Post-shock Volatility Forecasting Using Aggregated Shock Information

David Lundquist*, Daniel Eck†

Department of Statistics, University of Illinois at Urbana-Champaign

February 16, 2024

Abstract

We develop a procedure for forecasting the volatility of a time series immediately following a news shock. Adapting the synthetic prediction 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 the unobservable but easily-estimated realized volatility (RV). 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. A real-world application is provided, as are simulation results suggesting the conditions and hyperparameters under which our method thrives.

1 Introduction

Reacting to a seemingly unprecedented event might involve the question: what, if anything, does it resemble from the past? Such might be the case with event-driven investing strategies, where the identification of the event could arise via the news pages or corporate communications and hence contains a qualitative, narrative element [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] argue "[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 findings, 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. 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 $(r_t)_{t\in\mathbb{N}}$. The reasons for this are at least two-fold. First, 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

*davidl11@ilinois.edu

†dje13@illinois.edu

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. Second, the literature on return series predictability based on news has undergone a considerable shift over the past 50 years. Whereas once returns were thought to react to specific news events, now stock price movements are believed to be overwhelmingly noise-based (see Boudoukh et al. [2019] and references therein).

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 [Engle and Patton, 2001], prompting lines of inquiry to search for novel settings and challenges, as we do in this paper. No matter 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 Bauwens et al. [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 larger volatility than positive returns [Hansen et al., 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.

The same problem and more exists for Realized GARCH model, which incorporates so-called "realized measures" of volatility, like implied volatility. GARCH models have been shown slow to adapt to spikes in volatility [Andersen et al., 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]. Hansen et al. [2012] 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. We note also that 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 out-of-the-money.

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. Synthetic Volatility Forecasting 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]). 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: intercept corrections that attempt to account for future events, without hard-coding an ad-hoc adjustment into the model specification, and second, 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 et al. [2020] a introduces a similarity-based forecasting procedure for time-varying coefficients of a linear model. Foroni et al. [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.

Given the preceding review, we pause to take stock of several key components of forecast adjustments:

- Automated or non-automated
- The collection of data assembled to perform the correction
  - Internal (the time series itself)
  - External
- The parametric term to be corrected (e.g. intercept, coefficients), if any
- 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)

The foregoing provides a menu, of sorts, from which the researcher can construct an adjusted forecast procedure. Our procedure is an automated procedure for intercept adjustments using either 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 et al. [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] adapt 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 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.

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 scholarship 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-minization, and model interpretability. Guerrón-Quintana and Zhong [2017], citing previous work in the literature, 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. 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 [Timmermann, 2006]. 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. The key story behind this parameterization is that as the strength of the signal increases relative to the idiosyncratic error, the GARCH estimation of these effects increases in accuracy. From this, it follows that the aggregated shock estimate increases in accuracy.

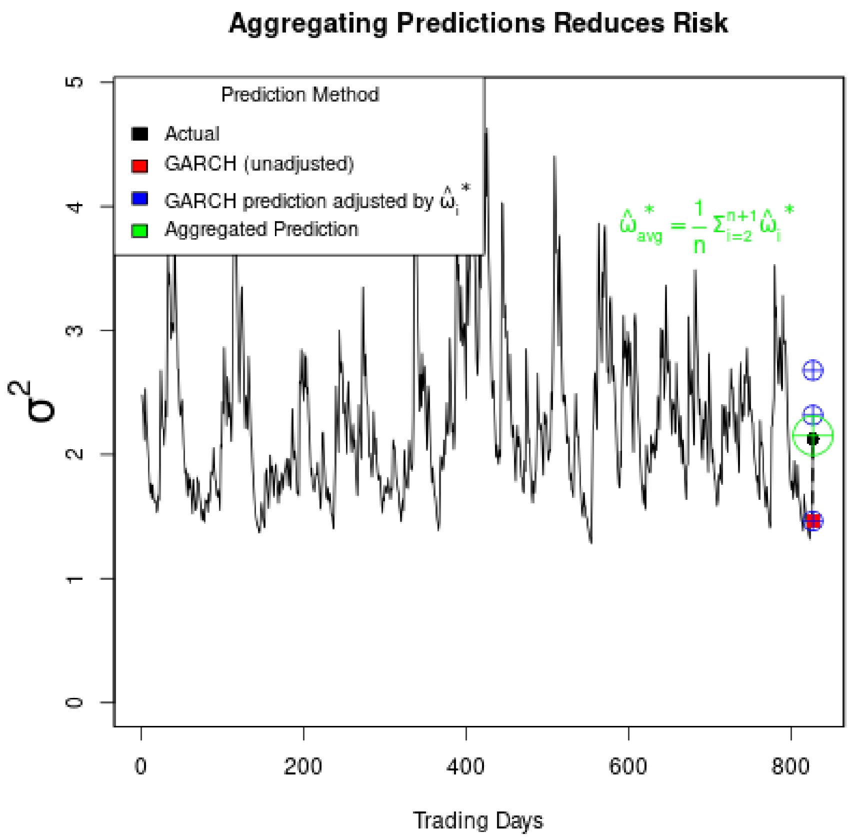


Figure 1: The time series experiences a volatility shock at a uniformly distributed point in the interval [502, 2520] trading days, corresponding to trading years on the interval [2,10]. We have truncated the series following the shock. The GARCH prediction fails even to approach the volatility spike at $T^* + 1$, as do some adjusted predictions. In contrast, the GARCH forecast adjusted by $\overline{\hat{\omega}^*}$, the arithmetic mean of the estimated shocks, recovers the shock in the time series under study.

2 Setting

2.1 A Primer on GARCH

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 [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 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 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. As Cont [2001] 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 $\mathbb{E}[a_t^2] = \text{Var}[a_t]$. 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]) is given by

$$
\begin{aligned}
&\sigma_t^2 = \omega + \sum_{k=1}^{m} \alpha_k a_{t-k}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 \\
&a_t = \sigma_t \epsilon_t \\
&\epsilon_t \overset{iid}{\sim} F \\
&\forall k, j, \alpha_k, \beta_j \geq 0 \\
&\forall t, \omega, \sigma_t > 0
\end{aligned}
$$

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

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

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t}, t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. Let $\mathbf{y}_{i,t} = (r_{i,t},\ \mathbf{x}_{i,t})$ where $r_{i,t}$ is a scalar response (typically the log-return of a financial asset) and $\mathbf{x}_{i,t}$ 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

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

2.2 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 shock distributions. 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.

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

$$\mathcal{M}_1\colon \begin{array}{l} \sigma^2_{i,t} = \omega_i + \omega^*_i D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \\ a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1 - D^{level}_{i,t}) + \epsilon^*_i D^{level}_{i,t}) \\ \omega^*_i = \mu_{\omega^*} + \delta' \mathbf{x}_{i,T_i^*} + \varepsilon_i, \end{array}$$

with random effects structure

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

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 $\mathbf{x}_{i,t} \in \mathbb{R}^p$. 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 further define the filtrations

$$\begin{aligned} \mathcal{F}^{\mathcal{M}_1}_{i,t} &= \{(\mathbf{x}_{i,t}, \omega^*_i, \epsilon_{i,t}, \epsilon^*_{i,t}, \varepsilon_{i,t})\colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\} \\ \mathcal{F}^{\mathcal{M}_2}_{i,t} &= \{(\mathbf{x}_{i,t}, \omega^*_i, \epsilon_{i,t}, \epsilon^*_{i,t}, \varepsilon_{i,t})\colon t = 1, \ldots, T_i, i = 2, \ldots, n+1\} \end{aligned}$$

2.3 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 that for each of the n donors, we posit p distinct covariates in the functional form of the shock. The volatility profile is defined as the $pn \times n$ matrix

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

Several comments are in order here. Although as shown, $\mathbf{V}_{p,n}$ displays 'balance' in that each of the p covariates is replicated n times, in practice, it may be that some donors do not possess or do not require such information.

2.4 Justifying the Shock Parameterization

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 et al. [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 an exogenous 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].

3 Post-shock Synthetic Volatility Forecasting Methodology

3.1 Forecasting

A GARCH model is an ARMA on the squares of the observed scalar time series [Tsay, 2005]. 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 present forecasts for GARCH(1,1), without loss of generality. For $h = 2, 3, 4, ...$, the conditional expectation is computed recursively, as is standard for iterative autoregressive forecasts.

We present two forecasts:

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

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

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 $(\{\hat{\omega}^*_i\}_{i=2}^{n+1}, \{\mathbf{x}_{i,T^*}\}_{i=2}^{n+1})$. We wish to recover weights $\{w_i\}_{i=2}^{n+1} \in \Delta^n$ that will minimize expected loss. Let $L(\cdot, \cdot)$ denote any loss function, not necessarily symmetric, mapping two quantities to the non-negative reals. Let $\hat{\sigma}^2$ denote any forecast for σ^2, where we consider an $h = 1$ step-ahead forecast, for simplicity. Formally, our optimization problem is

$$\underset{\{w_i\}_{i=2}^{n+1} \in \Delta^n}{\arg\min} \ \mathbb{E}[L(\sigma^2, \hat{\sigma}^2)]$$

We now 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 $\hat{\omega}^* \coloneqq \sum_{i=2}^{n+1} w_i \hat{\omega}^*_i$. 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, *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 risk minimization in section 4.

3.3 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^2_t 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 variance (RV).

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 Andersen and Teräsvirta [2009]. 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. We will proceed using $m = 77$, corresponding to the 6.5-hour trading day chopped into 5-minute blocks, with the first block ommitted in order to ignore unusual trading behavior at the start of the day.

3.4 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

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

Finally, we introduce the QL (quasi-likelihood) Loss [Brownlees et al., 2011]:

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

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 Brownlees et al. [2011] 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 $\mathbf{x}_{i,t}$. Assume also that $r \geq 2$, which is necessary in order to isolate the effects of the level shock $\epsilon^*_{i,t}$. Then

$$\sigma^2_{i,T^*+r+1} = \omega_i + \alpha_i a^2_{T^*+r} + \beta_i \sigma^2_{i,T^*+r} \tag{1}$$
$$= \omega_i + \alpha_i(\sigma_{i,T^*+r}\epsilon_{T^*+r})^2 + \beta_i \sigma^2_{i,T^*+r} \tag{2}$$
$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha_i(\epsilon_{T^*+r})^2 + \beta_i) \tag{3}$$

In (1), 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 [Zivot, 2009]. Indeed, if we assume unit variance for all $\epsilon_{i,t}$ except, of course, $\epsilon_{i,t}^*$, then we have

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

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

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

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), 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). □

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

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

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

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:

1. 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^*.

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.

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):

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

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 $\mathcal{M}_1$, 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 Han and Kristensen [2014] prevail.*

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

Lemma 1. *Under assumption 2, $\forall i, \{\omega^*_{i,t}\}_{t=0,...,T_i}$ 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 Han and Kristensen [2014]. □

4.2 Consistency of the Synthetic Volatility Forecast Function

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

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

$$\mathbb{E}_{T^*}[\sigma^2_{i,t+1}|\mathcal{F}_{T^*}] = \omega_i + \omega^*_i D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{8}$$

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

$$\hat{\sigma}^2_{i,t+1}|\mathcal{F}_{T^*} = \hat{\omega}_i + \hat{\omega}^*_i D^{vol}_{i,t} + \sum_{k=1}^{m_i} \hat{\alpha}_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \hat{\beta}_{i,j} \hat{\sigma}^2_{i,t-j} + \hat{\gamma}_i^T \mathbf{x}_{i,t} \tag{9}$$

which converges in probability to

$$\hat{\sigma}^2_{i,t+1}|\mathcal{F}_{T^*} = \omega_i + \omega^*_i D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j} \sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{10}$$

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

□

4.3 Risk Minimization Using Synthetic Volatility Forecasting

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

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

$$\begin{aligned}
&QL(\hat{\sigma}^2_{t+1,unadjusted}, \sigma^2_{t+1}) - QL(\hat{\sigma}^2_{t+1,adjusted}, \sigma^2_{t+1}) && (11)\\
&= (\frac{\hat{\sigma}^2_{t+1,unadjusted}}{\sigma^2_{t+1}} - \log \frac{\hat{\sigma}^2_{t+1,unadjusted}}{\sigma^2_{t+1}} - 1) - (\frac{\hat{\sigma}^2_{t+1,adjusted}}{\sigma^2_{t+1}} - \log \frac{\hat{\sigma}^2_{t+1,adjusted}}{\sigma^2_{t+1}} - 1) && (12)\\
&= \frac{\hat{\sigma}^2_{t+1,unadjusted} - \hat{\sigma}^2_{t+1,adjusted}}{\sigma^2_{t+1}} + \log \frac{\hat{\sigma}^2_{t+1,adjusted}}{\hat{\sigma}^2_{t+1,unadjusted}} && (13)
\end{aligned}$$

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 (13) can be expressed as

$$\begin{aligned}
&\frac{(\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2) - (\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2 + \hat{\omega}^*)}{\sigma^2_{t+1}} + \log \frac{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2 + \hat{\omega}^*}{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2} && (14)\\
&= \log \frac{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2 + \hat{\omega}^*}{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta}\sigma_t^2} - \frac{\hat{\omega}^*}{\sigma^2_{t+1}} && (15)
\end{aligned}$$

By letting $f(\hat{\omega}^*) = \log \frac{\hat{\omega}+\hat{\alpha}a_t^2+\hat{\beta}\sigma_t^2+\hat{\omega}^*}{\hat{\omega}+\hat{\alpha}a_t^2+\hat{\beta}\sigma_t^2} - \frac{\hat{\omega}^*}{\sigma_{t+1}^2}$, it's 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, $\overline{\omega^*}$, like 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.*
Then $QL(\hat{\sigma}^2_{t+1,unadjusted}, \sigma^2_{t+1}) - QL(\hat{\sigma}^2_{t+1,adjusted}, \sigma^2_{t+1}) \xrightarrow{p} \log \frac{\sigma^2_{t+1}}{\sigma^2_{t+1}-\mathbb{E}[\omega_i^*]} - \frac{\mathbb{E}[\omega_i^*]}{\sigma^2_{t+1}} \geq 0$. *Hence, a correctly specified model will outperform the unadjusted forecast asymptotically.*

Proof of Proposition The conclusion follows from fact that the model is correctly specified and consistency is guaranteed by prior results. Finally, the function $g : (-\infty, \mathbb{E}[\omega_i^*]) \to \mathbb{R}$ given by $g(x) = \log \frac{\sigma^2_{t+1}}{\sigma^2_{t+1}-x} - \frac{x}{\sigma^2_{t+1}}$ is nonnegative. □

5 Numerical Examples

5.1 Modeling Setup

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.

5.1.1 Most elementary simulation: one-day volatility shock

In order to investigate the Synthetic Volatility Forecasting method, our most elementary simulation uses $\mathcal{M}_2$. We vary only two parameters.

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:

$$\mathcal{M}_2 : \begin{array}{l} \sigma^2_{i,t} = \omega_i + \omega_i^* D^{vol}_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a^2_{i,t-k} + \sum_{j=1}^{s_i} \beta_{i,j}\sigma^2_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \\ a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1 - D^{level}_{i,t}) + \epsilon_i^* D^{level}_{i,t}) \\ \omega_i^* = \mu_{\omega^*} + \delta' \mathbf{x}_{i,T_i^*} + \varepsilon_i \\ D^{level}_{i,t} \equiv 0 \end{array}$$

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

Hypothesis: Method should do less well under level shock.

5.2 Peformance Metrics

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

5.3 Monte Carlo results

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.

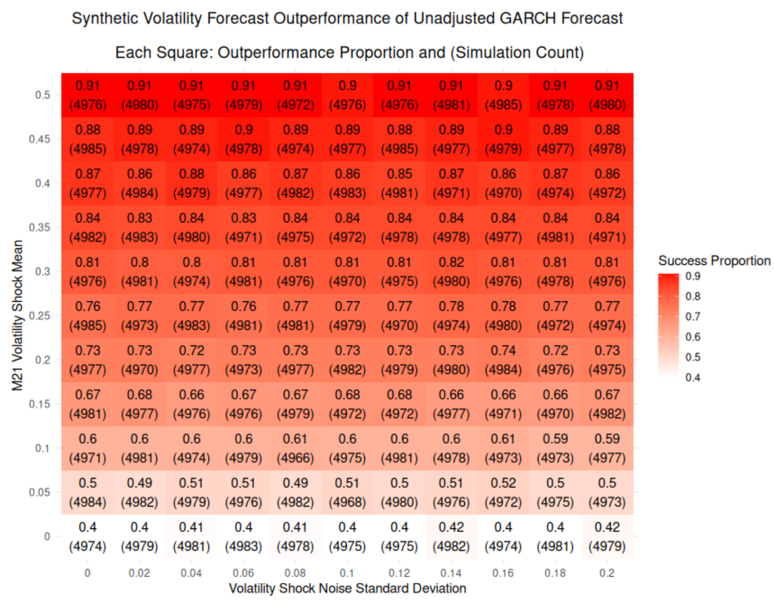


Figure 2: Fixed parameter values: $\alpha = .1, \beta = .82, \mu_x = 1, \sigma_x = .1$

If we switch the values of α and β, we see similar behavior, as in 3. 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^*.

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 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,

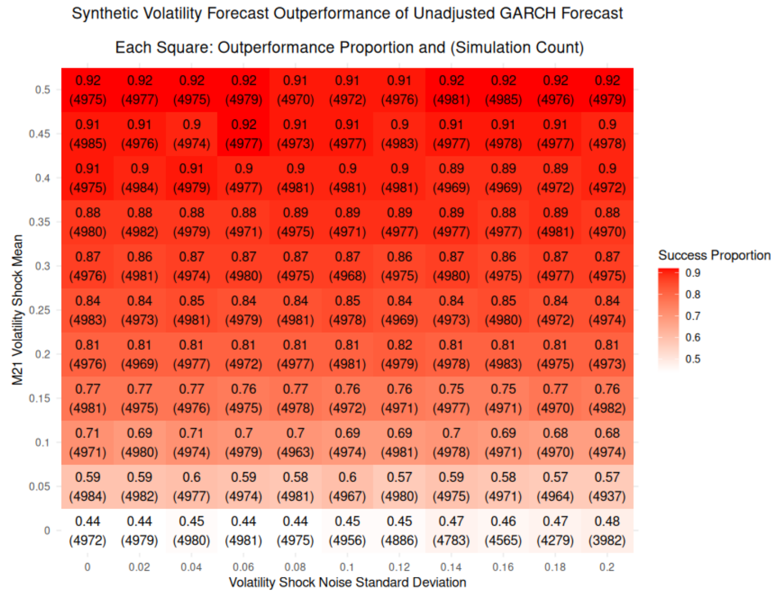
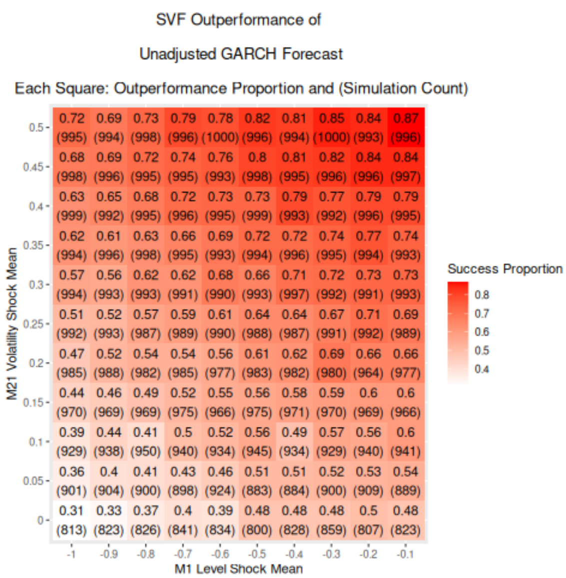
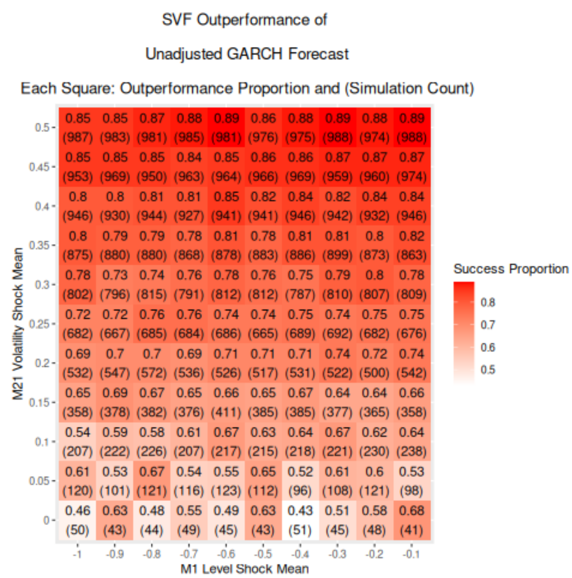


Figure 3: Fixed parameter values: $\alpha = .82, \beta = .1, \mu_x = 1, \sigma_x = .1$



(a) $\alpha = .1, \beta = .82, \mu_x = 1, \sigma_x = .1$



(b) $\alpha = .82, \beta = .1, \mu_x = 1, \sigma_x = .1$

Figure 4: A figure with two subfigures

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" [Blo, 2024]. Gleaning signals from financial options markets and betting markets, Wolfers and Zitzewitz [2016] 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. 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 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 spread between AAA and BAA corporate debt, widely considered a proxy for lending risk [Goodell and Vähämaa, 2013, Kane et al., 1996]. 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 [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 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 [Andersen and Teräsvirta, 2009], 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.

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

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

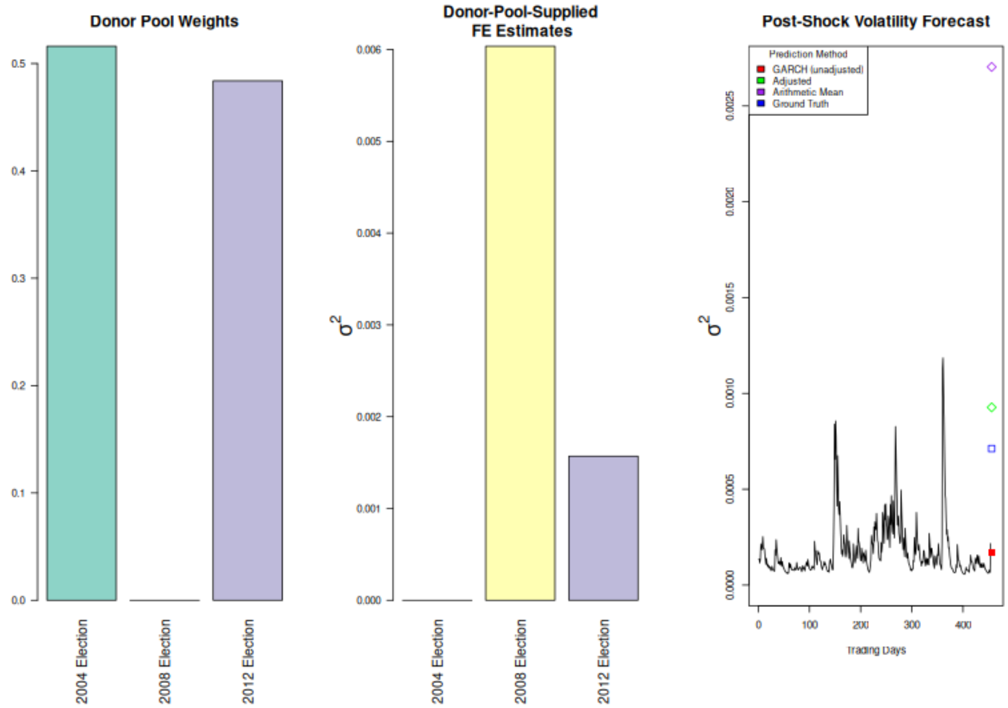


Figure 5: The volatility induced by the 2016 US election

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 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, 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 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 Steegen et al. [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 2004, 2012, and the time series under study.

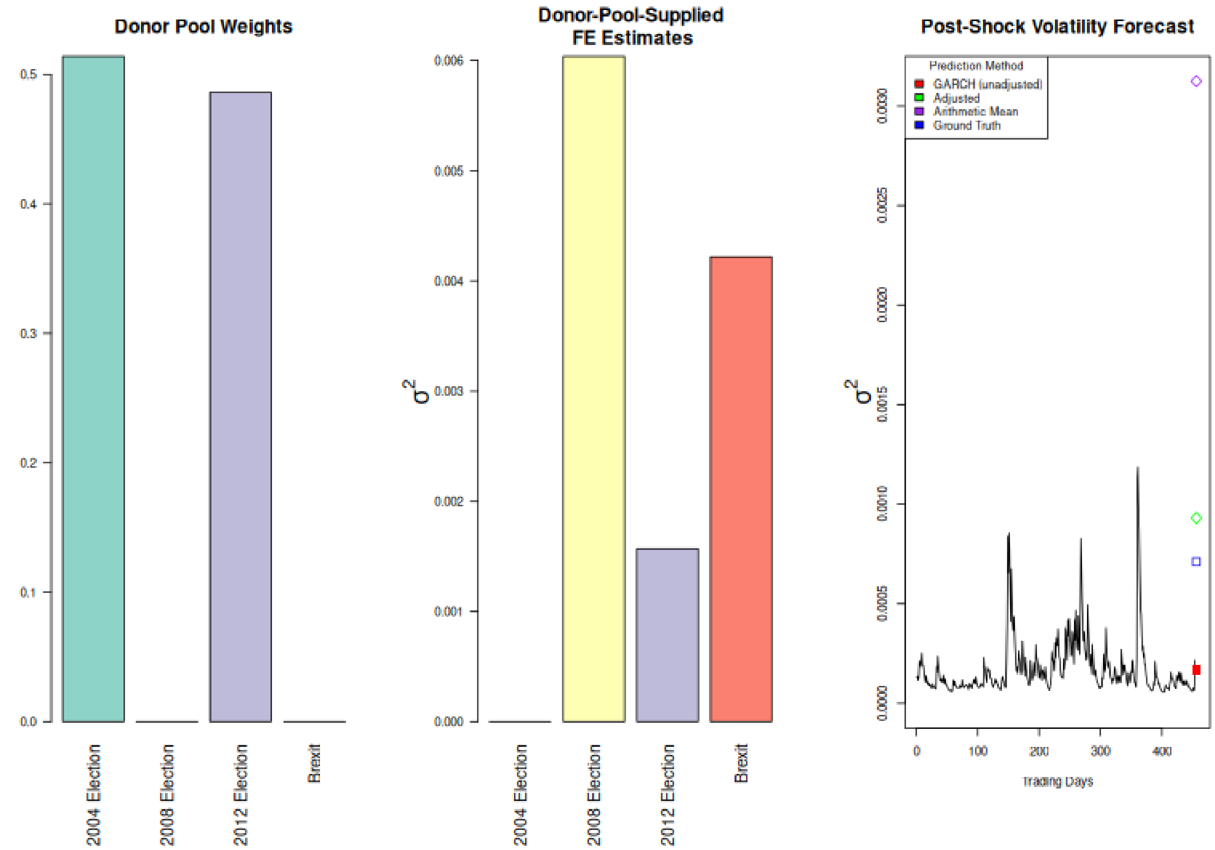


Figure 6: The volatility induced by the 2016 US election, Brexit excluded

7 Discussion

7.1 Connection to forecast combination methods

Because the method proposed uses a point in Δ^n, it is important to head-off possible confusion. What we are proposing is not a forecast combination method. What we are 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 like and unlike KNN regression?

1. We are not trying to learn a function, first and foremost. We are trying to estimate a parameter.

2. 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 et al., 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].

3. 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 $\{y_i\}_{i=1}^{N_{test}}$

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 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. 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 Synthetic Prediction 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 Synthetic Prediction, the determination of nearness begins with a qualitative step, i.e. curating the units between which we will calculate distances and from we will ultimately derive weights.

7.3 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.

citetabadie2022synthetic 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.

7.4 Why not use the covariates as covariates in the GARCH model alone or in the GARCH model as well?

The data-generating process does not specify that volatility series are well-modeled using ever-present exogenous covariates, and this is fundamentally due to the lack of theoretical justification for such an exogeneity assumption.

7.5 Do there exist optimal weights?

There are at least two senses of 'optimal' weights that one might be interested in here. First, we can think of optimal weights as a set $\{w_i\}_{i=2}^{n+1}$ such that $\omega_1 = \sum_{i=2}^{n+1} w_i\hat{\omega}_i$, i.e., ω_1 belongs to convex hull of the estimated shocks. There also exists a second sense of optimality, in which, for a given σ_t^2, the weights allow us to predict with zero loss.

7.6 Comments on the shock models

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.

7.7 Adjustment Using OLS on the Donor Shocks

Consider least-squares estimation for the weight vector $\vec{w}$:

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

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.

7.8 News shock literature review

We adapt the news shock framework of Kilian and Lütkepohl [2017]:

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

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 $z_{T^*+1} = \sigma_{T^*+1}\epsilon_{T^*+1}$

7.9 Improvements of this current work over its predecessor

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 $L_{\text{vol}}, L_{\text{level}} \in \mathbb{Z}^+$, 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 Lin and Eck [2021] 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):

$$a_{i,t}^2 = a_{i,t}^2 - \sigma_{i,t}^2 + \sigma_{i,t}^2 = \eta_{i,t} + \sigma_{i,t}^2 = \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j}\sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \tag{16}$$

$$= \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j}(a_{i,t-j}^2 - \eta_{i,t-j}) + \gamma_i^T \mathbf{x}_{i,t} \tag{17}$$

$$= \eta_{i,t} + \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{\max\{m_i,s_i\}} (\alpha_{i,k} + \beta_{i,k}) a_{i,t-k}^2 - \sum_{j=1}^{s_i} \beta_{i,j}\eta_{i,t-j} + \gamma_i^T \mathbf{x}_{i,t} \tag{18}$$

Then for any i, t, $\mathbb{E}[a_{i,t}^2]$ is nothing more than the expectation of the right-hand side of (18), 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 $\sigma_{i,t+1}^2 = w_i + (\alpha_i + \beta_i)\sigma_{i,t}^2 + \alpha_1\sigma_{i,t}^2[\epsilon_{i,t}^2 - 1] =$

$w_i + (\alpha_i + \beta_i)\sigma^2_{i,t} + \alpha_1\sigma^2_{i,t}[\epsilon^2_{i,t} - 1]$. 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 $\alpha_1\sigma^2_{i,t}[\epsilon^2_{i,t} - 1]$ 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.

7.10 Model Adjustment Using Synthetic Methods: A Global Overview

Here we distinguish what in this work is essential and inessential to synthetic forecasting.

1. k-dimensional random object to predict
2. correction term subject to shock, with DGP shared among donors
3.

7.11 News Sentiment

Within the vast literature on financial sentimental analysis, theree exists a sub-literature on financial volatility forecasting using news and news sentiment [Atkins et al., 2018]. The setting herein is one where the valence of the sentiment is reasonably well-understood, and hence the direction of volatility can be well-predicted, although not perfectly, by human intuition.

7.12 Uniqueness of the weight vector $\vec{w}$

8 Supplement

We analyze the real-world example with Brexit included.

References

Jan 2024. URL https://www.bloomberg.com/news/articles/2024-01-21/trump-s-2016-win-shook-markets-traders-won-t-get-fooled-again.

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, 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 Timo Teräsvirta. Realized volatility. In *Handbook of financial time series*, pages 555–575. Springer, 2009.

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

Adam Atkins, Mahesan Niranjan, and Enrico Gerding. Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4(2):120–137, 2018.

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.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223, 2001.

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.

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.

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.