Post-shock Volatility Forecasting Using Aggregated Shock Information

David Lundquist[*], Daniel Eck[†]

Department of Statistics, University of Illinois at Urbana-Champaign

January 26, 2024

Abstract

We develop a procedure for forecasting the volatility of a time series immediately following an exogenous shock. Adapting the synthetic prediction 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 three families of benchmarks: the unobservable but easily-estimated realized volatility (RV), implied volatility, and the empirical volatility over horizons of varying length. We also compare the performance of the adjusted forecast to the performance of the Realized GARCH forecast, which is known to react faster to rapidly-changing volatility than GARCH incorporating implied volatility. Finally, we combine Realized GARCH modeling with the synthetic prediction framework, using Realized GARCH in both the estimation of random effects as well as the forecast for the time series under study. Real-world illustrations are provided, as are simulation results suggesting the conditions under which our approach's hyperparameters can be tuned for best performance.

1 Introduction

Reacting to a seemingly unprecedented event might prompt the question: what, if anything, does it resemble from the past? Such could 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, performance and robustness, among many others.

In this work we focus on the second central moment of a time series, the volatility. The most important stochastic phenomenon of many 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. Yet 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. Modern portfolio management and theory often requires information about at least the first two moments of a return series, if not higher. Volatility modeling has grown immensely over the previous four decades to meet the needs of both theoreticians and practitioners (**?**), prompting lines of inquiry to search for novel settings and challenges, as we do in this paper. No matter

[*]davidl11@ilinois.edu
[†]dje13@illinois.edu

how a time series or its transformations are modeled, forecasting in the presence of exogenous shock events 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 larger volatility than positive returns (**?**). Problematically, such models depend upon the observation of a negative return to provide the most updated volatility forecast, and under the circumstances posited herein, no such return has been observed.

GARCH models have been shown slow to adapt to spikes in volatility (**?**). **?** explored the use of implied volatility as an exogenous regressor in a GARCH model, that is, using a so-called GARCH-X model (**?**). **?** propose Realized GARCH, which 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 that the market-implied volatility measure is not independent of the conditional variance posited by the GARCH model. 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 (**?**), by substituting modeling assumptions. Synthetic Volatility Forecasting proceeds under the assumption that similar unscheduled news events occasion volatility shocks arising from a common shock distribution.

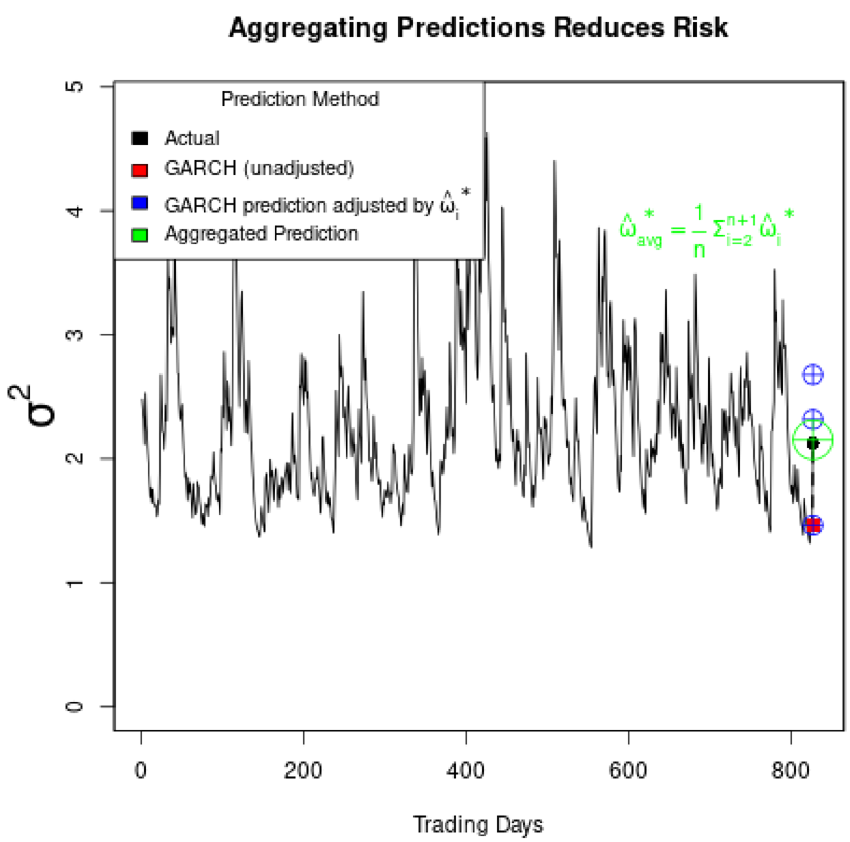

Figure 1: Shock aggregation using the arithmetic mean of donor fixed effect estimates.

As we see in Figure **??**, the arithmetic mean of estimated shock effects provides perhaps the most straightforward and intuitive way of aggregating information from similar events. Aggregating forecasts using the arithmetic mean is more than merely intuitive; it possesses an empirical pedigree and is known to be optimal under rather restrictive conditions (**?**). Synthetic Volatility Forecasting lacks these restrictions and instead assumes a credible, parsimonious parameterization in which the shock is an affine transformation of several key covariates.

2 Setting

We will suppose that a researcher has multivariate time series data $\mathbf{r}_{i,t}, t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. Let $\mathbf{r}_{i,t} = (r_{i,t},\ \mathbf{x}_{i,t},\ \mathbf{v}_{i,t})$ where $r_{i,t}$ is a scalar response, $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to the analyst prior to the observation of $ret_{1,t}$, and $\mathbf{v}_{i,t}$ is a vector of market-implied volatility metrics for the series i that are available prior to the market open at time t. Suppose that the analyst is interested in forecasting the volatility of $y_{1,t}$, the first time series in the collection. We require that each time series $\mathbf{r}_{i,t}$ is subject to an unscheduled 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) outside market hours. In contrast, the arrival of market-moving news does not obey any such restrictions.

We follow convention and define the daily log-return 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 developed in the mid-to-late 20th century (**?**) 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 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 constant-variance 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. Implicitly, we are committing ourselves to a mean-model for r_t in which μ_t, the expected daily log-return, may vary with time, while a_t is simply the noise. As **?** explains, such an assumption is justified by the empirical finding that returns lack significant autocorrelation. We thus derive a mean-zero process $(a_t)_{t \in \mathbb{N}}$ with the property that $\mathbb{E}[a_t^2] = \mathrm{Var}[a_t]$. 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 m past values of a_t^2. The GARCH(m,s) framework take this one step further but modeling σ_t^2 as a linear combination of the m past a_t^2 values and well as the s past values of σ_t^2. In functional form,

$$
\begin{aligned}
\sigma_t^2 &= \omega + \sum_{k=1}^{m} \alpha_k a_{t-k}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \\
a_t &= \sigma_t \epsilon_t
\end{aligned}
$$

Assuming further that σ_t^2 depends on a vector of exogenous covariates $\mathbf{x}_t$ (a so-called "GARCH-X"), we have

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

$$a_t = \sigma_t \epsilon_t$$

In light of the foregoing, we can rewrite the GARCH framework we are interested in as such

$$\sigma^2_{i,t} = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \begin{pmatrix} \mathbf{x}_{i,t} \\ \mathbf{v}_{i,t} \end{pmatrix}$$
$$a_{i,t} = \sigma_{i,t}\epsilon_{i,t}$$

2.1 Model setup

In this section, we will describe the assumed dynamic panel models for which post-shock aggregated estimators are provided. The basic structures of these models are the same for all time series in the analysis, the differences between them lie in the setup of the shock effect distribution. We first sharply distinguish between a volatility shock induced by a return series shock and a volatility shock directly affecting the volatility equation of a GARCH model, without mediation of the return series.

2.1.1 Mean Model and Volatility Model

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

$$\mathcal{M}_1\colon \sigma^2_{i,t} = \omega_i + \omega_i^* D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{1}$$
$$a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1 - D^{level}_{i,t}) + \epsilon_i^* D^{level}_{i,t}) \tag{2}$$
$$\omega_i^* = \mu_{\omega^*} + \varepsilon_i \tag{3}$$

where $D^{vol}_{i,t} = I(t \in \{T_i^*+1, ..., T_i^* + L_{i,vol}\})$, $D^{level}_{i,t} = I(t \in \{T_i^*+1, ..., T_i^* + L_{i,level}\})$ and $\mathbf{x}_{i,t} \in^p$, with $p \geq 1$. We assume that the $\mathbf{x}_{i,t}$ are fixed [1].[2][3] For $i = 1, \ldots, n+1$ and $t = 1, \ldots, T_i$, the random effects structure for $\mathcal{M}_1$ is:

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

Notice that $\mathcal{M}_1$ assumes that ω_i^* are i.i.d. with $\mathbb{E}[\omega_i^*] = \mu_{\omega^*}$ for $i = 1, \ldots, n+1$. We also consider a model where the shock effects are linear functions of covariates with an additional additive mean-zero error. For $i = 1, \ldots, n+1$, the random effects structure for this model (model $\mathcal{M}_2$) is:

[1] Is this necessary? Does it serve a purpose?

[2] Need to determine what difference it makes to a GARCH-X model, if any, for the covariates to be random.

[3] As for implied volatility as a covariate, the argument for regarding it as a non-random quantity is that we are not primarily concerned with what IV is trying to estimate. We're concerned with IV itself at a primitive quantity that affects volatility through market sentiment.

$$
\mathcal{M}_2\colon \begin{array}{l} \sigma^2_{i,t} = \omega_i + \omega^*_i D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j}\sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \\ a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1-D^{level}_{i,t}) + \epsilon^*_i D^{level}_{i,t}) \\ \omega^*_i = \mu_{\omega^*} + \delta' \mathbf{x}_{i,T^*_i} + \varepsilon_i, \end{array}
$$

with random effects structure

$$
\begin{aligned}
\omega^*_i &\overset{iid}{\sim} \mathcal{F}_{\omega^*} \text{ with } \mathrm{E}_{\mathcal{F}_{\omega^*}}(\omega^*) = \mu_{\omega^*}, \mathrm{Var}_{\mathcal{F}_{\omega^*_i}}(\omega^*_i) = \sigma^2_{\omega^*} \\
\epsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\epsilon} \text{ with } \mathrm{E}_{\mathcal{F}_{\epsilon}}(\epsilon) = 0, \mathrm{Var}_{\mathcal{F}_{\epsilon}}(\epsilon) = \sigma^2_{\epsilon} \\
\epsilon^*_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\epsilon^*} \text{ with } \mathrm{E}_{\mathcal{F}_{\epsilon^*}}(\epsilon^*_{i,t}) = \mu_{\epsilon^*}, \mathrm{Var}_{\mathcal{F}_{\epsilon^*}}(\epsilon^*_i) = \sigma^2_{\epsilon^*} \\
\delta &\overset{iid}{\sim} \mathcal{F}_{\delta} \text{ with } \mathrm{E}_{\mathcal{F}_{\delta}}(\delta) = \mu_{\delta}, \mathrm{Var}_{\mathcal{F}_{\delta}}(\delta_i) = \Sigma_{\delta} \\
\varepsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_{\varepsilon} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon}}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon}}(\varepsilon_{i,t}) = \sigma^2_{\varepsilon} \\
\omega^*_i &\perp\!\!\!\perp \epsilon_{i,t} \perp\!\!\!\perp \epsilon^*_{i,t} \perp\!\!\!\perp \delta \perp\!\!\!\perp \varepsilon_{i,t}
\end{aligned}
$$

We further define the parameter sets

$$
\begin{array}{lcl} \Theta & = & \{(\omega^*_i, \epsilon_{i,t}, \epsilon^*_{i,t}, \delta, \varepsilon_{i,t})\colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\}, \\ \Theta_1 & = & \{(\omega^*_i, \epsilon_{i,t}, \epsilon^*_{i,t}, \varepsilon_{i,t})\colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\}, \end{array} \tag{4}
$$

where Θ and Θ_1 can be adapted to $\mathcal{M}_1$ by dropping δ. We assume this for notational simplicity.

2.2 Volatility Profile of a Time Series

In this section we make a novel contribution to the synthetic prediction framework by constructing a profile of a time series' volatility. Suppose we have D donors and for each of those D donors, q distinct measurements and/or covariates of volatility. Abusing notation for RV (I need some clever way to index (a) date, (b) donor, (c) how many days were used to get the measurement, we have...

$$
\mathbf{V}_{q,D} = \begin{pmatrix}
\alpha_{T^*,1} & \alpha_{T^*,2} & \cdots & \alpha_{T^*,D} \\
\beta_{T^*,1} & \beta_{T^*,2} & \cdots & \beta_{T^*,D} \\
\vdots & \vdots & \ddots & \vdots \\
RV_{T^*,1} & RV_{T^*,2} & \cdots & RV_{T^*,D} \\
RV_{T^*-1,1} & RV_{T^*-1,2} & \cdots & RV_{T^*-1,D} \\
\vdots & \vdots & \ddots & \vdots \\
IV_{T^*,1} & IV_{T^*,2} & \cdots & IV_{T^*,D} \\
IV_{T^*-1,1} & IV_{T^*-1,2} & \cdots & IV_{T^*-1,D} \\
\vdots & \vdots & \ddots & \vdots \\
AbsoluteReturn_{T^*,1} & AbsoluteReturn_{T^*,2} & \cdots & AbsoluteReturn_{T^*,D} \\
AbsoluteReturn_{T^*-1,1} & AbsoluteReturn_{T^*-1,2} & \cdots & AbsoluteReturn_{T^*-1,D} \\
\vdots & \vdots & \ddots & \vdots \\
Volume_{T^*,1} & Volume_{T^*,2} & \cdots & Volume_{T^*,D} \\
Volume_{T^*-1,1} & Volume_{T^*-1,2} & \cdots & Volume_{T^*-1,D} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta RV_{T^*,1} & \Delta RV_{T^*,2} & \cdots & \Delta RV_{T^*,D} \\
\Delta RV_{T^*-1,1} & \Delta RV_{T^*-1,2} & \cdots & \Delta RV_{T^*-1,D} \\
\vdots & \vdots & \ddots & \vdots
\end{pmatrix}
$$

3 Post-shock Synthetic Volatility Forecasting Methodology

3.1 Ground Truth Estimators

The price series of a financial asset is a sequence of observable random variables; in particular, for any t, P_t is realized at t and henceforth no longer random, under the assumption that $(P_t)_{t\in\mathbb{N}}$ is adapted to $\mathcal{F}_t$. Time series econometrics has a vast inventory of approaches for modeling P_t. In contrast, 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. Approaches come in two basic forms: model-independent and model-dependent. The naive approach of constructing a rolling unbiased estimator of the variance of a price series is analogous to using a rolling average to estimate the mean of a price series. Implicitly, these approaches assume a simple, if unrealistic, covariance structure. Additionally, they offer very little in the way of forward guidance (predictiveness) regarding the series at hand. For that reason, in this section, we propose three families of estimators of the ground truth that we aim to forecast.

3.1.1 Market-Implied Volatility

Here we introduce implied volatility, a quantity that is derived using the equilibrium price of call and put options on the underlying financial asset.

We note also that Black-Scholes implied volatility is a biased estimator of volatility (**??**), with the bias increasing in times of crises when options are out-of-the-money.

3.1.2 Realized Volatility

Suppose we examine K units of of time, where each unit is divided into m intervals of length 1/m. We follow the notation of **?**. Let $p_t = \log(P_t)$, and let $r(t, 1/m) = p_t - p_{t-1/m}$. We estimate the variance of a log-return series using Realized Volatility, denoted $RV_{K,m}$, using

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

Assuming that the K units $r(t, 1) = p_t - p_{t-1}$ are such that $r(t, 1) \overset{iid}{\sim} N(\mu, \delta^2)$, 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.

3.1.3 Historical Volatility

As discussed in section **??**, daily returns exhibit insigificant autocorrelation. Such a finding could be used to motivate using the empirical volatility. This is just the unbiased estimator of log-return σ^2 over the preceding M days.

3.2 Aggregation Mechanism

Here we explain how we use the volatility profile to arrive at a set of nonnegative weights that sum to 1. These weights are then used to compute $\hat{\omega}^* = \sum_{i=2}^{n+1} w_i\hat{\omega}_i$. Since the $\{w_i\}_{i=2}^{n+1}$ are computed using information from $\mathcal{F}_{T_i^*}$, the set $\{w_i\}_{i=2}^{n+1}$ is deterministic.

In **?**, the authors advance 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 so-called donor series. 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 1-step-ahead forecasts take inspiration from distance-based-weighting, pooling shock estimates from similar series according to the series' similarity to the series under study. The use of fixed-effect estimation for structural shocks has a pedigree (**?** cited in **?**). Their approach does not take into account the ARCH effects commonly observed in time series, especially financial times series, leaving unaccounted the variability that accompanies predictions of a heteroskedastic time series. In this present study, we focus on only volatility forecasting. We furthermore depart from the synthetic prediction 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.

3.3 Loss Functions

We are interested in a point forecast for σ^2_{1,T^*+h}, $h = 1, 2, ..., H$, the h-step ahead conditional variance for the time series under study, up to a forecast length of H. The forecast performance evaluation uses three distinct loss functions, each computed using three families of estimators for the ground truth that we seek. 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, one loss function of interest in this study is the 1-step-ahead MSE using Synthetic Volatility Forecasting and Realized Volatility:

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

In more generality, for an multihorizon volatility forecast with forecast length H, the loss function is

$$MSE^H_{method,groundtruth} = \frac{1}{H}\sum_{h=1}^{H}(\hat{\sigma}^2_{h,method} - \hat{\sigma}^2_{h,groundtruth})^2$$

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

$$\text{MAPE}^H_{method,groundtruth} = \frac{1}{H}\sum_{h=1}^{H}\frac{|\hat{\sigma}^2_{h,method} - \hat{\sigma}^2_{h,groundtruth}|}{\hat{\sigma}^2_{h,groundtruth}}$$

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

$$\text{QL}^H_{method,groundtruth} = \frac{1}{h}\sum_{h=1}^{H}(\frac{\hat{\sigma}^2_{h,method}}{\hat{\sigma}^2_{h,groundtruth}} - \log\frac{\hat{\sigma}^2_{h,method}}{\hat{\sigma}^2_{h,groundtruth}} - 1)$$

What distinguishes QL Loss is that it is multiplicative rather than additive. This serves as a basis for some of its virtues, both practical and theoretical. As **?** explains, "[a]mid volatility turmoil, large MSE 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."

4 Properties of Volatility Shock 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 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 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 $\sigma^2_{i,t}$ that should be familiar to scholars of GARCH 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 regressor $\mathbf{x}_{i,t}$. Consider also the case where $r \geq 2$, which is necessary in order to isolate the effects of the level shock $\epsilon^*_{i,t}$. Then

$$\sigma^2_{i,T^*+r+1} = \omega_i + \alpha_i a^2_{T^*+r} + \beta_i \sigma^2_{i,T^*+r} \tag{5}$$
$$= \omega_i + \alpha_i(\sigma_{i,T^*+r}\epsilon_{T^*+r})^2 + \beta_i \sigma^2_{i,T^*+r} \tag{6}$$
$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha_i(\epsilon_{T^*+r})^2 + \beta_i) \tag{7}$$

In (**??**), observe that ω^*_i 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 ω^*_i 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 (**?**). Indeed, if we assume unit variance for all $\epsilon_{i,t}$ except, of course, $\epsilon^*_{i,t}$, then we have

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

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 $L_{level} \geq 1$ at some time T^*+1. Then for any r such that $r \geq max\{L_{i,vol}, L_{i,level}\} + 1$,*

$$\mathbb{E}[\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r}] = \omega_i + \sigma^2_{i,T^*+r}(\alpha+\beta) \tag{8}$$
$$\tag{9}$$

Proof of Proposition We claim

$$\mathbb{E}[\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r}] = \mathbb{E}[\omega_i + \alpha a^2_{T^*+r} + \beta\sigma^2_{i,T^*+r}|\mathcal{F}_{T^*+r}] \tag{10}$$
$$= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T^*+r}\epsilon_{T^*+r})^2|\mathcal{F}_{T^*+r}] + \beta\sigma^2_{i,T^*+r} \tag{11}$$
$$= \omega_i + \alpha\sigma^2_{i,T^*+r} + \beta\sigma^2_{i,T^*+r} \tag{12}$$
$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha+\beta) \tag{13}$$

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 (**??**). 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 (**??**) above. Using the unit-variance assumption regarding ϵ_{T^*+r}), we can compute explicitly the expectation in (**??**). Finally, by rearranging terms, we arrive at equation (**??**). □

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 one to see that

$$\begin{aligned}
\mathbb{E}[\sigma^2_{i,T^*+r+1}|\mathcal{F}_{T^*+r}] &= \mathbb{E}[\omega_i + \alpha a^2_{T^*+1} + \beta\sigma^2_{i,T^*+1}|\mathcal{F}_{T^*+1}] && (14)\\
&= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T^*+r}\epsilon^*_{T^*+1})^2|\mathcal{F}_{T^*+1}] + \beta\sigma^2_{i,T^*+1} && (15)\\
&= \omega_i + \alpha\sigma^2_{i,T^*+1}(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta\sigma^2_{i,T^*+1} && (16)\\
&= \omega_i + \sigma^2_{i,T^*+1}(\alpha(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta) && (17)
\end{aligned}$$

For each of the p entries in δ, the kth entry should have the property that
$\mathbb{E}[x_{i,t,k} \cdot \delta_k] > 0$
In principle, you could have an $\mathcal{M}_{21}$ shock with a covariate that, in expectation, has a negative marginal contribution to the cumulative shock.

Moreover, after both shocks have been exhausted, their influence disappears quickly. This short-memory effect has implications for the method being developed herein:

1. 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^*.

2. An operator of the method needs some idea of how long the operator expects the shock in the time series under study to be. Such an idea will guide her trust in the choice of k in the k-step ahead forecast the operator produces. 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.

3. There may be different risks associated with over/underestimating level shock and vol shock lengths.

In many settings, it is reasonable to model a volatility shock as occuring without a rupture in the mean-zero, i.i.d. sequence $\epsilon_{i,t}$. In cases like this, outsize movement in the observable random variable $a_{i,t}$ is due solely to the shock ω^*_i. Using the ARMA representation of a GARCH(m,s) model, we can see clearly how the random effect ω^*_i increases the expectation of the nonnegative random variable $a^2_{i,t}$. The unconditional mean of an ARMA model tells that the random effects simply contribute to the intercept term (see Tsay p. 132):

$$\mathbb{E}[a^2_{i,t}] = \frac{\omega_i + \mathbb{E}[\omega^*_i]}{1 - \sum_{k=1}^{\max(\mathrm{m,s})}(\alpha_{i,k} + \beta_{i,k})}$$

However, since it is not known a priori for which t the effect ω^*_i will be nonzero, this fact is of little practical guidance.

4.1 Consistency of the Synthetic Volatility fixed-effect estimators

Proposition 2. *Assume*

1. *$\forall i$, $\{a_{i,t}\}_{t=0,...,T_i}$ obeys a GARCH-X(m,s) process with volatility shocks found in either $\mathcal{M}_1$ or $\mathcal{M}_2$, where T_i is the length of the ith series.*

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

3. *Let the conditions in Assumption 0 of **?** prevail.*

Then $\forall i, \hat{\omega}^*_{i,t} \xrightarrow{p} \mathbb{E}[\omega^*_i]$.

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

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

4.2 Consistency of the Synthetic Volatility prediction function

Proposition 3. *Under the conditions in Proposition **??**, $\hat{\sigma}^2_{1,T^*+r} \xrightarrow{p} \sigma^2_{1,T^*+r}, r=1, k=max\{m,s\}$.*

Proof of Proposition Recall the conditional expectation of the variance for the GARCH-X model:

$$\mathbb{E}_{T^*}[\sigma^2_{i,t+1}|\mathcal{F}_{T^*}] = \omega_i + \omega_i^* D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j}\sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{18}$$

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

$$\hat{\sigma}^2_{i,t+1}|\mathcal{F}_{T^*} = \hat{\omega}_i + \hat{\omega}_i^* D^{vol}_{i,t} + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \hat{\beta}_{i,j}\hat{\sigma}^2_{i,t-j} + \hat{\gamma}_i^T \mathbf{x}_{i,t} \tag{19}$$

which converges in probability to

$$\hat{\sigma}^2_{i,t+1}|\mathcal{F}_{T^*} = \omega_i + \omega_i^* D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j}\sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{20}$$

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

□

4.3 Forecasting

Here we make reference to Tsay p. 133 and explain, very similarly to Tsay, that the prediction function for a GARCH(m,s) is very similar to an ARMA prediction function. Modeled after section 2.2 here https://arxiv.org/pdf/2008.11756.pdf

OR we could cite **?**

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

5 Numerical Examples

Having introduced the model parameters above, we introduce estimation techniques that can be varied

5.1 Modeling Setup

5.2 Peformance Metrics

5.3 Most elementary simulation setup

5.4 Monte Carlo results

Recall an $\mathcal{M}_2$ model on the volatility, which is characterized by an exogenous shock to the volatility equation generated by an affine function of the covariates:

$$\mathcal{M}_2\colon \begin{array}{l} \sigma_{i,t}^2 = \omega_i + \omega_i^* D_{i,t}^{vol} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \\ a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1 - D_{i,t}^{level}) + \epsilon_i^* D_{i,t}^{level}) \\ \omega_i^* = \mu_{\omega^*} + \delta' \mathbf{x}_{i,T_i^*} + \varepsilon_i, \end{array}$$

5.4.1 Most elementary simulation setup

In order to investigate the Synthetic Volatility Forecasting method

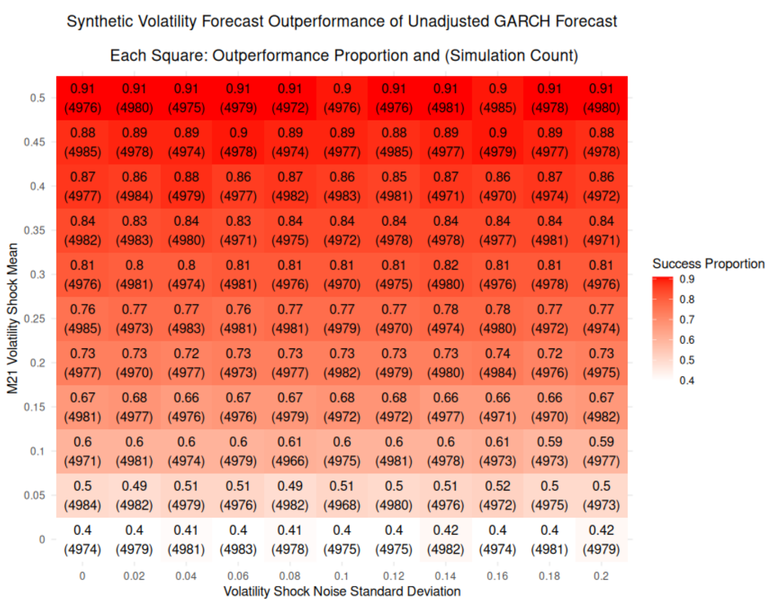

Figure 2: Fixed parameter values: $\alpha = .1, \beta = .82, \mu_x = 1, \sigma_x = .1$

In Figure **??**, when only two parameters are varied, the shock signal and the shock noise, we observe several encouraging phenomena:

- For any column selected, an increasing trend exists as the shock signal increases.
- 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 anemic.

As we see in Figure **??**

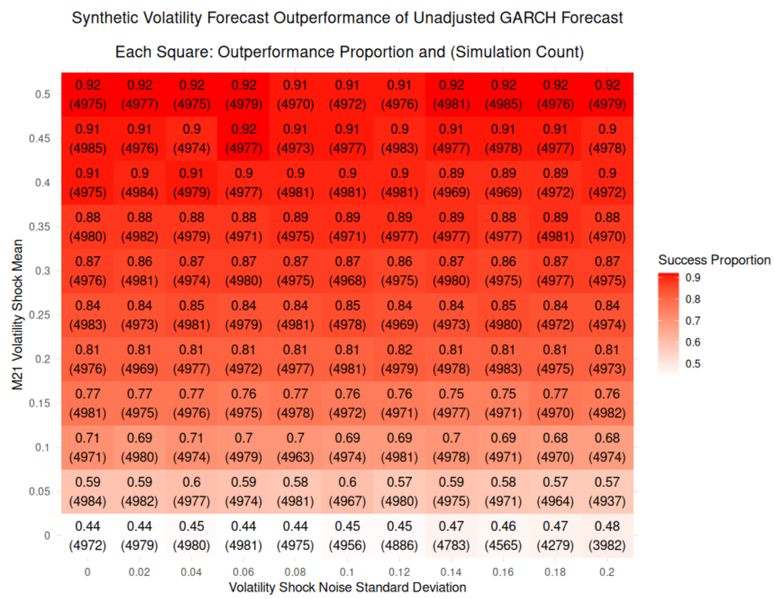

Figure 3: Fixed parameter values: $\alpha = .82, \beta = .1, \mu_x = 1, \sigma_x = .1$

5.4.2 Dual shock: both a level and volatility shock at $T^* + 1$

Hypothesis: Method should do less well under level shock.

5.4.3 Length of the Volatility Shock

5.4.4 Length of the Level Shock

5.4.5 Shorter Series

5.4.6 Fat tails in innovations

5.4.7 Lengthening the measurement period

5.4.8 Changing the optimization norm for distance-based weighting

5.4.9 Asymmetric GARCH

6 Real Data Example

We show the applicability of Synthetic Volatility Forecasting 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

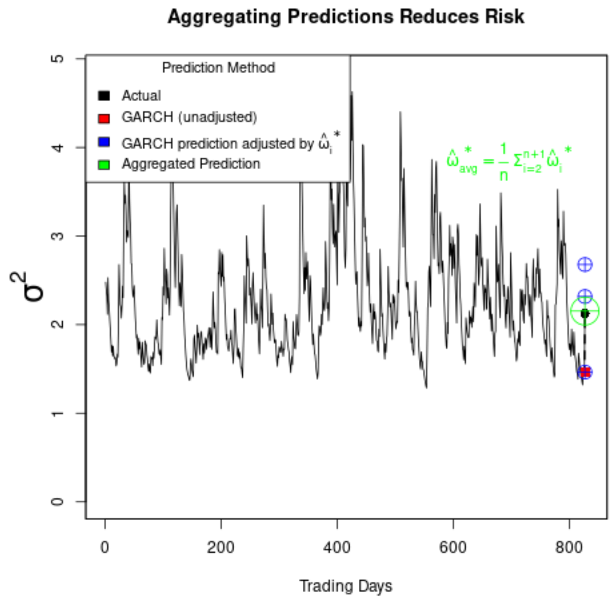

Figure 4: Dual shock: both a level and volatility shock at $T^* + 1$

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 upon a Trump victory. 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.

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[4] (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

[4]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.

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 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.

3. **Donor pool construction** Synthetic Control, an 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 (**??**). 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 faces a decidedly more difficult task 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 (**?**), 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.

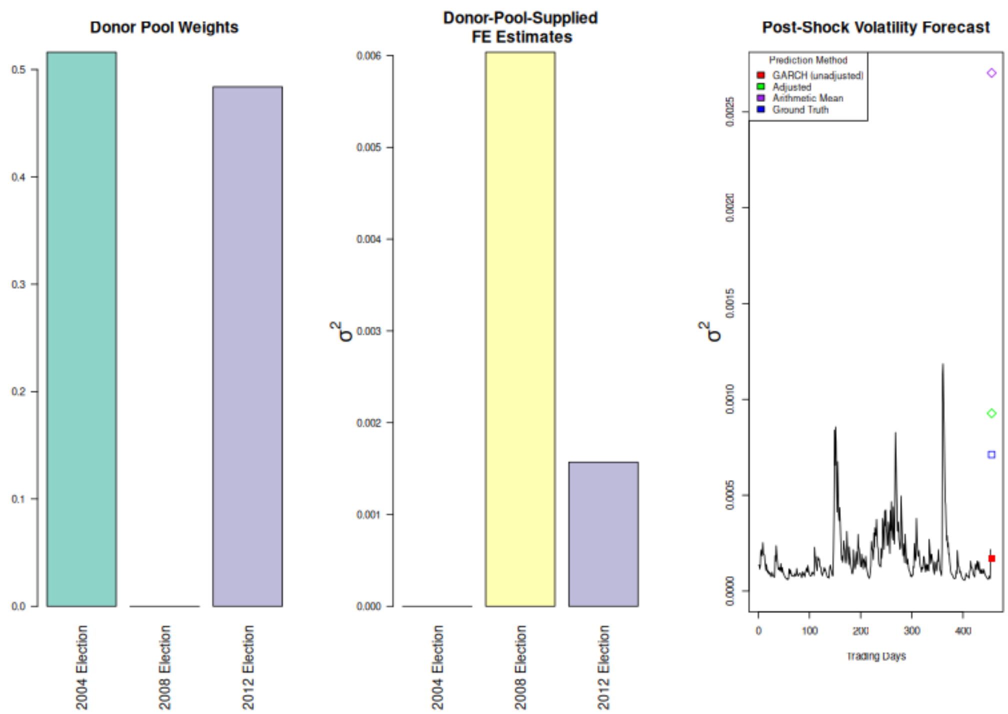

Figure 5: The volatility induced by the 2016 US election

We now discuss the three subplots in Figure **??** in order from left to right. On the left, we see that the distanced-based weighting places nearly equal weight on the 2004 and 2012 elections, with only neglible weight on the 2008 election. Assuming an approximately correct specification of the covariates, this is interpreted to mean that even of the 2016 US election had a general climate of risk and tension less extreme than 2008 and more similar to the 2004 and 2012 elections. In the middle plot, we notice that the fixed effect estimates for 2008 and 2012, 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 one data point, they are theoretically high in variance. On the right, we observe in black the fitted values of σ^2 given 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, the Synthetic Volatility Forecasting method comes reasonably close the ground truth. The prediction is not only directionally correct; it 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 because simple averaging ignores the radically different conditons on the even 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 no singular value represents more than 50% of the sum of singular values. Indeed, the first four singular values descend from 50% to 25% to 21% to 5% of the the cumulative variation, indicating a low concentration or redundancy of information. Second, we follow **?** 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 2004, 2012, and the time series under study.

7 Discussion

7.1 Connection to forecast combination methods

What we are combining, if anything, is subcomponents of forecasts, not forecasts themselves. However, from a broader perspective, forecast combination is an inapt term for what is being proposed here. First, the donors are not forecasts.

7.2 Why is the proposal here unlike KNN regression?

In KNN regression, the hyperparameter K must be learned. In Synthetic Volatility Forecasting, 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.

7.3 Should we gather as many donors as possible and pick them quantitatively?

It would be counter to the method proposed herein 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.

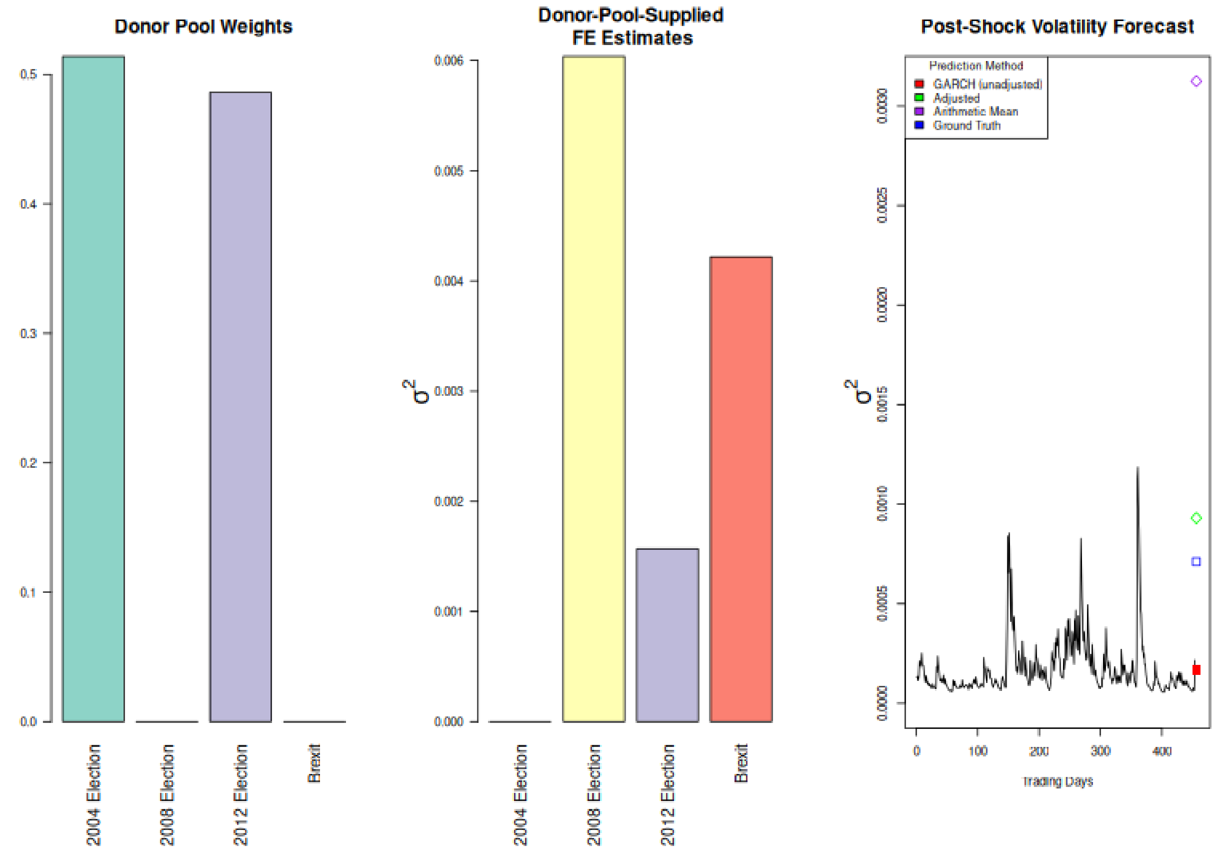

Figure 6: The volatility induced by the 2016 US election, Brexit excluded

7.4 News shock literature review

We adapt the news shock framework of **?**:

$$z_{T^*+1}^{\text{irregular news shock}} := z_{T^*+1}^{\text{irregular news}} - \mathbb{E}_{T^*}[z_{T^*+1}^{\text{irregular news}}]$$

In our adapted schema, irregular news shocks are $\mathcal{F}_{T^*}$-zero-expectation events governed by a GARCH-X process. As such, $z_{T^*+1}^{\text{irregular news shock}}$ admits of the decomposition $z_{T^*+1} = \sigma_{T^*+1}\epsilon_{T^*+1}$

7.5 Improvements of this current work over its predecessor

This present work extends **?** 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 $L_{\text{vol}}, L_{\text{level}} \in \mathbb{Z}^+$, and the weights applied to the donor pool can come from more exotic subsets of $\mathbb{R}^n$.

At a technical level, the most substantial insight related to **?** is that the GARCH model provides an elegant forecast function for squared observations of the time series under study. To see this, for each i, t, define $\eta_{i,t} := a_{i,t}^2 - \sigma_{i,t}^2$, and by adding $\eta_{i,t}$ to both sides of the conditional variance specification of $\mathcal{M}_1$ and $\mathcal{M}_2$, we obtain the standard ARMA representation of a GARCH(m,s):

$$a_{i,t}^2 = a_{i,t}^2 - \sigma_{i,t}^2 + \sigma_{i,t}^2 = \eta_{i,t} + \sigma_{i,t}^2 = \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \tag{21}$$

$$= \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} (a_{i,t-j}^2 - \eta_{i,t-j}) + \gamma_i^T \mathbf{x}_{i,t} \tag{22}$$

$$= \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{\max\{m_i,s_i\}} (\alpha_{i,k} + \beta_{i,k}) a_{i,t-k}^2 - \sum_{j=1}^{s_i} \beta_{i,j} \eta_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{23}$$

Then for any i, t, $\mathbb{E}[a_{i,t}^2]$ is nothing more than the expectation of the right-hand side of (**??**), where $\mathbb{E}[\eta_{i,t}] = 0$ for any i, t. Hence $\mathbb{E}[a_{i,t}^2] = \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{\max\{m_i,s_i\}} (\alpha_{i,k} + \beta_{i,k}) a_{i,t-k}^2 + \gamma_i^T \mathbf{x}_{i,t}$

1. 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?

2. See Tsay p. 133: if t is the last index that has been observed, then our two-step-ahead forecast is made easier by rewriting the volatility equation as $\sigma_{i,t+1}^2 = w_i + (\alpha_i + \beta_i)\sigma_{i,t}^2 + \alpha_1 \sigma_{i,t}^2 [\epsilon_{i,t}^2 - 1] = w_i + (\alpha_i + \beta_i)\sigma_{i,t}^2 + \alpha_1 \sigma_{i,t}^2 [\epsilon_{i,t}^2 - 1]$. Usually, we assume unit variance (though not necessarily a N(0,1)). However, what is the variance is smaller than 1? Then in conditional expectation, the term $\alpha_1 \sigma_{i,t}^2 [\epsilon_{i,t}^2 - 1]$ will be negative. So it seems that for level shock, we want the signal to noise ratio to be large because the mean is high, not because the variance is low.

7.6 Why not parameterize the volatility shock as a change in coefficients?

1. Swamped by shock: If the goal is near-term forecasting, then we don't really care how those coefficients may be changing. The assumption is that if any changes occur, their influence cannot account for the large shock. Given that the GARCH parameters for a stationary GARCH process must sum to less than 1, there is an upper bound to the forecastable change.

2. Practically infeasible: If we did that, we would need quite a few time points to reliably estimate the change in the GARCH parameters. We don't have that.

8 Supplement

We analyze the real-world example with Brexit included