

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

David Lundquist\*, Daniel J. Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

July 3, 2024

## Abstract

We develop a procedure for forecasting the volatility of a time series immediately following a news shock. Adapting the similarity-based framework of ?, 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 aggregation of these estimates is done in service of adjusting the  $h$ -step-ahead 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). A real-world application is provided, as are simulation results suggesting the conditions and hyperparameters under which our method thrives.

## 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 [?]. 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, ? 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.

In this work we focus on the second central moment of a time series, the volatility. One of the most important stochastic phenomena of 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 [?], 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 ? 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  $(r_t)_{t \in \mathbb{N}}$  over time.

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 ? 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

---

\*davidl11@illinois.edu

†dje13@illinois.edu

positive returns [?]. Asymmetry and mixed-frequency are employed in ? 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 [?], which 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 [??], 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 [?].

The approach herein sidesteps the functional complexity posited by Realized GARCH, with its minimum nine parameters to estimate [?], 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 an appropriate procedure for pooling those estimates. The use of fixed effect estimation for the study of structural shocks has a pedigree in macroeconomic analysis (? cited in ?; see also discussion of deterministic exogenous events in ?). 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 rare events (see ? 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 [??]. 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. ? use weighted subsets of a scalar series’ own past to correct forecast errors by an additive term. ? introduces a similarity-based forecasting procedure for time-varying coefficients of a linear model. ? 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, involves an optimization step inspired by the causal inference literature. In particular, in ?, 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 either the treatment or control unit. 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. ? adapt these methods for the purpose of prediction. Their one-step-ahead forecasts use distance-based-weighting to pool shock estimates from similar series according to the donor series’ similarity to the series under study. Their approach does not take into account the ARCH effects commonly observed in time series, especially financial time series, leaving unaccounted for the variability that accompanies predictions of a heteroskedastic time series. Outside of ?, 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 confirm 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 most closely-followed democratic political event on earth, the 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 Election in the United States. 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.

## 2 Setting

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

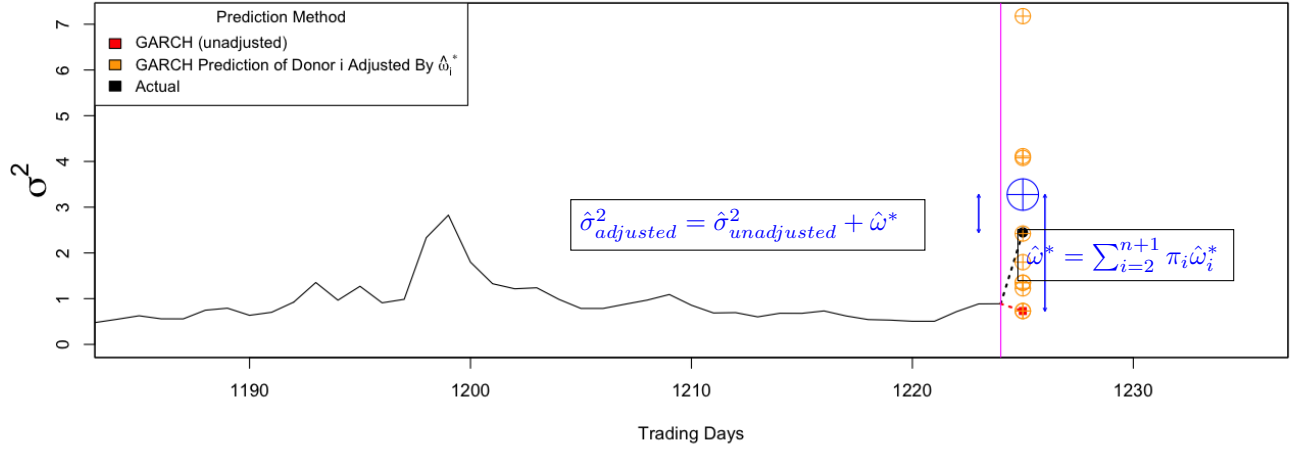


Figure 1: The simulated time series experiences a volatility shock between trading days 1223 and 1224. We have truncated the series after  $T^* + 1 = 1224$ . The GARCH prediction, in red, fails even to approach the volatility spike at  $T^* + 1$ , as do some adjusted predictions, which are orange. In contrast, the GARCH forecast adjusted by  $\hat{\omega}^*$ , a convex combination of the estimated shocks in the donor pool, achieves directional correctness as well as a smaller absolute loss in its prediction.

### 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  $\text{ARIMA}(p, d, q)$  models [?] provides a framework for modeling the autoregressive structure of  $r_t$ , all under the umbrella of frequentist statistics. 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 [?] and GARCH [?] models provide 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  $\mu_t$  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 so-called ARCH effects, that is, tests for the clustering of volatility. These tests explore the alternative hypothesis that  $\sigma_t^2$  is not only a time-varying parameter but furthermore a function of past squared residuals of the mean model. In particular, the  $\text{ARCH}(m)$  model is an autoregressive model in which  $\sigma_t^2$  is a deterministic function of the past  $m$  values of  $r_t^2$ . The  $\text{GARCH}(m, s)$  framework take this one step further by modeling  $\sigma_t^2$  as a linear combination of the past  $m$  values of

$r_t^2$  and well as the past  $s$  values of  $\sigma_t^2$ . In functional form, a GARCH process (sometimes called a strong GARCH process [?, p. 19]) is given by

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

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

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

## 2.2 Model setup

We will suppose that a researcher has multivariate time series data  $\mathbf{y}_{i,t} = (r_{i,t}, \mathbf{v}_{i,t})$ ,  $t = 1, \dots, T_i$ ,  $i = 1, \dots, n+1$ , 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 a 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 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  $i = 1, \dots, n+1$ , and let  $T_i^*$  denote the largest time index prior to the arrival of the news shock, with  $T_i^* < T_i$ . Let  $\delta, \mathbf{v}_{i,t} \in \mathbb{R}^p$ . For  $t = 1, \dots, T_i$  and  $i = 1, \dots, n+1$ , the model  $\mathcal{M}_1$  is defined as

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

with error structure

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

where  $D_{i,t}^{\text{return}} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,\text{return}}\})$  and  $D_{i,t}^{\text{vol}} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,\text{vol}}\})$  and  $L_{i,\text{return}}, L_{i,\text{vol}}$  denote the lengths of the log return and volatility shocks, respectively. Let  $\mathcal{M}_0$  denote the subclass of  $\mathcal{M}_1$  models such that  $\delta \equiv 0$ . Note that  $\mathcal{M}_0$  assumes that  $\omega_i^*$  have no dependence on the covariates and are i.i.d. with  $\mathbb{E}[\omega_i^*] = \mu_{\omega^*}$ .

## 2.3 Volatility Profile of a Time Series

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

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

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

where RV and IV denote realized volatility and implied volatility, respectively,  $\Delta$  denotes the lag operator between consecutive realizations of a variable, and *Volume* denotes the quantity of traded shares or other units of an asset. Covariates chosen for inclusion in a given volatility profile may be levels, log differences in levels, percentage changes in levels, or absolute values thereof, among many choices.

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

## 3 Methodology for Similarity-based Parameter Correction

### 3.1 Forecasting

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

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

A GARCH model is an ARMA on the squares of the observed scalar time series  $a_t^2$  [??, 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 GARCH model is, just like for an ARMA model, the conditional expectation function,  $\mathbb{E}[\sigma_{i,T^*+h}^2 | \mathcal{F}_{T^*}]$ , or practically speaking, the estimate thereof,  $\hat{\mathbb{E}}[\sigma_{i,T^*+h}^2 | \mathcal{F}_{T^*}]$  [?]. Here we have presented one-step-ahead forecasts

for a GARCH-X(1,1), without loss of generality. For  $h = 2, 3, 4, \dots$ , the conditional expectation is computed recursively, as is standard for iterative autoregressive forecasts.

### 3.2 Excess Volatility Estimators

The problem of aggregating estimated donor shocks begins with the data constraints. Taking the estimated shocks as a given, we essentially observe the pair  $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ . We wish to recover weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$  leading to favorable forecasting properties. These weights are used to compute  $\hat{\omega}^* := \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ , our forecast adjustment term. Since the weights  $\{\pi_i\}_{i=2}^{n+1}$  are computed using  $\mathcal{F}_{T_i^*}$ , the set  $\{\pi_i\}_{i=2}^{n+1}$  is deterministic, *modulo* any stochastic ingredient in the numerical methods employed to approximate  $\mathbf{v}_{1,T^*}$  using a convex combination of donor covariates. We say more about the properties of  $\omega_i^*$  in section 4.

Following ??, let  $\|\cdot\|_{\mathbf{S}}$  denote any semi-norm on  $\mathbb{R}^p$ , and define

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

In the forecast combination literature, it is discussed whether the weights employed to aggregate forecasts strive toward and meet various optimality criteria [??]. In this work, there are at least two senses of optimal weights that one might be interested in. First, we can think of optimal weights as a set  $\{\pi_i\}_{i=2}^{n+1}$  such that  $\omega_1 = \sum_{i=2}^{n+1} \pi_i \omega_i$ , i.e.,  $\omega_1$  is recovered perfectly, as it belongs to convex hull of the estimated shocks. However,  $\omega_1$  is never revealed to the practitioner, and hence there is no way of verifying the extent to which this condition is satisfied.

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

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

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

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

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

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

### 3.3 Ground Truth Estimators

The time-varying parameter  $\sigma_t^2$  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 [?]. The realized variance itself can be decomposed into the sum of a continuous component and a jump component, with the latter being less predictable and less persistent [?], cited in ?, two factors that further motivate the method employed herein.

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

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

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  $\tilde{r}(t, 1) \stackrel{iid}{\sim} N(\mu, \delta^2)$ , 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.

### 3.4 Loss Functions

We are interested in point forecasts for  $\sigma_{1,T^*+h}^2 | \mathcal{F}_{T^*}$ ,  $h = 1, 2, \dots$ , 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

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

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

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

Finally, we introduce the QL (quasi-likelihood) Loss [?]:

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

What distinguishes QL Loss is that it is multiplicative rather than additive. This has benefits, both practical and theoretical. As ? 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

The model  $\mathcal{M}_1$  is defined by a volatility equation and mean equation, as is any GARCH model. The choice to model the volatility shock  $\omega_i^*$  as an additive random effect is straightforward. However, the choice to model the level effect  $\epsilon_{i,t}^*$  as a temporary rupture in the otherwise i.i.d. sequence of innovations  $\epsilon_{i,t}$  stands in need of deeper justification. 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  $T^*$  and  $T^* + 1$ , we do not have an easy way to express a conditional distribution of

the innovation  $\epsilon_{T^*+1}$  given the overnight arrival of information. Using  $\epsilon_{i,t}^*$  thus breaks this impasse. This defense also explains why we do not parameterize the level shock at  $T^* + 1$  as a sum of two shocks,  $\epsilon_{i,T^*+1}$  and  $\epsilon_{i,T^*+1}^*$ , which would represent the level shock as generated by two independent sources of stochasticity. To do so would be inelegant and would also lack motivation as a practical level. While we want to model the shock at  $T^* + 1$  as potentially large in absolute value, we also want to retain the property of a unitary source of noise.

Note that under the popular GARCH(1,1), a dual level-volatility shock has an marginal effect on the conditional variance  $\sigma_{i,t}^2$  that reflects the geometric decay of innovations in autoregressive models. As usual, assume  $\alpha + \beta < 1$ . Furthermore, assume that both the volatility shock  $\omega_i^*$  and the level shock  $\epsilon_{i,t}^*$  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

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

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

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

By repeated substitution, in conditional expectation, the shock is  $\mathcal{O}((\alpha + \beta)^r)$ . 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  $r \geq \max\{L_{i,vol}, L_{i,level}\} + 1$ ,*

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

**Proof of Proposition** We claim

$$\mathbb{E}[\sigma_{i,T^*+r+1}^2 | \mathcal{F}_{T^*+r}] = \mathbb{E}[\omega_i + \alpha a_{T^*+r}^2 + \beta \sigma_{i,T^*+r}^2 | \mathcal{F}_{T^*+r}]\tag{3}$$

$$= \omega_i + \mathbb{E}[\alpha (\sigma_{i,T^*+r} \epsilon_{T^*+r})^2 | \mathcal{F}_{T^*+r}] + \beta \sigma_{i,T^*+r}^2\tag{4}$$

$$= \omega_i + \alpha \sigma_{i,T^*+r}^2 + \beta \sigma_{i,T^*+r}^2\tag{5}$$

$$= \omega_i + (\alpha + \beta) \sigma_{i,T^*+r}^2 .\tag{6}$$

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 (3). 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 (4) above. Using the unit-variance assumption on  $\epsilon_{T^*+r}$ , we can compute explicitly the expectation, yielding (5). Finally, by rearranging terms, we arrive at equation (6).  $\square$

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



$$\begin{aligned}
\mathbb{E}[\sigma_{i,T^*+2}^2 | \mathcal{F}_{T^*+1}] &= \mathbb{E}[\omega_i + \alpha a_{T^*+1}^2 + \beta \sigma_{i,T^*+1}^2 | \mathcal{F}_{T^*+1}] \\
&= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T^*+1} \epsilon_{T^*+1}^*)^2 | \mathcal{F}_{T^*+1}] + \beta \sigma_{i,T^*+1}^2 \\
&= \omega_i + \alpha \sigma_{i,T^*+1}^2 (\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta \sigma_{i,T^*+1}^2 \\
&= \omega_i + \sigma_{i,T^*+1}^2 (\alpha(\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta) .
\end{aligned}$$

where  $(\alpha(\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta)$  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 over/underestimating level shock and vol shock lengths. Estimation of effects in donor pool 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  $\omega^*$ . Second, a practitioner of the method needs some idea of how long the operator expects the respective shocks in the time series under study to be. There are couple of obvious strategies: take all the donors, and over all the donor shock lengths, take the minimum. Alternatively, one could take the maximum.

## 4.1 Consistency of the fixed effect estimators

**Proposition 2.** *Assume*

1. For each  $i$ ,  $\{a_{i,t}\}_{t=0,\dots,T_i}$  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  $i$ th series.
2. For each  $i$ ,  $\{\omega_{i,t}^*\}_{t=0,\dots,T_i}$  is potentially non-zero at  $\{T_i^* + 1, \dots, T_i^* + k\}$ ,  $\omega_{i,T^*+1}^* \equiv \dots \equiv \omega_{i,T^*+k}^*$ , and zero otherwise, where the arrival of  $T_i^*$  is governed by a time-invariant distribution on  $\{a_{i,t}\}_{t=0,\dots,T_i-1}$ .
3. The conditions in Assumption 0 of ? prevail.

Then for any  $i$ ,  $\hat{\omega}_{i,t}^* \xrightarrow{p} \omega_{i,T^*+1}^*$  as  $t \rightarrow \infty$ .

**Lemma 1.** Under assumption 2,  $\forall i$ ,  $\{\omega_{i,t}^*\}_{t=0,\dots,T_i}$  is a strictly stationary series.

**Proof of Lemma** 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  $\{\omega_{i,t}^*\}_{t=0,\dots,T_i}$  is governed by a distribution  $F_{\{\omega_{i,t}^*\}_{t=0,\dots,T_i}}$  that is invariant to shifts in time.  $\square$

**Proof of Proposition** The result follows from the consistency proof of the QMLE in GARCH-X models, as established by ?.  $\square$

## 4.2 Consistency of the Forecast Function

**Proposition 3.** *Assume*

1. All conditions listed in Proposition 2.
2. There exist weights  $\{\pi_i\}_{i=2}^{n+1}$  such that  $\mathbf{v}_{1,T^*} = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T^*}$ .

Then  $\hat{\sigma}_{adjusted}^2 \xrightarrow{p} \sigma_{1,T^*+1}^2$  as  $t \rightarrow \infty$ .

**Proof of Proposition** Recall the conditional expectation of the variance for the GARCH-X( $m, s$ ) model:

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

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

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

which converges in probability to

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

as  $t \rightarrow \infty$  by a simple application of Slutsky's Theorem. □

### 4.3 Asymptotic Loss

We now evaluate the loss and risk of our method under two scenarios: first, under arbitrary distribution of  $\sigma_{t+1}^2$ , and then second, under the assumption that the data-generating process is correctly specified.

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

$$\begin{aligned} & QL(\hat{\sigma}_{t+1,unadjusted}^2, \sigma_{t+1}^2) - QL(\hat{\sigma}_{t+1,adjusted}^2, \sigma_{t+1}^2) \\ &= \left( \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - 1 \right) - \left( \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} - 1 \right) \\ &= \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} + \log \frac{\hat{\sigma}_{t+1,unadjusted}^2}{\hat{\sigma}_{t+1,adjusted}^2} \\ &= \frac{\sigma_{t+1}^2 (\hat{\sigma}_{t+1,adjusted}^2 - \hat{\sigma}_{t+1,unadjusted}^2)}{\hat{\sigma}_{t+1,adjusted}^2 \hat{\sigma}_{t+1,unadjusted}^2} + \log \frac{\hat{\sigma}_{t+1,unadjusted}^2}{\hat{\sigma}_{t+1,adjusted}^2}. \end{aligned} \quad (7)$$

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 (7) can be expressed as

$$\frac{\sigma_{t+1}^2 \hat{\omega}_{t+1}^*}{\hat{\sigma}_{t+1,adjusted}^2 \hat{\sigma}_{t+1,unadjusted}^2} + \log \frac{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta} \sigma_t^2}{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta} \sigma_t^2 + \hat{\omega}_{t+1}^*}$$

It is easily verified that as  $\hat{\omega}_{t+1}^* \rightarrow 0^+$ , 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 the volatility  $\sigma_{t+1}^2$  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,  $\bar{\omega}^*$ , 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  $t \rightarrow \infty$*

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

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

**Proof of Proposition** First, note that the function  $g : (-\infty, \omega_i^*) \rightarrow \mathbb{R}$  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  $\sigma_{adjusted}^2$  is guaranteed Proposition (3). □

## 5 Numerical Examples

In this section, we demonstrate the effectiveness of the proposed method using Monte Carlo simulations. In order to investigate the forecasting method presented herein, all of our simulations will use  $\mathcal{M}_1$  volatility models and  $\mathcal{M}_0$  models on the returns, i.e.  $D_{i,t}^{return} \equiv 0$ . 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.

### 5.1 Fixed Parameters

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

### 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.  $\mu_V, \sigma_V$  govern the elements of the volatility profile.  $\mu_\delta$  interacts with  $V$  via the dot-product operation, of course.  $\mu_{\omega^*}$  is a location parameter for the volatility shock, and  $\sigma_u$  is the idiosyncratic noise of the volatility shock.

We add an important note about the parameter  $\mu_\delta$ , which governs a vector  $\delta$  of length  $p$  with elements that increase monotonically in proportion to  $1, \dots, p$ , 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  $\mu_\delta$ . The heterogeneity of the entries is critical to simulating the plausible heterogeneity that will exist in the random effects structure. Absent a heterogeneous vector  $\delta$ , the shock would not vary systematically with regard to each variable chosen for the volatility profile, which 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,  $\mu_\delta$  and  $\sigma_u$ , in grids while increasing the mean of the covariate vectors  $\mathbf{v}$  over three plots. Recall that  $\delta$  is a parameter shared between donors, whereas the vectors  $i$  in the volatility profile are i.i.d. multivariate normal. In Figure 2a, we see very little variation across the grid. However, in Figures 2b and 2c, where  $\mu_v$  is increased, we observe an expected phenomenon: for any fixed column of the plots, the performance of the adjusted forecast improves with increasing  $\mu_\delta$ . 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  $\sigma_u$  leads to a decline in performance. However, this is not observed for smaller levels of  $\mu_\delta$ , where the signal is too weak.

The second category of simulations examines  $\mu_\delta$  and  $\mu_v$ , while increasing the noise  $\sigma_u$ . 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,  $\mu_{\omega^*}$ , in particular, its interaction with  $\sigma_u$ . We can think of  $\mu_{\omega^*}$  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  $\mu_{\omega^*}$ , the outperformance of the adjusted forecast declines with increase noise  $\sigma_u$ , but this effect is absent for smaller values of  $\mu_{\omega^*}$ .

## Signal and Noise

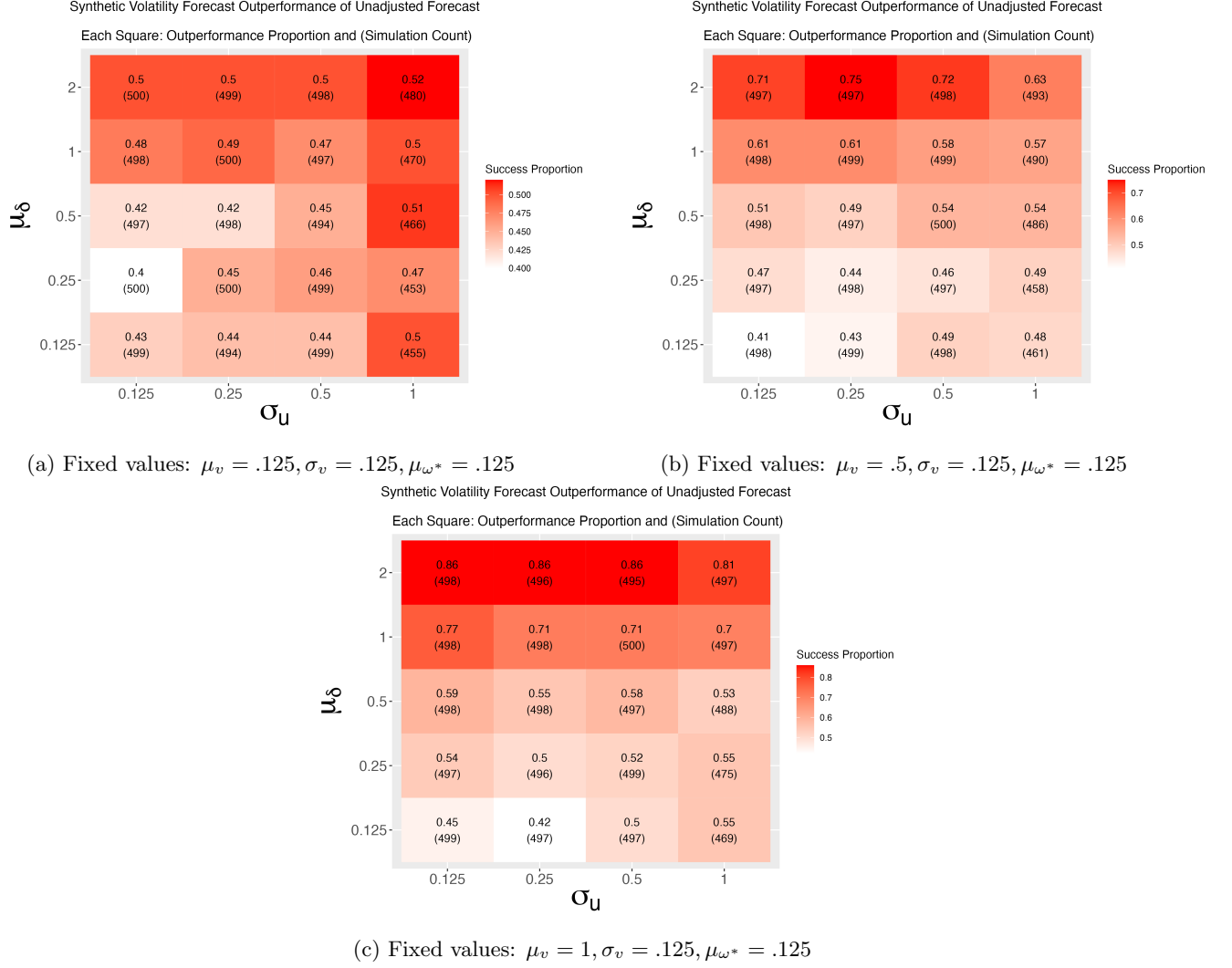


Figure 2: In the progression from 2a to 2b to 2c, we see that increasing  $\mu_\delta$  leads to improved performance of the adjusted forecast, and this effect is intensified as  $\mu_v$  increases. However, it is not consistently true that an increasing noise undermines the performance of the adjusted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

### Interaction between Signal and Volatility Profile Means

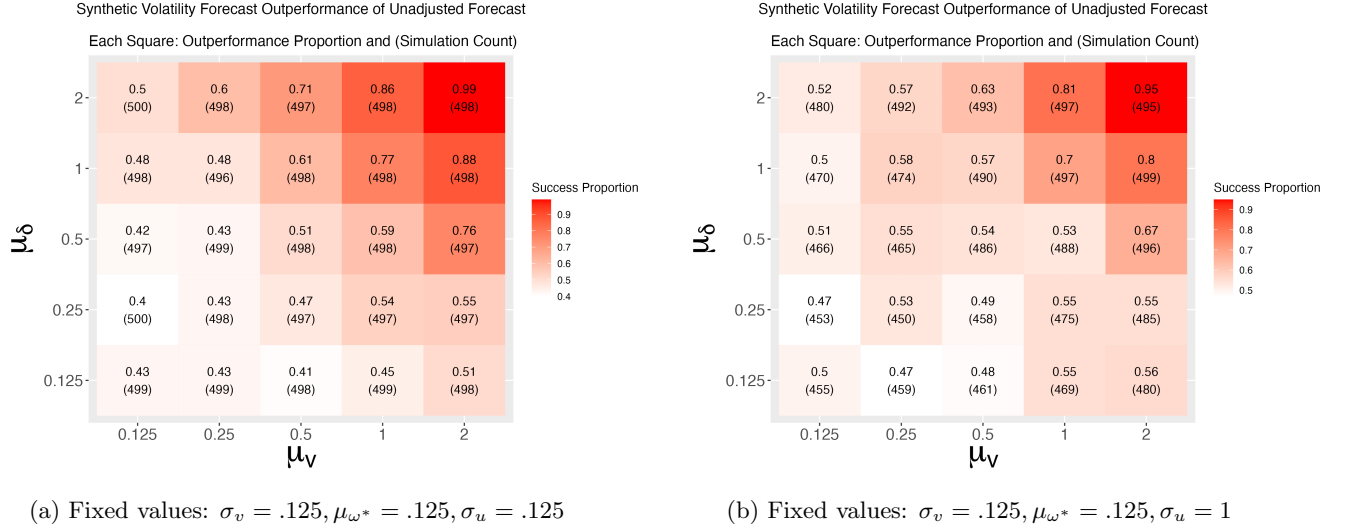


Figure 3: We compare the interaction of  $\mu_\delta$  and  $\mu_v$  at two different levels of shock noise. In the low-noise regime, the peak performance of the adjusted forecast is higher, and the ascent is faster along both dimensions. However, in the high-noise regime, the adjusted forecast performs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

### Interaction between Shock Intercept and Shock Noise

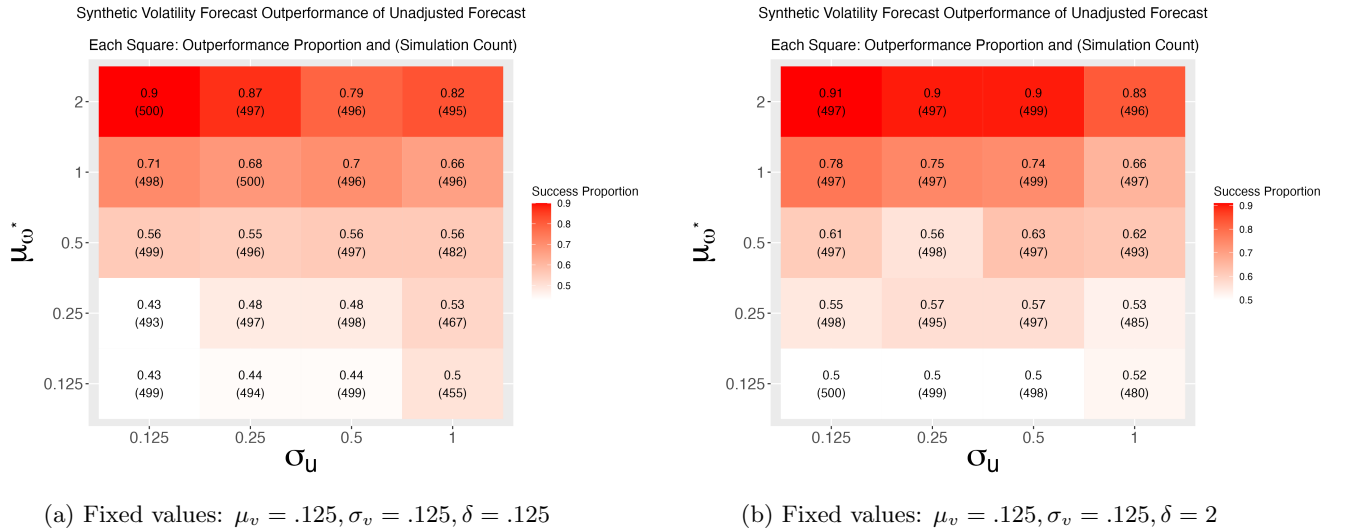


Figure 4: The interaction between the intercept  $\mu_{\omega^*}$  and  $\sigma_u$  suggests that the parameters behave as expected. However, a larger value of  $\delta$  in 4b attenuates the effect of increasing noise.

## 6 Real Data Example

We show the applicability of our method using a real data example that sits at the crossroads of financial trading and electoral politics. In the spring of 2016 in the United States, the Republican Party’s primary election process narrowed down candidates until Donald J. Trump cleared the threshold of votes to win the nomination formally at the party’s convention that summer. He would go on to face the Democratic Party’s nominee, Hillary Rodham Clinton.

From an ex-ante perspective, several qualities of the 2016 US election cycle as well as the candidates themselves made the election difficult to prognosticate. The Electoral College permits victory without a majority or even plurality of the popular vote, which can render presidential races more competitive than a raw vote total would, elevating the uncertainty surrounding the country’s future leadership. The election featured no incumbent, ruling out any incumbent-advantage of the empirical, “statistical” kind distinguished by ?. The Republican Party candidate espoused unorthodox, populist positions on matters such as healthcare, trade, and foreign policy, some of which could be considered rare in either of the major two parties. Additionally, Donald J. Trump, lacking any experience in government — either electoral or appointed service — possessed neither a voting record nor any on-the-job performance for voters to judge or his opponents to attack. As one financial industry professional commented, comparing the 2016 election to the upcoming 2024 election, “this time the markets will be aware of both possibilities and price them to some extent — we wouldn’t expect the same volatility as we saw in 2016 after the election” [?]. Gleaning signals from financial options markets and betting markets, ? predicted that markets would decline prodigiously following a Trump victory in November 2016. Finally, the election outcome delivered significant “news”, in the econometric sense of the word, in the simple sense that it was not predicted. ? found support for the theory that the polling-implied probabilities of election outcomes encode information about future macroeconomic conditions, which is itself reflected in market volatility. In its final post before the election result, acclaimed forecasting outfit 538, headed by economist Nate Silver, predicted a Clinton victory with a probability of .714, more than 2-to-1 odds [?], suggesting that Trump’s victory was at least somewhat surprising. The lack of uncertainty exerts downward pressure on volatility [??], setting the table for a surprise.

For all of these reasons and more, the aftermath of the 2016 presidential election meets the standard of an interesting and notable event for which a quantitative researcher might seek a volatility point prediction. On a more technical level, the election outcome was not known until the evening of election day, well after the closing of financial markets at 4pm Eastern Time. This satisfies the condition that the shock be not yet digested by liquid markets. We therefore proceed to make the following technical specifications in order to predict the volatility of financial services ETF IYG<sup>1</sup> (an ETF composed of American financial majors JPMorgan, Bank of American, etcetera) on Wednesday November 9th, 2016.

1. **Model choice** We assume a GARCH(1,1) for the daily log return series of IYG in each donor. As argued in ?, 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 3-month, 5-year, 10-year, and 30-year US Treasuries, as well as the log return of the most recently available monthly spread between AAA and BAA corporate debt, widely considered a proxy for lending risk [??]. We also include the log return in the trading volume of the ETF IYG itself, which serves as a proxy for panic. Finally, we include the squares of the demeaned log return of IYG for the 30 trading days preceding 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. The three US presidential elections are the only presidential elections since the advent of the ETF IYG. 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.

---

<sup>1</sup>It has been noted that GARCH effects are more attenuated in aggregated returns [?], which suggests against using the S&P 500 or similar indices as an example.

4. **Choice of estimator for volatility** We use the sum of squared 5-minute log returns of IYG on November 9th, 2016, otherwise known as the Realized Volatility estimator of volatility [?], 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 [?]. 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) [?]. The spread between AAA and BAA yields is provided by Federal Reserve Economic Data (FRED) and accessed via the quantmod package.

We now discuss the three subplots in Figure 5 in order from left to right. On the left, we see that distanced-based weighting places nearly equal weight on the 2004 and 2012 elections, with only negligible weight on the 2008 election, when financial market conditions were extreme across nearly every dimension. 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 the 2004 and 2012 elections. In the middle plot, we notice that the fixed effect estimates for 2008 and 2012 are considerable, with only 2004 registering a near-zero fixed effect estimate. The fixed effect estimates quantify the amount of surprise the US election results delivered (strictly speaking, not only the presidential race but all November elections in the US with the ability to influence financial markets) under the assumption of a GARCH(1,1). As estimates gleaned from only one data point per time series, they are theoretically high in variance. On the right, we observe in black the  $\hat{\sigma}^2$  yielded by the 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 reasonably close to the ground truth. The prediction is not only directionally correct, i.e. we predict a volatility spike where there is one; the prediction far 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. In the prediction presented here, the singular values descend from 62% to 23% to 15% of the cumulative variation, indicating a moderate concentration or redundancy of information in the three donors. Second, we follow [?] in executing a multiverse analysis. In particular, 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. These results include an added donor representing Brexit.

## 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 [?] principally by substituting the 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 [?] 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  $\sigma_t^2$  is an input.

This work as well as [?] itself can be viewed as a generalization of [?] where the rare event probability  $\lambda$  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 [?] 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  $\omega^*$  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

? (cited in ?) 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 [?]. In contrast, large  $p$  is not a problem in and of itself for synthetic control — in fact, asymptotic results exist for  $p$  [?].

As is pointed out in ?, 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  $\omega^*$ , denoted  $\bar{\omega}^*$ , 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 shock effects.

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 ? 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 ?, 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. ? 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. For more on this question, see the Supplement, where the donors and the volatility profile are treated with a leave-one-out analysis.

## 7.3 The Nature and Estimation of Volatility Shocks

Not all of the volatility of an asset return may be related to news [?]. 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. ? 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 [?]. 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 ???. There are at least two reasons that we do not herein 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 ?.

## 8 Supplement

### 8.1 Leave-one-out: How we analyzed a multiverse of 50 predictions

Given the option of redoing our analysis, with nine covariates and four 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 fifty 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, 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 more often closer to the bottom of Table 1 is, *ceteris paribus*, a more important covariate for this prediction task, since by dropping it, a larger loss results. The covariates that thus stand out are the demeaned log return of IYG, the VIX, the Debt Risk Spread (spread between between AAA and BAA corporate debt), and last but not least, the choice of dropping none of the covariates. The demeaned log return of IYG, of which we use the 30 days preceeding the  $T_i^*$ , suggests an unusually strong relevance in matching donors to the time series under study. The Debt Risk Spread is also deserving of additional comment and attention, perhaps alongside the poor performance of the log return of the VIX. The Debt Risk Spread is available at a monthly frequency via FRED (add citation), whereas the VIX is available daily via Yahoo Finance. It may be the case that the true conditional shock distribution includes low-frequency data as well as data in levels (like the VIX), while the daily changes in the VIX are absent.

### 8.3 Sensitivity to Donors Chosen

The most glaring result visible in Table 1 regarding donor selection is the poor performance of the 2008 US Election. In hindsight, this is unsurprising, given the large estimated fixed effect for November 5th, 2008 as well as the swirl of complex events occurring in financial markets around that time. It is possible that the GARCH(1,1) used to fit the run-up to that election is underparameterized. The weight given to the 2008 US election is near-zero, but in our method, unsuitable donors can influence the distance-based weighting by affecting that dispersion of one or more of the  $p$  covariates present in the volatility profile. Recall that for each of the  $p$  covariates, we transform the  $n + 1$ -vector of variables to Z-scores. The lesson here is that more careful pre-quantitative inspection of the donors may be warranted as well as ordinary outlier analysis for the volatility profile. As an additional note, dropping the 2008 Election vastly improves the performance of the arithmetic mean forecast. This provides a cautionary tale against using the arithmetic mean estimator for the conditional shock effect.

The donors most conducive to good predictions in this task are the 2004 and 2008 US elections. The estimated fixed effects tell a story of only modest surprise due to these election results: the estimate is nearly zero for 2004, while 2012 is nonzero but small. Given that the presidential election results of these two did not deliver much surprise, it is possible that the surprise could be due more so to non-presidential races.

We discuss one last remarkable phenomenon. The prediction that dropped the preceding 30 days of demeaned log returns of IYG as well as the 2012 Election yielded the same QL Loss as the unadjusted forecast. By inspecting the donor weight and donor fixed effect estimates, we can see that nearly zero adjustment was made because the 2004 Election is given nearly all the weight.

Table 1: Adjusted Forecasts Ranked by QL Loss with Brexit as June 22nd, 2016

QL Loss of the Adjusted Forecast	Omitted Covariate	Omitted Donor
0.0003	Log Return of VIX	2016-06-22
0.0013	Log Return of IRX	2008-11-04
0.0013	Log Return of TYX	None
0.0013	Log Return of TYX	2016-06-22
0.0017	Log Return of TYX	2008-11-04
0.0019	Log Return of TNX	2016-06-22
0.0020	Log Return of TNX	2008-11-04
0.0027	Log Return of FVX	2008-11-04
0.0029	VIX	2008-11-04
0.0034	Log Return of TNX	None
0.0036	Log Return of CL.F	2016-06-22
0.0037	Log Return of VIX	2008-11-04
0.0041	Debt Risk Spread	2016-06-22
0.0043	None	2008-11-04
0.0047	Log Return of FVX	2016-06-22
0.005	NA (Median Forecast)	NA (Median Forecast)
0.0053	VIX	2016-06-22
0.0062	Log Return of CL.F	2008-11-04
0.0068	NA (Average Forecast)	NA (Average Forecast)
0.0070	None	2016-06-22
0.0075	Log Return of FVX	None
0.0077	Log Return of IRX	2016-06-22
0.0102	Log Return of IRX	None
0.0103	Debt Risk Spread	2008-11-04
0.0113	Log Return of VIX	None
0.0135	VIX	None
0.0135	IYG	2008-11-04
0.0136	None	None
0.0137	Log Return of CL.F	None
0.0154	Debt Risk Spread	None
0.0268	IYG	2016-06-22
0.0568	Log Return of TNX	2012-11-06
0.0670	IYG	None
0.0748	Log Return of VIX	2012-11-06
0.0805	Log Return of TNX	2004-11-02
0.0808	Log Return of TYX	2004-11-02
0.0816	Log Return of VIX	2004-11-02
0.0851	Debt Risk Spread	2012-11-06
0.0864	None	2012-11-06
0.0872	VIX	2012-11-06
0.0899	Log Return of IRX	2004-11-02
0.0900	Log Return of CL.F	2012-11-06
0.0907	Log Return of IRX	2012-11-06
0.0913	VIX	2004-11-02
0.0916	Log Return of FVX	2004-11-02
0.0917	None	2004-11-02
0.0925	Debt Risk Spread	2004-11-02
0.0935	Log Return of CL.F	2004-11-02
0.0993	Log Return of TYX	2012-11-06
0.1142	Log Return of FVX	2012-11-06
0.2143	IYG	2004-11-02
2.3529	IYG	2012-11-06
2.3529	All	All

### 8.3.1 Brexit

In the pre-quantitative analysis of donors, there was only one competitor to the setup that was ultimately selected and presented in Section 6, and that competitor donor pool used Brexit. The inclusion of Brexit is based on the fundamental belief that the conditional shock distribution governing IYG’s volatility shocks may be shared among US elections and some political events elsewhere, like the Brexit referendum of July 23rd, 2016. That referendum shared some important qualities with the 2016 US elections, which we have discussed above.

We now turn, however, to why the inclusion of Brexit failed to perform nearly as well our primary specification. First, note time zone differences between the UK and the US make difficult the analysis of after-hour shocks on US markets. Additionally, like many election results, that the “Leave” side of the referendum would prevail was not revealed at any discrete time, of course, but became more certain as observation of voting turnout and informal vote tallies across the UK progressed [?]. This is in spite of UK news outlets adhering to various rules and guidelines regarding reportage of polling results [?]. These facts pose a challenge for any method that uses daily data. Of prime concern is whether we should consider  $T^*$  for Brexit to have occurred after market hours on June 23rd, 2016, Eastern Standard Time, which ignores the steady percolation of information and uncertainty that attend election days, or alternatively, consider the shock to have occurred after hours on June 22nd, 2016, which implies that  $T^* + 1$  is June 23rd, 2016, and hence the reaction of IYG to Brexit can be estimated and extracted from late-day

Table 2: Adjusted Forecasts Ranked by QL Loss with Brexit as June 23rd, 2016

QL Loss of the Adjusted Forecast	Omitted Covariate	Omitted Donor
0.0003	Log Return of VIX	2016-06-23
0.0013	Log Return of TYX	2016-06-23
0.0019	Log Return of TNX	2016-06-23
0.0036	Log Return of CL.F	2016-06-23
0.0041	Debt Risk Spread	2016-06-23
0.0047	Log Return of FVX	2016-06-23
0.0054	VIX	2016-06-23
0.0070	None	2016-06-23
0.0079	Log Return of IRX	2016-06-23
0.0268	IYG	2016-06-23
0.0362	IYG	None
0.1113	IYG	2008-11-04
0.2401	Log Return of TNX	2012-11-06
0.2420	Log Return of FVX	2012-11-06
0.2553	Log Return of TNX	None
0.2678	IYG	2012-11-06
0.2717	Log Return of FVX	None
0.2725	Log Return of TYX	None
0.2766	Log Return of CL.F	None
0.2881	VIX	None
0.2884	Debt Risk Spread	None
0.2905	None	None
0.2954	Log Return of IRX	None
0.2964	Log Return of TYX	2012-11-06
0.3086	Log Return of IRX	2012-11-06
0.3125	NA (Median Forecast)	NA (Median Forecast)
0.3164	VIX	2012-11-06
0.3186	Log Return of CL.F	2012-11-06
0.3219	None	2012-11-06
0.3257	Debt Risk Spread	2012-11-06
0.3289	VIX	2008-11-04
0.3344	Log Return of CL.F	2008-11-04
0.3125	NA (Average Forecast)	NA (Average Forecast)
0.3475	Log Return of VIX	2012-11-06
0.3581	Log Return of IRX	2008-11-04
0.3586	None	2008-11-04
0.3617	Log Return of FVX	2008-11-04
0.3619	Debt Risk Spread	2008-11-04
0.3642	Log Return of TNX	2008-11-04
0.3695	Log Return of TYX	2008-11-04
0.3801	Log Return of VIX	None
0.3902	Log Return of VIX	2008-11-04
0.5510	IYG	2004-11-02
0.7452	Log Return of FVX	2004-11-02
0.7532	Log Return of TNX	2004-11-02
0.7591	Log Return of CL.F	2004-11-02
0.7616	Debt Risk Spread	2004-11-02
0.7681	None	2004-11-02
0.7683	VIX	2004-11-02
0.7753	Log Return of IRX	2004-11-02
0.8034	Log Return of VIX	2004-11-02
0.8618	Log Return of TYX	2004-11-02
2.3529	All	All

trading on that date. Given this dilemma, we opt for dropping Brexit from our predictive model completely.

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

Forecast combination is technique with a vast, sprawling literature that we have referred to above. 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 50 forecasts and the median of all 50 forecasts. The results are presented in Tables 1 and 2.

# Real Data Example: 2016 Election with Brexit as June 22nd, 2016

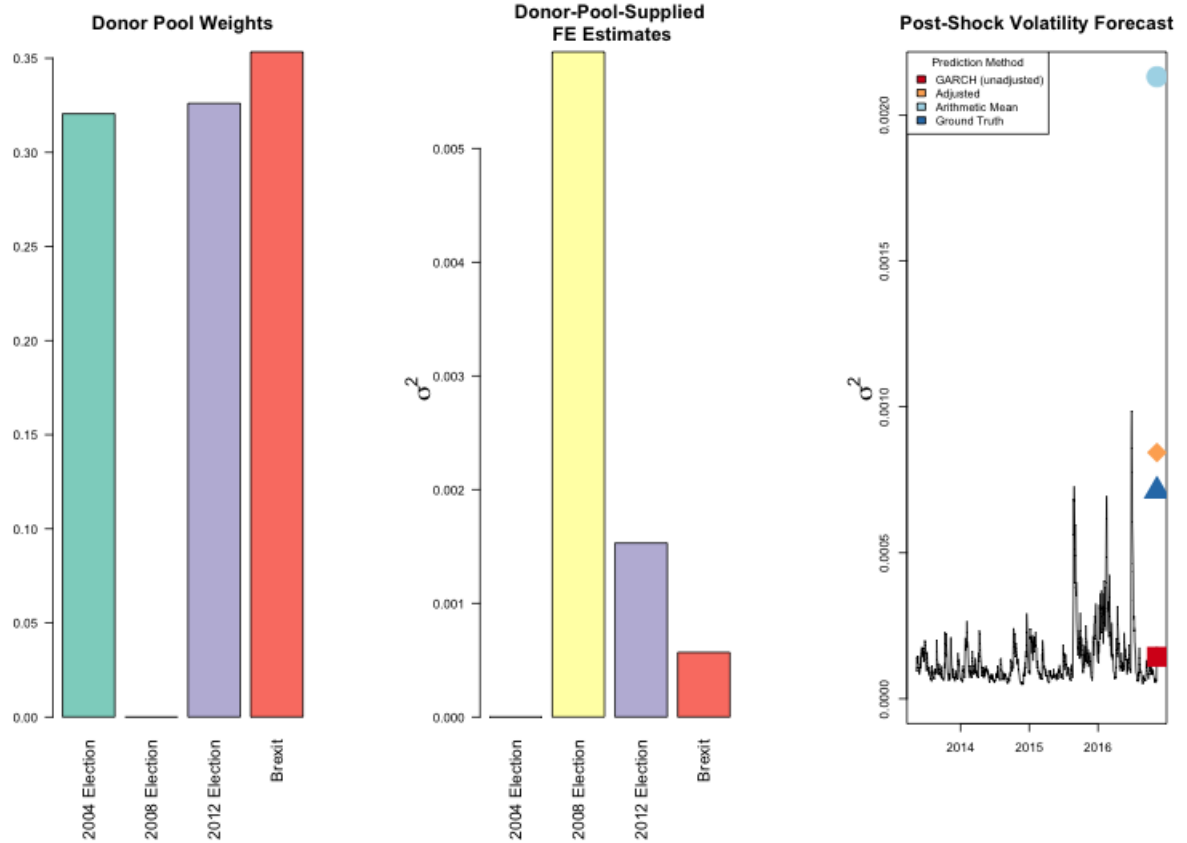


Figure 5: When Brexit is taken to be June 22nd, 2016, it shares equal weight with the 2004 and 2008 elections, while its modest fixed effect estimate is the smallest of the three nonzero estimates.

# Real Data Example: 2016 Election with an Brexit as June 23rd, 2016

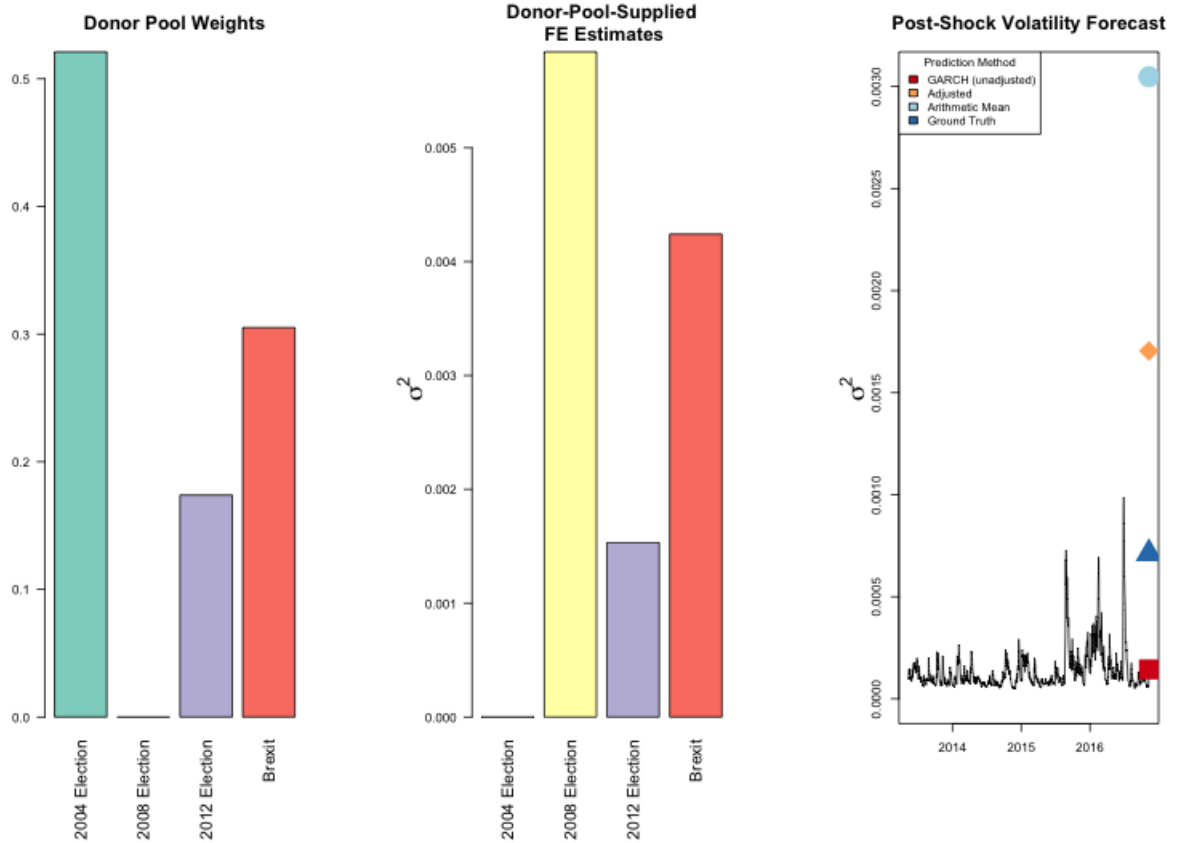


Figure 6: When Brexit is taken to be June 23rd, 2016, its own weight is unchanged but others' weights are. The large fixed effect estimate of Brexit contributes to a substantial overprediction.