17% weight it had in the full analysis. The donors that fill that void then have the opportunity to vault the prediction much closer to the ground truth. More fundamentally, however, we should ask why it is that 2004 is such a poor donor, beyond the fact that its fixed effect is small. At a deeper level, it could be that the 2004 election delivered such little surprise or is so far distant, temporally, that its inclusion in our donor pool is not fully justified.

8.4. Can we combine forecasts to outperform the individual forecasts?

We consider a very simple forecast combination procedure – in fact, we consider two: the average forecast and the median forecast. While forecast combination can be justified in any number of ways, here we invoke forecast combination as a way to robustify forecasts against misspecification. We combine in two simply ways, using the mean of all 63 forecasts and the median of all 63 forecasts. The forecast losses are all specifications – including the average and median forecasts – are presented in Table 1.

8.5. Formal Results

Proof of Proposition 1 We claim

$$\mathbb{E}[\sigma^2_{i,T_i^*+r+1}|\mathcal{F}_{T_i^*+r}] = \mathbb{E}[\omega_i + \alpha a^2_{T_i^*+r} + \beta\sigma^2_{i,T_i^*+r}|\mathcal{F}_{T_i^*+r}] \tag{5}$$

$$= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T^*+r}\epsilon_{T_i^*+r})^2|\mathcal{F}_{T_i^*+r}] + \beta\sigma^2_{i,T_i^*+r} \tag{6}$$

$$= \omega_i + \alpha\sigma^2_{i,T_i^*+r} + \beta\sigma^2_{i,T_i^*+r} \tag{7}$$

$$= \omega_i + (\alpha+\beta)\sigma^2_{i,T_i^*+r} \ . \tag{8}$$

The volatility equation of a GARCH(1,1) dictates that for any r, the one-step-ahead volatility is given by the expression inside the expectation in (5). By the mean-model assumption of a GARCH(1,1), we have $a_{i,t} = \sigma_{i,t}\epsilon_{i,t}$, and hence by substituting $\sigma_{i,t}\epsilon_{i,t}$ for $a_{i,t}$, we arrive at equation (6) above. Using the unit-variance assumption on $\epsilon_{T_i^*+r}$, we can compute explicitly the expectation, yielding (7). Finally, by rearranging terms, we arrive at equation (8). $\square$

Proof of Lemma 1 Since the shock is assumed to arrive uniformly at random for each i, $1 \leq i \leq n+1$, and last for a discrete number of indices, the sequence $\{\omega_{i,t}^*\}_{t=0,\ldots,T_i}$ is governed by a distribution $F_{\{\omega_{i,t}^*\}_{t=0,\ldots,T_i}}$ that is invariant to shifts in time. $\square$

Proof of Proposition 2 The result follows from the consistency proof of the QMLE in GARCH-X models, as established by Han and Kristensen (2014). $\square$

Proof of Proposition 3 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:

$$\mathbb{E}[\sigma_{i,t+1}^2|\mathcal{F}_t] = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t+1-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t+1-j}^2 + \gamma_i^T \mathbf{v}_{i,t+1} + \omega_{i,*}^* \ .$$

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

$$\hat{\sigma}_{i,t+1}^2|\mathcal{F}_t = \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a_{i,t+1-k}^2 + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \hat{\sigma}_{i,t+1-j}^2 + \hat{\gamma}_i^T \mathbf{v}_{i,t+1} + \hat{\omega}_{i,*}^* \ ,$$

which converges in probability to

$$\hat{\sigma}_{i,t+1}^2|\mathcal{F}_t = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t+1-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t+1-j}^2 + \gamma_i^T \mathbf{v}_{i,t+1} + \omega_{i,*}^*$$

as $t \to \infty$ by a simple application of Slutsky's Theorem. $\square$

Proof of Proposition 4 First, note that the function $g : (-\infty, \omega_i^*) \to \mathbb{R}$ given by $g(x) = \frac{x}{\sigma_{t+1}^2 - x} + \log \frac{\sigma_{t+1}^2 - x}{\sigma_{t+1}^2}$ is nonnegative, convex, obtains a minimum at $x = 0$, and being continuous, preserves consistency. The conclusion follows from fact that the model is correctly specified and consistency of $\hat{\sigma}_{t+1,adjusted}^2$ is guaranteed Proposition (3). $\square$

Table 1: Adjusted volatility forecasts for the 2016 US presidential election ranked by QL loss. This table contains all forecasts made by applying simultaneous leave-one-out approaches made to the covariates and the donors forming the donor pool. The forecast from our analysis in the main text is included in yellow. The mean and median of all forecasts is also included in grey.

QL Loss of the Adjusted Forecast	Omitted Covariate	Omitted Donor
0.0001	Log Return of ^IRX	None
0.0007	NA (Average Adjusted Forecast)	NA (Average Adjusted Forecast)
0.0007	Log Return of CL=F	2004-11-02
0.0008	Log Return of ^IRX	2004-11-02
0.0034	Log Return of ^TNX	None
0.0060	Log Return of ^TNX	2004-11-02
0.0061	Log Return of ^VIX	2016-02-19
0.0077	Raw^VIX	2004-11-02
0.0079	None	2004-11-02
0.0080	Log Return of CL=F	2008-11-04
0.0099	Log Return of ^VIX	2004-11-02
0.0099	NA (Median Adjusted Forecast)	NA (Median Adjusted Forecast)
0.0115	Log Return of CL=F	2016-02-19
0.0127	None	2016-02-19
0.0137	Raw^VIX	2016-02-19
0.0151	Log Return of ^TNX	2016-02-19
0.0160	Log Return of ^TYX	2016-02-19
0.0193	Raw^VIX	2016-06-23
0.0194	Log Return of ^VIX	2016-06-23
0.0216	Log Return of ^IRX	2016-06-23
0.0219	None	None
0.0222	Log Return of ^FVX	2016-02-19
0.0232	Log Return of ^VIX	2008-11-04
0.0237	Log Return of ^TNX	2016-06-23
0.0240	Log Return of ^IRX	2016-02-19
0.0245	Log Return of ^TNX	2008-11-04
0.0251	None	2016-06-23
0.0280	Log Return of CL=F	2016-06-23
0.0315	Log Return of ^FVX	2008-11-04
0.0322	None	2008-11-04
0.0360	Log Return of ^TYX	2008-11-04
0.0393	Log Return of ^VIX	2012-11-06
0.0396	Raw^VIX	2008-11-04
0.0473	Log Return of ^IRX	2008-11-04
0.0484	IYG	2004-11-02
0.0499	Log Return of ^FVX	2016-06-23
0.0542	Log Return of ^TYX	2016-06-23
0.0578	IYG	2012-11-06
0.0595	Log Return of ^VIX	None
0.0596	Raw^VIX	None
0.0598	Log Return of ^TYX	None
0.0598	IYG	None
0.0598	Log Return of ^FVX	None
0.0604	Log Return of ^TYX	2004-11-02
0.0627	Log Return of ^FVX	2004-11-02
0.0651	Log Return of CL=F	None
0.0692	Log Return of ^VIX	2016-06-13
0.0758	Log Return of ^IRX	2012-11-06
0.0922	IYG	2008-11-04
0.0924	Log Return of ^TNX	2012-11-06
0.0990	None	2012-11-06
0.1038	Raw^VIX	2012-11-06
0.1117	Log Return of ^IRX	2016-06-13
0.1147	Log Return of CL=F	2012-11-06
0.1241	Raw^VIX	2016-06-13
0.1356	IYG	2016-06-23
0.1359	Log Return of ^TYX	2012-11-06
0.1480	Log Return of ^FVX	2012-11-06
0.1621	None	2016-06-13
0.1630	Log Return of ^TNX	2016-06-13
0.1928	Log Return of CL=F	2016-06-13
0.2102	Log Return of ^FVX	2016-06-13
0.2435	Log Return of ^TYX	2016-06-13
0.3013	IYG	2016-02-19
1.0261	IYG	2016-06-13
2.3529	All	All

References

Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, 2003.

Alberto Abadie and Jaume Vives-i Bastida. Synthetic controls in action. *arXiv preprint arXiv:2203.06279*, 2022.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105 (490):493–505, 2010.

Torben G Andersen and Luca Benzoni. Stochastic volatility. *CREATES Research Paper*, (2010-10), 2010.

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

Torben G Andersen, Tim Bollerslev, and Francis X Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701–720, 2007.

Torben Gustav Andersen and Luca Benzoni. Realized volatility, working paper 2008-14. 2008.

Anne Applebaum. Britain's decision to leave the e.u. is a warning to america, 2016. URL https://www.washingtonpost.com/opinions/global-opinions/after-brexit-what-will-and-wont-happen/2016/06/24/c9f7a2f6-39f1-11e6-8f7c-d4c723a2becb_story.html?postshare=9121466767991000. Published: 2016-06-24.

Josh Barro. 3 reasons brexit does not portend a trump victory — and one reason to worry, 2016. URL https://www.businessinsider.com/what-does-brexit-mean-for-trump-2016-6?op=1. Published: 2016-06-24.

Luc Bauwens, Arie Preminger, and Jeroen VK Rombouts. Regime switching garch models. 2006.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Jacob Boudoukh, Ronen Feldman, Shimon Kogan, and Matthew Richardson. Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, 32(3):992–1033, 2019.

George Box. Box and jenkins: time series analysis, forecasting and control. In *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*, pages 161–215. Springer, 2013.

Christian T Brownlees, Robert F Engle, and Bryan T Kelly. A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915*, 2011.

Alex Chinco, Adam D Clark-Joseph, and Mao Ye. Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492, 2019.

Bent J Christensen and Nagpurnanand R Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150, 1998.

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.

Stephen Collins. Trump on brexit: America is next, 2016. URL https://www.cnn.com/2016/06/25/politics/donald-trump-brexit-scotland. Published: 2016-06-25.

Fulvio Corsi, Francesco Audrino, and Roberto Renó. Har modeling for realized volatility forecasting. 2012.

Giovanni De Luca et al. Forecasting volatility using high-frequency data. *Statistica Applicata*, 18, 2006.

Yiannis Dendramis, George Kapetanios, and Massimiliano Marcellino. A similarity-based approach for macroeconomic forecasting. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(3):801–827, 2020.

Kathryn ME Dominguez and Freyan Panthaki. What defines 'news' in foreign exchange markets? *Journal of International Money and Finance*, 25(1):168–198, 2006.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

Robert F Engle and Andrew J Patton. What good is a volatility model? *Quantitative finance*, 1(2):237, 2001.

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.

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.

Peter R Hansen and Asger Lunde. A forecast comparison of volatility models: does anything beat a garch (1,1)? *Journal of applied econometrics*, 20(7):873–889, 2005.

Peter Reinhard Hansen, Zhuo Huang, and Howard Howan Shek. Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906, 2012.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

D Hendry and M Clements. On a theory of intercept corrections in macroeconometric forecasting. 1994.

Katty Kay. Five reasons brexit could signal trump winning the white house, 2016. URL https://www.bbc.com/news/election-us-2016-36564808. Published: 2016-06-20.

Will Kenton. Event-driven investing strategies and examples, 2022. URL https://www.investopedia.com/terms/e/eventdriven.asp.

Lutz Kilian and Helmut Lütkepohl. *Structural vector autoregressive analysis*. Cambridge University Press, 2017.

Jilei Lin and Daniel J Eck. Minimizing post-shock forecasting error through aggregation of outside information. *International Journal of Forecasting*, 2021.

Jonathan Martin and Alexander Burns. Is 'brexit' the precursor to a donald trump presidency? not so fast, 2016. URL https://www.nytimes.com/2016/06/25/us/politics/is-brexit-the-precursor-to-a-donald-trump-presidency-not-so-fast.html. Published: 2016-06-24.

Stewart Mayhew. Implied volatility. *Financial Analysts Journal*, 51(4):8–20, 1995.

Robert F Phillips. Forecasting in the presence of large shocks. *Journal of Economic Dynamics and Control*, 20(9-10):1581–1608, 1996.

Christina D Romer and David H Romer. Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual*, 4:121–170, 1989.

Barbara Rossi. Advances in forecasting under instability. In *Handbook of economic forecasting*, volume 2, pages 1203–1324. Elsevier, 2013.

Barbara Rossi. Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them. *Journal of Economic Literature*, 59(4):1135–1190, 2021.

Jeffrey A Ryan, Joshua M Ulrich, Wouter Thielen, Paul Teetor, Steve Bronder, and Maintainer Joshua M Ulrich. Package 'quantmod'. 2015.

Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712, 2016.

Genaro Sucarrat. garchx: Flexible and Robust GARCH-X Modelling. MPRA Paper 100301, University Library of Munich, Germany, May 2020. URL https://ideas.repec.org/p/pra/mprapa/100301.html.

Allan Timmermann. Forecast combinations. *Handbook of economic forecasting*, 1:135–196, 2006.

Ruey S Tsay. *Analysis of financial time series*, volume 543. John wiley & sons, 2005.

Erin Wachowicz. Wharton research data services (wrds). *Journal of Business & Finance Librarianship*, 25(3-4):184–187, 2020.

Lu Wang, Feng Ma, and Guoshan Liu. Forecasting stock volatility in the presence of extreme shocks: Short-term and long-term effects. *Journal of Forecasting*, 39(5):797–810, 2020.

Xiaoqian Wang, Rob J Hyndman, Feng Li, and Yanfei Kang. Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547, 2023.

Graham K Wilson. Brexit, trump and the special relationship. *The British Journal of Politics and International Relations*, 19(3):543–557, 2017.

Eric Zivot. Practical issues in the analysis of univariate garch models. In *Handbook of Financial Time Series*, pages 113–155. Springer, 2009.