Post-shock Volatility Forecasting Using Aggregated Shock Information

David Lundquist*

Department of Statistics, University of Illinois at Urbana-Champaign

December 3rd, 2021

Abstract

We develop a novel procedure for forecasting the volatility of a time series under study immediately following an exogenous news shock. Adapting the synthetic prediction framework of Lin and Eck [2021], we exploit series that have experienced similar shocks. We extract volatility signals from these series in order to aggregate properly their shock-induced excess volatilities. The volatility spikes are modeled as random effects and estimated as fixed effects. The aggregation of these estimates is 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) of the time series under study, 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 the standard GARCH model by permitting the inclusion of metrics like 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

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 and daily return series may exhibit behavior that makes inconvenient and uninterpretable the traditional methods of time series analysis. In contrast, the volatility of the return series may be less noisy, less structurally dependent on external factors, and more easily modeled as a weakly stationary process. Yet even if one could model and forecast price series and return series with high accuracy, that would not necessarily tell us much about the variability of such forecasts nor enlighten us about the evolution of the variability of $(P_t)_{t\in\mathbb{N}}$ over time. Modern portfolio management and theory requires information about at least the first two moments of a return series, if not higher. For these reasons and others, volatility modeling has come to dominate financial econometrics over the past four decades.

When $(P_t)_{t\in\mathbb{N}}$ is positive, as one would expect of a price series, the percentage change over any interval of length $k \in \mathbb{N}$ can be approximated by $\log(\frac{P_{t+k}}{P_t})$, owing to the fact that for any nonnegative real number y, $\log(y+1) \le y$. Equality holds at exactly y = 0, corresponding to the case where $P_{t+k} = P_t$. It is easily verified that the approximation works well when the ratio $\frac{P_{t+k}}{P_t} \in (.85, 1.15)$, a return absolute value of 15% or less. The daily log return series $\log(\frac{P_{t+1}}{P_t})$, however, not only provides an approximation to the daily return; the logarithm of a price series has the helpful property of being supported on $\mathbb{R}$. Partly due
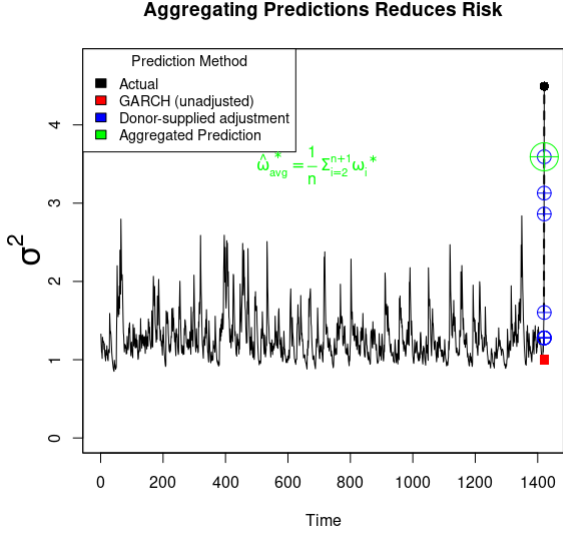
---

*davidl11@ilinois.edu

Figure 1: An example of aggreation using arithmetic mean

to that, it earns a pedigree for its role in geometric Brownian motion, the foundation of much of modern financial mathematics, including the Black-Sholes differential equation [Tsay, 2005]. Additionally, the log transformation is a tried-and-true variance-reduction technique for heteroskedastic data [Faraway, 2016] and non-stationary time series. In many applications, the log-return of a positive time series will allow us to reject the null of a unit-root process using the methods pioneered in Dickey and Fuller [1979] and Dickey and Fuller [1981], thus paving the way for the vast family of autoregressive-moving-average models.

However, no matter how a time series or its transformations is modeled, forecasting in the prescence of exogenous shock events requires a methodological framework that sensibly incorporates relevant information that has yet to manifest itself in market price or derivative quantities like volatility. In this setting, regime-change models 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]. As a separate problem, GARCH models have been shown slow to adapt to spikes in volatility (find first authors to point this out). Engle [2002] explored the use of implied volatility as an external regressor in a GARCH model, that is, using so-called GARCH-X models. 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.

The approach herein can be viewed as an attempt to sidestep the functional complexity posited by Realized GARCH by substituting modeling assumptions. Synthetic volatility forecasting proceeds under the assumption that similar exogenous shock events occasion volatility shocks arising from a common shock distribution.

2   Setting

Figure 2 shows a simple example.

We follow convention and define the daily log-return as $r_t = \log(\frac{P_{t+1}}{P_t})$. 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 and maximum-likelihood estimation of parameters. These models assume a certain dependence structure between r_t and $(r_k)_{k \leq t}$, as well as all observations in between, 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, while a_t is simply the noise. As Cont [2001] explains, such an assumption is justified by the empirical finding that returns lack significant autocorrelation. We 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 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,

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

$$a_t = \sigma_t \epsilon_t$$

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

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

$$a_t = \sigma_t \epsilon_t$$

We will suppose that a researcher has multivariate time series data $\mathbf{y}_{i,t}$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$. We let $\mathbf{y}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t}, \mathbf{v}_{i,t})$ where $y_{i,t}$ is a scalar response, $\mathbf{x}_{i,t}$ is a vector of covariates that are revealed to the analyst prior to the observation of $y_{1,t}$, 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 $\mathbf{y}_{i,t}$ is subject to a news event immediately following $T_i^* \leq T_i + 1$ and before witnessing $T_i^* + 1$. To define an interesting setting, we will suppose that $T_1^* = T_1 + 1$, and $1 < T_i^* < T_i + 1$ for $i \geq 2$. We will suppose that $\mathbf{x}_{i,t=T_i^*}$ and $\mathbf{v}_{i,t=T_i^*}$ are observed before the shock is manifested on $y_{i,t=T_i^*+1}$.

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

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

$$a_{i,t} = \sigma_{i,t} \epsilon_{i,t}$$

3

We are interested in a multihorizon 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 maximum forecast horizon 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 subscript "method, ground truth denote the loss function L and a pair, (prediction method, ground truth estimator), for an h-step-ahead forecast. For example, a loss function of interest in this study is the 1-step-ahead MSE

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

In more generality, for an multihorizon volatility forecast with maximum 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 [Brownlees et al., 2011]:

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

What distinguishes QL Loss is that it is multiplicative rather than additive. This serves as a basis for some of its virtues. 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."

2.1 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 $\mathcal{F}_t$. Much of financial econometrics is devoted to decomposing P_t into random and non-random components as well as understanding the covariance structure of P_t, difficult though that may be in some circumstances. 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. 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.

2.1.1 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 [Christensen and Prabhala, 1998].

Find citation that Black-Scholes implied volatility is especially biased in times of a crisis.

2.1.2 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 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.

2.1.3 Historical Volatility

As discussed in section 2, 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.

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...

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

2.3 Aggregation Mechanism

Here we explain how we use the volatility profile to arrive at a set of nonnegative weights that sum to 1.

In Abadie et al. [2010], 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. Lin and Eck [2021] 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. 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.

2.4 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 at time T*

2.4.1 Volatility Shock 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

$$\mathcal{M}_1 : \sigma_{i,t}^2 = \omega_i + \omega_i^* D_{i,t}^{vol} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t} \tag{1}$$

$$a_{i,t} = \sigma_{i,t}(\epsilon_{i,t}(1 - D_{i,t}^{level}) + \epsilon_i^* D_{i,t}^{level}) \tag{2}$$

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

$$\omega_i^* \overset{iid}{\sim} \mathcal{F}_{\omega^*} \text{ with } \mathrm{E}_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \mu_{\omega^*}, \mathrm{Var}_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \sigma_{\omega^*}^2$$

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

$$\varepsilon_{i,t} \overset{iid}{\sim} \mathcal{F}_{\varepsilon} \text{ with } \mathrm{E}_{\mathcal{F}_{\varepsilon}}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_{\varepsilon}}(\varepsilon_{i,t}) = 1$$

$$\omega_i^* \perp\!\!\!\perp \epsilon_i^* \perp\!\!\!\perp \gamma_i \perp\!\!\!\perp \varepsilon_{i,t}.$$

WHAT FOLLOWS IS (a) defense of the modeling choices, (b) remarks on the models' implications and subtleties

---

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

[2]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 as a primitive quantity that affects volitility through market sentiment. 'It is what it is.'

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. This would be inelegant and would also lack motivation as a practical level. While we want to model the shock at $T^* + 1$ as large in absolute value, we also want to retain the property of a unitary source of noise.

Note that under the popular GARCH(1,1), a dual level-volatility shock has an marginal effect on the conditional variance $\sigma_{i,t}^2$ that 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

$$\sigma_{i,T^*+r+1}^2 = \omega_i + \alpha_i a_{T^*+r}^2 + \beta_i \sigma_{i,T^*+r}^2 \tag{1}$$

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

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

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

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

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

Proposition 1. *Let a_t be a mean-zero time series obeying a GARCH(1,1) specification 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 $r - 1 \geq max\{L_{i,vol}\}, L_{i,level}\}$,*

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

$$\tag{2}$$

Proof of Proposition   We claim

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

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

$$= \omega_i + \alpha\sigma^2_{i,T^*+r} + \beta\sigma^2_{i,T^*+r} \tag{3}$$

$$= \omega_i + \sigma^2_{i,T^*+r}(\alpha + \beta) \tag{4}$$

$$\tag{5}$$

The volatility equation of a GARCH(1,1) dictates that for any r, the one-step-ahead volatility is given by $\sigma^2_{i,T^*+r+1} = \omega_i + \alpha a^2_{i,T^*+r} + \beta\sigma^2_{i,T^*+r}$. By the mean-model assumption of a GARCH(1,1), we have $a_{i,t} = \sigma_{i,t}\epsilon_{i,t}$. By substitution, we arrive at equation (2) above. Using the unit-variance assumption regarding ϵ_{T^*+r}), we can compute explicitly the expectation in (2). Finally, by rearranging terms, we arrive at equation (4).   □

In other words, 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

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

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

$$= \omega_i + \alpha\sigma^2_{i,T^*+1}(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta\sigma^2_{i,T^*+1} \tag{3}$$

$$= \omega_i + \sigma^2_{i,T^*+1}(\alpha(\mu^2_{\epsilon^*} + \sigma^2_{\epsilon^*}) + \beta) \tag{4}$$

For each of the p entries in delta, the kth entry should have the property that
$\mathbb{E}[x_{i,t,k} * \delta_k] > 0$
In principle, you could have an M21 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:

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

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.

2.5 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.

2.6 Improvements of this current work over its predecessor

1. $y_{i,t}$ have shocks in both volatility and level.

2. The shocks can have length greater than 1.

3. 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. However, 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?

4. We may want a loss function that is asymmetric, i.e. penalizes upward and downward misses differently.

5. Lin+Eck 2021 requires exogenous information to arrive strictly between T^* and $T^* + 1$. However, we may want to relax this requirement so as to permit the exogenous information to be concurrent with the trading session or end of session on $T^* + 1$.

6. By offering more linear combination methods for the donor pool shocks (e.g. unit ball, affine hull, convex cone), we give the operator the ability to use donors such that the donor outcome series is negatively correlated with the time series under study.

7. See Tsay p. 94: 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.

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:

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

where the added random effects are

$$\omega_i^* \overset{iid}{\sim} \mathcal{F}_{\omega^*} \text{ with } \mathrm{E}_{\mathcal{F}_{\omega_i^*}}(\omega^*) = \mu_{\omega^*}, \mathrm{Var}_{\mathcal{F}_{\omega_i^*}}(\omega_i^*) = \sigma_{\omega^*}^2$$

	$\Sigma_\delta = I_p$	$\Sigma_\delta \neq I_p$
$\Sigma_{x_{i,T^*}} = I_p$	Easiest to model	Captures reasonable possibility of non-zero correlation in random effects
$\Sigma_{x_{i,T^*}} \neq I_p$	Captures empirically-verified phenomenon of non-zero correlation in covariates	Implausibly complicated to model

	Vershynin (2018)	Lin and Eck (2021)	SynthVolForecasting
observed quantity	y_i	NA	NA
estimated quantity	NA	α_i	ω_i^*
signal	$A_i x$	$\mu + \delta' x_{i,T^*}$	$\mu + \delta' x_{i,T^*}$
prior information	$x \in T$, T a convex	across donors, linear shock function equal in distribution	across donors, linear shock function equal in distribution
key assumption(s)	A_i independent $\forall i$, isotropic too?	x_{i,T^*} independent across donors; independent entries of x_{i,T^*}	x_{i,T^*} independent across donors; independent entries of x_{i,T^*}
noise	w	$\tilde{\varepsilon}$	$\tilde{\varepsilon}$
strategy	Observe y_i; then estimate x with small expected L2 loss	Estimate α_i for $i = 2, ..., n+1$; then take convex combination to predict α_1	Estimate ω_i^* for $i = 2, ..., n+1$; then take convex combination to predict ω_1^*

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

$$\delta \overset{iid}{\sim} \mathcal{F}_\delta \text{ with } \mathrm{E}_{\mathcal{F}_\delta}(\delta) = \mu_\delta, \mathrm{Var}_{\mathcal{F}_\delta}(\delta_i) = \Sigma_\delta$$

$$\varepsilon_{i,t} \overset{iid}{\sim} \mathcal{F}_\varepsilon \text{ with } \mathrm{E}_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = 0, \mathrm{Var}_{\mathcal{F}_\varepsilon}(\varepsilon_{i,t}) = 1$$

$$\tilde{\varepsilon}_i \overset{iid}{\sim} \mathcal{F}_{\tilde{\varepsilon}} \text{ with } \mathrm{E}_{\mathcal{F}_{\tilde{\varepsilon}}}(\tilde{\varepsilon}_i) = 0, \mathrm{Var}_{\mathcal{F}_\varepsilon}(\varepsilon_i) = \sigma_{\tilde{\varepsilon}}^2$$

$$\omega_i^* \perp\!\!\!\perp \epsilon_i^* \perp\!\!\!\perp \gamma_i \perp\!\!\!\perp \delta \perp\!\!\!\perp \varepsilon_{i,t} \perp\!\!\!\perp \tilde{\varepsilon}_i.$$

We further define the parameter sets

$$\begin{aligned}
\Theta &= \{(\omega_i^*, \epsilon_i^*, \gamma_i, \delta, \varepsilon_{i,t}, \tilde{\varepsilon}_i) : t = 1, \ldots, T_i, i = 2, \ldots, n+1\}, \\
\Theta_1 &= \{(\omega_i^*, \epsilon_i^*, \gamma_i, \delta, \varepsilon_{i,t}, \tilde{\varepsilon}_i) : t = 1, \ldots, T_i, i = 1\},
\end{aligned} \tag{1}$$

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

2.7 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

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

3 Simulation Setup and Results

Having introduced the model parameters above, we introduce estimation techniques that can be varied
   Figure 2 shows a simple example.

1. Additional measurement periods

Figure 2: An example of aggreation using arithmetic mean

2. Optimization norm

We also introduce simulation parameters that can be varied but are not random effects

1. Length of the volatility shock

2. Length of the level shock

4 Connection to Signal Recovery

References

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 Gustav Andersen, Richard A Davis, Jens-Peter Kreiß, and Thomas V Mikosch. *Handbook of financial time series*. Springer Science & Business Media, 2009.

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

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.

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

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

David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

David A Dickey and Wayne A Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, pages 1057–1072, 1981.

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.

Julian J Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2016.

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.

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

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