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 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.

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.

In this work we focus on the second central moment of a time series, the volatility. One of the most important stochastic phenomenon 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 (Tsay, 2005), easily-interpreted, and often at least weakly stationary. 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. The literature on return series predictability based on news has undergone a considerable shift over the past 50 years (see Boudoukh, Feldman, Kogan, and Richardson (2019) and references therein). Whereas once returns were thought to react to specific news events, now stock price movements are believed to be overwhelmingly noise-based.

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). Problematically, such models 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), which incorporates observable measures of volatility, known as "realized measures", like implied volatility. 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 are likely to be out-of-the-money. GARCH models have been shown slow to adapt to spikes in volatility (Andersen, Bollerslev, Diebold, and Labys, 2003). Engle (2002) explored the use of implied volatility as an exogenous regressor in a GARCH model, that is, using a so-called GARCH-X model (Sucarrat, 2020). Realized GARCH aims to solve both the asymmetry problem as well as the slow-reaction problem by introducing a "measurement equation" as a tool for modeling the contribution to the conditional variance made by the market-implied volatility measure. The key insight is that a realized measure of volatility will often fail to be independent of the log return of the asset. Therefore, including the external measure must be done in a way that accounts for this dependence.

The approach herein can be viewed as an attempt to sidestep the functional complexity posited by Realized GARCH, with its minimum nine parameters to estimate (Sharma et al., 2016), by substituting modeling assumptions. The method herein proceeds under the assumption that similar news events occasion volatility shocks arising from a common 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 (Romer and Romer (1989) cited in Kilian and Lütkepohl (2017); see also discussion of determinstic exogenous events in Engle and Patton (2001)). 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 automated or non-automated, a variety of choices for the collection of data assembled to perform the correction, whether the data is internal (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 an 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 Abadie, Diamond, and Hainmueller (2010), the authors build upon previous work in causal inference whereby a treatment effect can be estimated by creating a synthetic time series that that represents either the treatment or control unit. The synthetic unit is constructed using a convex combination of the so-called donor series. The particular convex combination employed is a function of the distance between the time series under study and the donors. Lin and Eck (2021) adapted these methods for the purpose of prediction. Their 1-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 times series, leaving unaccounted for the variability that accompanies predictions of a heteroskedastic time series. In this present study, we focus on only volatility forecasting. We furthermore depart from Lin and Eck (2021) framework by weighting series not by their covariates (which would be most appropriate for estimating the parameters of time series' mean model) but by their volatility profile.

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. Likewise, we are not familiar with any prior work that attempts to account for anticipated nonzero errors using an explicit parametric adjustment. This current work is a contribution to the forecast adjustment literature but also aims for the virtues of robustness, risk-minimization, and model interpretability, akin to the departures from nearest neighbors discussed in Abadie and L'hour (2021). Guerrón-Quintana and Zhong (2017), citing previous work, discuss that 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. By using convex combinations of donor pool quantities to adjust forecasts, we are generalizing Nearest-Neighbor methods while also providing robustness, risk-minization, and model interpretability.

In order to motivate our procedure, we show how the aggregation of estimated excess volatilities from donors in the donor pool works when the correction function is the arithmetic mean. As we see in Figure 1, the arithmetic mean of excess volatilities provides perhaps the most straightforward and intuitive way of aggregating information from similar events. Taking the average also coincides with the K-Nearest-Neighbor estimator, where K is equal to the number of donors in the donor pool. Although using the arithmetic mean is intuitive and enjoys some empirical track record, its grounding in theory is much less impressive (Timmermann, 2006). 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. Setting

2.1. A Primer on GARCH

We define the log return of an asset between $t-1$ and t as $r_t = \log(\frac{P_{t+1}}{P_t})$, where P_t denotes the price at time t. The class of ARIMA(p, d, q) models (Box, 2013) provides a framework for postulating and quantifying the autoregressive structure of r_t, all within the framework 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

Figure 1: The time series experiences a volatility shock at a uniformly distributed point in the interval [756, 2520] trading days, corresponding to trading years on the interval [3,10]. We have truncated the series after $T^* + 1$. The GARCH prediction fails even to approach the volatility spike at $T^* + 1$, as do some adjusted predictions. In contrast, the GARCH forecast adjusted by $\overline{\widehat{\omega}^*}$, the arithmetic mean of the estimated shocks, recovers the shock in the time series under study.

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 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 (when 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 so-called 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 fitted using a linear combination of the past m values of r_t^2. The GARCH(m, s) framework take this one step further but 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 $\mathbf{x}_t$, we have a GARCH-X(m, s). The volatility equation then becomes

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

2.2. Model setup

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t} = (r_{i,t}, \mathbf{x}_{i,t})$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n + 1$, where $r_{i,t}$ is a scalar response (typically the log return of a financial asset) and $\mathbf{v}_{i,t}$ is a vector of covariates such that $\mathbf{x}_{i,t}|\mathcal{F}_{i,t-1}$ is deterministic. Suppose that the analyst is interested in forecasting the volatility of $r_{1,t}$, the first time series in the collection. 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

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

Let $I(\cdot)$ be an indicator function. Let T_i denote the time length of the time series i for $i = 1, \ldots, n+1$, and let T_i^* denote the largest time index prior to the arrival of the news shock, with $T_i^* < T_i$. Let $\delta, \mathbf{x}_{i,t} \in \mathbb{R}^p$. For $t = 1, \ldots, T_i$ and $i = 1, \ldots, n + 1$, the model $\mathcal{M}_1$ is defined as

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

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

$$\omega_{i,t}^* = D_{i,t}^{vol}[\mu_{\omega^*} + \delta' \mathbf{x}_{i,t} + u_{i,t}],$$

6

with error structure

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

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

$$u_{i,t} \stackrel{iid}{\sim} \mathcal{F}_{u} \text{ with } \mathrm{E}_{\mathcal{F}_{u}}(u) = 0, \mathrm{Var}_{\mathcal{F}_{u}}(u) = \sigma_{u}^{2}$$

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

where $D_{i,t}^{return} = I(t \in \{T_i^* + 1, ..., T_i^* + L_{i,return}\})$ and $D_{i,t}^{vol} = I(t \in \{T_i^* + 1, ..., T_i^* + L_{i,vol}\})$ and $L_{i,return}, L_{i,vol}$ 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 ω_i^* have no dependence on the covariates and are i.i.d. with $\mathbb{E}[\omega_i^*] = \mu_{\omega^*}$.

Consider the connection between the model above and the factor model from Abadie et al. (2010), where an untreated unit is governed by:

$$Y_{i,t}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \varepsilon_{i,t} \ , \tag{2}$$

which happens to nest the GARCH model's volatilty equation (putting aside that σ_t is latent in the GARCH model) as well as as the ARMA representation of a GARCH model, where

$\delta_t \sim \omega$, a location parameter shared across donors

$\boldsymbol{\theta}_t \sim \boldsymbol{\alpha}_k$, a vector of ARCH parameters and other coefficients shared across donors

$\mathbf{Z}_i \sim \boldsymbol{a}_{i,t-k}$, a vector of observable quantities specific to each donor

$\boldsymbol{\lambda}_t \sim \boldsymbol{\beta}_j$, a vector of GARCH parameters shared across donors

$\boldsymbol{\mu}_i \sim \boldsymbol{\sigma}_{i,t-j}^2$, a vector of latent quantities specific to each donor

and $\varepsilon_{i,t}$ is idiosyncratic noise, uncorrelated across time and donors.

2.3. Volatility Profile of a Time Series

In this section, we make a novel contribution to 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, whereas no such restriction exists in (2).

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

$$
\mathbf{V}_t = \begin{pmatrix}
\hat{\alpha}_{1,t} & \hat{\alpha}_{t,2} & \cdots & \hat{\alpha}_{t,n} \\
\hat{\beta}_{1,t} & \hat{\beta}_{t,2} & \cdots & \hat{\beta}_{t,n} \\
\vdots & \vdots & \ddots & \vdots \\
RV_{1,t} & RV_{2,t} & \cdots & RV_{n,t} \\
RV_{1,t-1} & RV_{2,t-1} & \cdots & RV_{n,t-1} \\
\vdots & \vdots & \ddots & \vdots \\
IV_{1,t} & IV_{2,t} & \cdots & IV_{n,t} \\
IV_{1,t-1} & IV_{2,t-1} & \cdots & IV_{n,t-1} \\
\vdots & \vdots & \ddots & \vdots \\
|r_{1,t}| & |r_{2,t}| & \cdots & |r_{n,t}| \\
|r_{1,t-1}| & |r_{2,t-1}| & \cdots & |r_{n,t-1}| \\
\vdots & \vdots & \ddots & \vdots \\
Volume_{1,t} & Volume_{2,t} & \cdots & Volume_{n,t} \\
Volume_{1,t-1} & Volume_{2,t-1} & \cdots & Volume_{n,t-1} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta RV_{1,t} & \Delta RV_{2,t} & \cdots & \Delta RV_{n,t} \\
\Delta RV_{1,t-1} & \Delta RV_{2,t-1} & \cdots & \Delta RV_{n,t-1} \\
\vdots & \vdots & \ddots & \vdots
\end{pmatrix},
$$

where RV and IV denote realized volatility and implied volatility, respectively. 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

We present two forecasts:

Forecast 1: $\hat{\sigma}^2_{unadjusted} = \hat{\mathbb{E}}[\sigma^2_{1,T_1^*+1}|\mathcal{F}_{T^*}] = \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \sigma^2_{i,t-j} + \hat{\gamma}_i^T \mathbf{x}_{i,t}$

Forecast 2: $\hat{\sigma}^2_{adjusted} = \hat{\mathbb{E}}[\sigma^2_{1,T^*_1+1}|\mathcal{F}_{T^*}] + \hat{\omega}^* = \hat{\omega}_i + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k}a^2_{i,t-k} + \sum_{j=1}^{s_i} \hat{\beta}_{i,j}\sigma^2_{i,t-j} + \hat{\gamma}_i^T \mathbf{x}_{i,t} + \hat{\omega}^*$.

A GARCH model is an ARMA on the squares of the observed 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 GARCH model is, just like for an ARMA model, the conditional expectation function, $\mathbb{E}[\sigma^2_{i,T^*+h}|\mathcal{F}_{T^*}]$, or practically speaking, the estimate thereof, $\hat{\mathbb{E}}[\sigma^2_{i,T^*+h}|\mathcal{F}_{T^*}]$ (Zivot, 2009). Here we have presented one-step-ahead forecasts for a GARCH-X(1,1), without loss of generality. For $h = 2, 3, 4, ...$, 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 $\{\pi_i\}_{i=2}^{n+1}$ 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 $\mathbf{x}_{1,T^*}$ using a convex combination of donor covariates. We will say more about the properties of ω_i^* in section 4.

Following Abadie and Gardeazabal (2003); Abadie et al. (2010), 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 (Timmermann, 2006; Wang, Hyndman, Li, and Kang, 2023). 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 $\omega_1 = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i$, 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 satifised.

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 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 aggregating and weighting (not combining) is 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 been 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, where the model family is not in doubt but the parameter values and how they prevail must be learned.

A question naturally arises regarding the uniqueness of weights. Remarks by Lin and Eck (2021); Abadie and Vives-i Bastida (2022) apply here. 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 matrix, and let Conv$(\cdot)$ denote the convex hull of a set of 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, Caratheodory applies (Abadie and Vives-i Bastida, 2022)	Fit not perfect, not necessarily unique, Caratheodory applies

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

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

One immediately visible problem is that this optimization problem is an optmization problem over p-vectors $\vec{w}$ — i.e. a linear combination of the covariates, whereas what we seek is an n-vector — a linear combination of donors.

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 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 $\tilde{r}(t, \frac{1}{m}) = p_t - p_{t-\frac{1}{m}}$. We estimate the variance of ith 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 $\tilde{r}(t,1) \overset{iid}{\sim} N(\mu, \delta^2)$, 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.

3.4. Loss Functions

We are interested in point forecasts for $\sigma_{1,T^*+h}^2 | \mathcal{F}_{T^*}$, $h = 1, 2, ...$, 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, the 1-step-ahead MSE using our method and Realized Volatility is

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

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

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

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

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

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, "[a]mid volatility turmoil, large MSE

11

losses will be a consequence of high volatility without necessarily corresponding to deterioration of forecasting ability. The QL avoids this ambiguity, making it easier to compare losses across volatility regimes." For this reason, we proceed to evaluate the method, both in simuations and real data examples, using the QL loss.

4. Properties of Volatility Shock and Shock Estimators

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.

There is an important question about the stability of the GARCH parameters under the presence of a shock. 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, we do leave over that similarity-based estimators for the GARCH coefficients could be produced, for example, by adapting the methods of Dendramis et al. (2020).

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 ω_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 ϵ_{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, ϵ_{i,T^*+1} and ϵ_{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 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 ω_i^* and the level shock $\epsilon_{i,t}^*$ are of length one only, and consider a circumstance with no exogenous covariate $\mathbf{x}_{i,t}$. Assume also that $r \geq 2$, which is necessary in order to isolate the effects of the level shock $\epsilon_{i,t}^*$. Then

$$\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r} = \omega_i + \alpha_i a^2_{T^*+r} + \beta_i \sigma^2_{i,T^*+r} \tag{3}$$

$$= \omega_i + \alpha_i(\sigma_{i,T^*+r}\epsilon_{T^*+r})^2 + \beta_i\sigma^2_{i,T^*+r}$$

$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha_i(\epsilon_{T^*+r})^2 + \beta_i) \ .$$

In (3), observe that $\omega^*_{i,t}$ and $\epsilon^*_{i,t}$ each appear at most once, through the term $\sigma^2_{T^*+r}$. This might lead one to suspect geometric decay of the shocks $\omega^*_{i,t}$ and ϵ^*_i. Such a suspicion is easier to substantiate by examining the conditional expectation of the variance, $\mathbb{E}[\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r}]$, 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

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

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

$$= \omega_i + \alpha\sigma^2_{i,T^*+r} + \beta\sigma^2_{i,T^*+r} \qquad \text{(Due to the unit variance assumption)}$$

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

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 $r \geq max\{L_{i,vol}, L_{i,level}\} + 1$,*

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

Proof of Proposition   We claim

$$\mathbb{E}[\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r}] = \mathbb{E}[\omega_i + \alpha a^2_{T^*+r} + \beta\sigma^2_{i,T^*+r}|\mathcal{F}_{T^*+r}] \tag{4}$$

$$= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T^*+r}\epsilon_{T^*+r})^2|\mathcal{F}_{T^*+r}] + \beta\sigma^2_{i,T^*+r} \tag{5}$$

$$= \omega_i + \alpha\sigma^2_{i,T^*+r} + \beta\sigma^2_{i,T^*+r} \tag{6}$$

$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha + \beta) \ . \tag{7}$$

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 (4). By the mean-model assumption of a GARCH(1,1),

13

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 (5) above. Using the unit-variance assumption regarding ϵ_{T^*+r}, we can compute explicitly the expectation, yielding (6). Finally, by rearranging terms, we arrive at equation (7). $\qquad\square$

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

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

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 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 ω^*. 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 minimium. Alternatively, one could take the maximum.

4.1. Consistency of the fixed effect estimators

Proposition 2. *Assume*

1. *For each i, $\{a_{i,t}\}_{t=0,\ldots,T_i}$ 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, $\{\omega^*_{i,t}\}_{t=0,\ldots,T_i}$ is potentially non-zero at $\{T^*_i + 1, \ldots, T^*_i + k\}$, $\omega^*_{i,T^*+1} \equiv \ldots \equiv \omega^*_{i,T^*+k}$, and zero otherwise, where the arrival of T^*_i is governed by a time-invariant distribution on $\{a_{i,t}\}_{t=0,\ldots,T_i-1}$.*

3. *The conditions in Assumption 0 of Han and Kristensen (2014) prevail.*

Then for any i, $\hat{\omega}^*_{i,t} \xrightarrow{p} \omega^*_i$, and by Slutsky, consistency is preserved under any convex combination of the estimates due to the i.i.d assumption on $\{\omega_{i,t}\}^{n+1}_{i=1}$.

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

Proof of Lemma   Since the shock is assumed to arrive uniformly at random for each i, $1 \le i \le n+1$, 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. $\qquad\square$

4.2. Consistency of the Forecast Function

Proposition 3. *Assume*

1. *All conditions from the previous proposition.*

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

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

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

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

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

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

which converges in probability to

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

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

$\qquad\square$

4.3. Asymptotic Loss

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

For 1-step-ahead forecast of σ_{t+1}^2, consider the difference

$$QL(\hat{\sigma}_{t+1,unadjusted}^2, \sigma_{t+1}^2) - QL(\hat{\sigma}_{t+1,adjusted}^2, \sigma_{t+1}^2) \tag{8}$$

$$= \left(\frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - 1 \right) - \left(\frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} - 1 \right) \tag{9}$$

$$= \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,unadjusted}^2} - \frac{\sigma_{t+1}^2}{\hat{\sigma}_{t+1,adjusted}^2} + \log \frac{\hat{\sigma}_{t+1,unadjusted}^2}{\hat{\sigma}_{t+1,adjusted}^2} \tag{10}$$

$$= \frac{\sigma_{t+1}^2(\hat{\sigma}_{t+1,adjusted}^2 - \hat{\sigma}_{t+1,unadjusted}^2)}{\hat{\sigma}_{t+1,adjusted}^2 \hat{\sigma}_{t+1,unadjusted}^2} + \log \frac{\hat{\sigma}_{t+1,unadjusted}^2}{\hat{\sigma}_{t+1,adjusted}^2} . \tag{11}$$

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 (11) can be expressed as

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

It is easily verified that as $\hat{\omega}^* \to 0^+$, the difference in the losses goes to zero. On the other hand, as $\hat{\omega}^*$ becomes large, the difference in the losses turns negative, with the lesson being that $\hat{\omega}^*$ must be in appropriate proportion to the volatility σ_{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, $\overline{\omega^*}$, 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 $QL(\hat{\sigma}_{t+1,unadjusted}^2, \sigma_{t+1}^2) - QL(\hat{\sigma}_{t+1,adjusted}^2, \sigma_{t+1}^2) \xrightarrow{p} \frac{\mathbb{E}[\omega_i^*]}{\sigma_{t+1}^2 - \mathbb{E}[\omega_i^*]} + \log \frac{\sigma_{t+1}^2 - \mathbb{E}[\omega_i^*]}{\sigma_{t+1}^2} \geq 0$. *Hence, a correctly specified* $\mathcal{M}_1$ *model will outperform the unadjusted forecast asymptotically.*

Proof of Proposition First, note that the function $g : (-\infty, \mathbb{E}[\omega_i^*]) \to \mathbb{R}$ 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, g preserves consistency. The conclusion follows from fact that the model is correctly specified and consistency is guaranteed Proposition (3).

$\square$

5. Numerical Examples

In this section, we demonstrate the effectiveness of the proposed method using Monte Carlo simulations. The first simulation setup will use a $\mathcal{M}_2$ model on the volatility. The second setup will be a dual shock, i.e. a shock to both the level and volatility.

In order to investigate the forecasting method presented herein, our most elementary simulation uses $\mathcal{M}_1$. We vary only two parameters. Recall an $\mathcal{M}_1$ model on the volatility, which is characterized by an exogenous shock to the volatility equation generated by an affine function of the covariates:

$$\mathcal{M}_1: \quad \begin{aligned} &\sigma_{i,t}^2 = \omega_i + \omega_{i,t}^* + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \\ &a_{i,t} = \sigma_{i,t}((1 - D_{i,t}^{return})\epsilon_{i,t} + D_{i,t}^{return}\epsilon_i^*) \\ &\omega_{i,t}^* = D_{i,t}^{vol}[\mu_{\omega^*} + \delta'\mathbf{x}_{i,t} + u_{i,t}] \\ &D_{i,t}^{return} \equiv 0 \end{aligned}$$

Consider now a shock to both the volatility and the level beginning at $T^* + 1$. In particular, we will explore the performance of the method when a $\mathcal{M}_1$ shock hits the volatility, while an $\mathcal{M}_0$ shock hits the level. There is a good reason we explore an $\mathcal{M}_0$ shock to the level: we have no method to account for an $\mathcal{M}_1$ shock, except for the method of Lin and Eck (2021), which fails to account for heteroskedastic innovations.

Hypothesis: Method should do less well under a concommitant volatility and level shock. A dual shock is in essence a squared observable with two independent sources of randomness.

5.1. Evaluative Framework and Monte Carlo Results

We compare adjusted and unadjusted forecasts using QL Loss, calculating the fraction of the simulations that the adjusted forecast dominates the unadjusted forecast.

In Figure 2, when only two parameters are varied, the volatility shock signal and the volatility shock noise, we observe several encouraging phenomena. First, for any column selected, an increasing trend exists as the shock signal increases. Second, for almost all small values of the shock signal, the outperformance rate hovers around .5, supporting the hypothesis that in the absence of a signal, any level of noise renders the method no better at GARCH than a flip of a coin.

If we switch the values of α and β, we see similar behavior, as in Figure 3. Here we cite Francq and Zakoian (2019, p. 22) in motivating simulations for large α values and then large β values. However, for small values of the shock mean, increasing noise does lead to fewer converged simulations, likely due to large negative realizations of the noise term, which in turn lead to near-zero and even negative estimates of the terms ω_i^*.
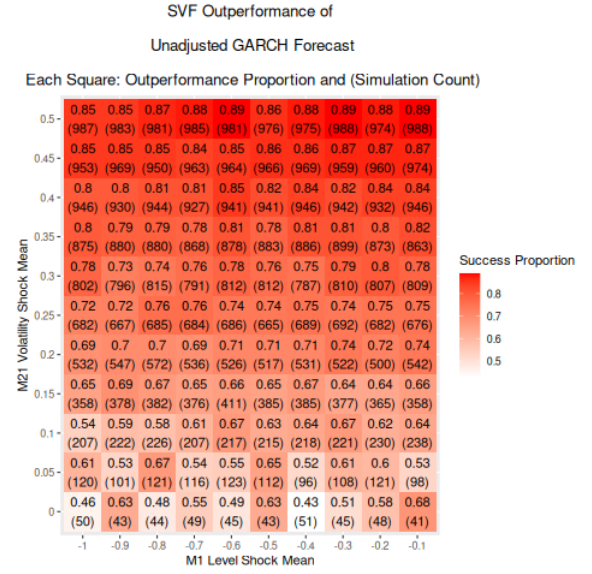
Figure 2: Fixed parameter values: $\alpha = .1, \beta = .82, \mu_x = 1, \sigma_x = .1$

Figure 3: Fixed parameter values: $\alpha = .82, \beta = .1, \mu_x = 1, \sigma_x = .1$
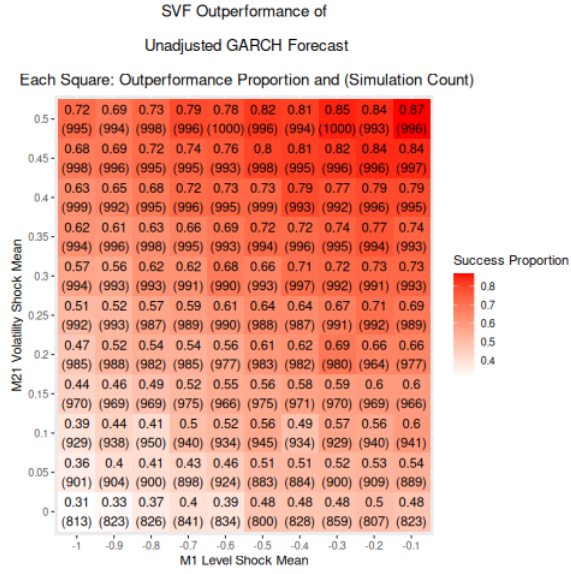
(a) $\alpha = .1, \beta = .82, \mu_x = 1, \sigma_x = .1$ (b) $\alpha = .82, \beta = .1, \mu_x = 1, \sigma_x = .1$

Figure 4: Shocks to both the mean function and the volatility equation

19

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 Mayhew (2008). 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" (News, 2024). Gleaning signals from financial options markets and betting markets, Wolfers and Zitzewitz (2016) 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. Goodell and Vähämaa (2013) 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 (Silver, 2016), suggesting that Trump's victory was at least somewhat surprising.

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[1] (an ETF composed

---

[1]It has been noted that GARCH effects are more attenuated in aggregated returns (Zivot, 2009), which suggests

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 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 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 (Goodell and Vähämaa, 2013; Kane, Marcus, and Noh, 1996). 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** Synthetic Control, as a tool of causal inference, often goes about weighting control units by first identifying a natural set of donors or standard donors such as the untreated units within a set of subnational units like US states or Spanish provinces (Abadie and Gardeazabal, 2003; Abadie et al., 2010). While such a procedure does not necessarily preclude considered judgments (e.g. should Canadian provinces be used as donors for a treated US state?), as a tool of small-n prediction, distanced-based weighting may favor somewhat different considerations in constructing a donor pool. For our purposes, 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.

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 (Andersen

---

against using the S&P 500 or similar indices as an example.

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.
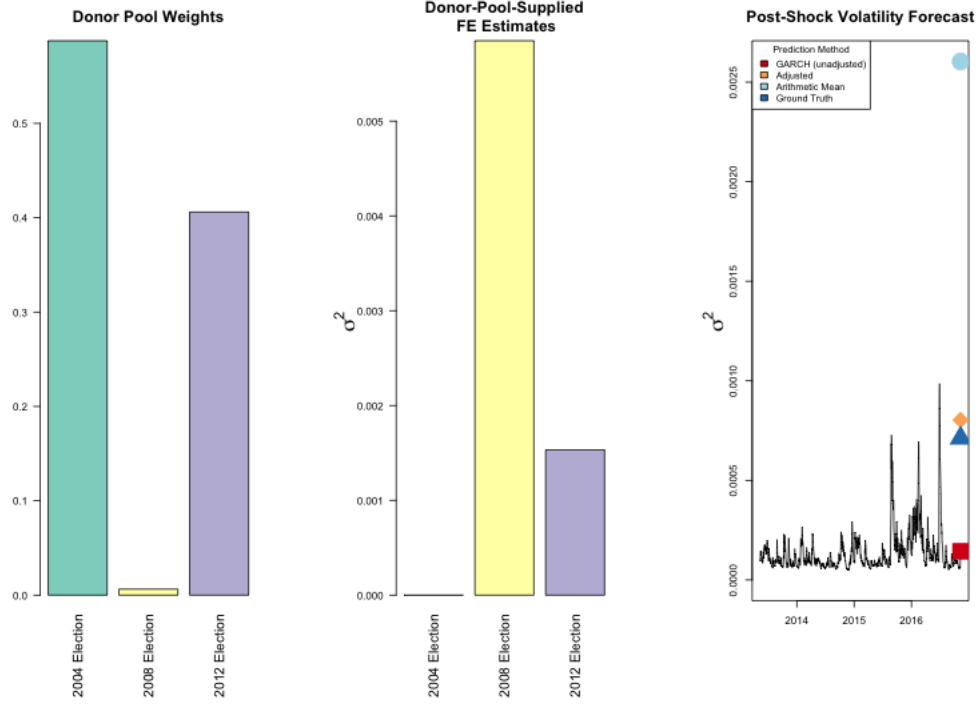

Figure 5: The volatility induced by the 2016 US election

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 neglible 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 points, all indicated by 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 the ground truth. The prediction is not only directionally correct, i.e. we predict a volatility spike where there is one; the prediction

22

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 conditons 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 the cumulative variation, indicating a moderate concentration or redundancy of information in the three donors. Second, we follow Steegen, Tuerlinckx, Gelman, and Vanpaemel (2016) in a 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. Additionally, in the supplement, we show that with Brexit added as a donor, the results are unchanged, due to the fact that Brexit's covariates do not lie in the convex hull of the 2004 Election, the 2012 Election, and the time series under study.

7. Discussion

This present work extends Lin and Eck (2021) principally by substituting the GARCH model for an AR(1), i.e. modeling both the mean and volatility of a univariate time series. We permit both shocks in the mean model and volatility model of GARCH. As ancillary improvements, the current work permits the shock periods to be of any length, formally $L_{\text{vol}}, L_{\text{return}} \in \{0, 1, 2, ...\}$, and the weights applied to the donor pool can come from more exotic subsets of $\mathbb{R}^n$. At a more technical level, the most substantial insight related to Lin and Eck (2021) is that the GARCH model provides an elegant forecast function for squared observations of the time series under study. However, this use-case becomes less attractive in situations where the sign of the time series under study following the shock is uncertain.

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? For how long is the forecast reliable? Should we take a convex combination of the donor pool shock lengths? Or mean? Or minimum of the donor pool shock lengths? unlike KNN in that 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 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 Lin and Eck (2021) and KNN. In KNN, the exogenous variables are taken as a given and 'nearness' to object-to-predict 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.

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.

7.1. Future directions

- High-frequency data

- Should we fit n models or only 1 or something in between?

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, 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

As mentioned in **??**

8.3. Sensitivity to Donors Chosen

We note that dropping the 2008 Election vastly improves the performance of the arithmetic mean forecast.

8.3.1. Brexit

We analyze the real-data example with Brexit included and using the daily return in the Pound-Dollar exchange rate as a covariate.

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

Note that by inspecting the fit of the distance-based weighting approximation quantitatively, we can gain some degree of foresight into how well the forecast will be.

- Table of predictions

- Table of losses

- How often does average or median forecast dominate?

Donor Pool Weights

Donor-Pool-Supplied FE Estimates

Post-Shock Volatility Forecast
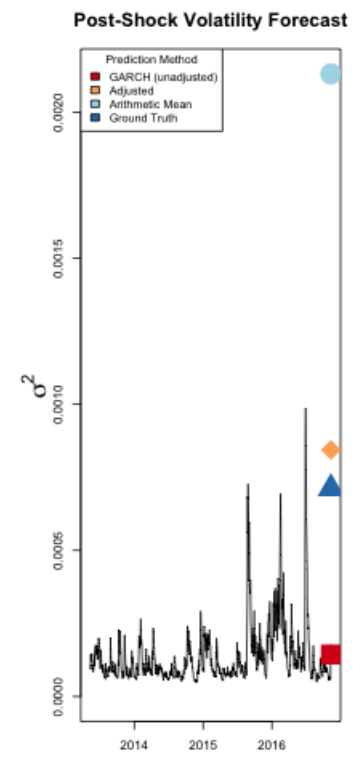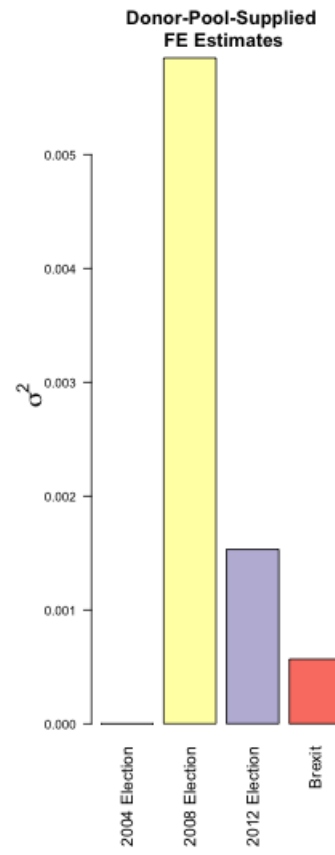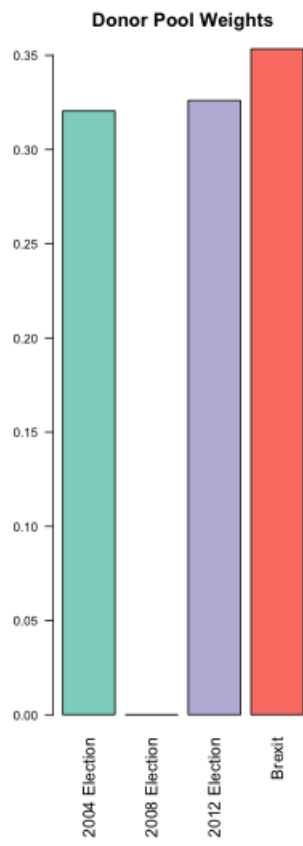
QLLoss.AdjustedForecast.	dropped_covariate	dropped_donor
0.0003	X.VIX_log_ret	2016-06-22
0.0013	X.IRX_log_ret	2008-11-04
0.0013	X.TYX_log_ret	none
0.0013	X.TYX_log_ret	2016-06-22
0.0017	X.TYX_log_ret	2008-11-04
0.0019	X.TNX_log_ret	2016-06-22
0.0022	X.TNX_log_ret	2008-11-04
0.0026	X.FVX_log_ret	2008-11-04
0.0031	X.VIX_raw	2008-11-04
0.0033	X.TNX_log_ret	none
0.0035	CL.F_log_ret	2016-06-22
0.0037	X.VIX_log_ret	2008-11-04
0.0041	Debt_Risk_Spread	2016-06-22
0.0043	none	2008-11-04
0.0047	X.FVX_log_ret	2016-06-22
0.0053	X.VIX_raw	2016-06-22
0.0060	CL.F_log_ret	2008-11-04
0.0070	none	2016-06-22
0.0075	X.FVX_log_ret	none
0.0078	X.IRX_log_ret	2016-06-22
0.0102	Debt_Risk_Spread	2008-11-04
0.0102	X.IRX_log_ret	none
0.0112	X.VIX_log_ret	none
0.0135	X.VIX_raw	none
0.0136	IYG	2008-11-04
0.0137	none	none
0.0138	CL.F_log_ret	none
0.0154	Debt_Risk_Spread	none
0.0269	IYG	2016-06-22
0.0568	X.TNX_log_ret	2012-11-06
0.0670	IYG	none
0.0746	X.VIX_log_ret	2012-11-06
0.0800	X.VIX_log_ret	2004-11-02
0.0806	X.TNX_log_ret	2004-11-02
0.0807	X.TYX_log_ret	2004-11-02
0.0851	Debt_Risk_Spread	2012-11-06
0.0866	none	2012-11-06
0.0868	X.VIX_raw	2012-11-06
0.0901	CL.F_log_ret	2012-11-06
0.0901	X.IRX_log_ret	2004-11-02
0.0906	X.IRX_log_ret	2012-11-06
0.0912	X.VIX_raw	2004-11-02
0.0913	none	2004-11-02
0.0916	X.FVX_log_ret	2004-11-02
0.0920	Debt_Risk_Spread	2004-11-02
0.0935	CL.F_log_ret	2004-11-02
0.0995	X.TYX_log_ret	2012-11-06
0.1143	X.FVX_log_ret	2012-11-06
0.2144	IYG	2004-11-02
2.3530	IYG	2012-11-06

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 Jérémy L'hour. A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834, 2021.

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.

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.

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 Engle. New frontiers for arch models. *Journal of Applied Econometrics*, 17(5):425–446, 2002.

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.

John W Goodell and Sami Vähämaa. Us presidential elections and implied volatility: The role of political uncertainty. *Journal of Banking & Finance*, 37(3):1108–1117, 2013.

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.

Alex Kane, Alan J Marcus, and Jaesun Noh. The p/e multiple and market volatility. *Financial Analysts Journal*, 52 (4):16–24, 1996.

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.

David R Mayhew. Incumbency advantage in us presidential elections: The historical record. *Political Science Quarterly*, 123(2):201–228, 2008.

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

Bloomberg News. Election 2024: What will markets do with trump victory over biden? https://www.bloomberg.com/news/articles/2024-01-21/trump-s-2016-win-shook-markets-traders-won-t-get-fooled-again, Jan 2024.

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.

Prateek Sharma et al. Forecasting stock market volatility using realized garch model: International evidence. *The Quarterly Review of Economics and Finance*, 59:222–230, 2016.

Nate Silver. 2016 election forecast, Nov 2016. URL https://projects.fivethirtyeight.com/2016-election-forecast/.

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.

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.

Justin Wolfers and Eric Zitzewitz. What do financial markets think of the 2016 election. *Unpublished manuscript. brook. gs/2njI8U3*, 2016.

Eric Zivot. Practical issues in the analysis of univariate garch models. In *Handbook of financial time series*, pages 113–155. Springer, 2009.