Synthetic Volatility Forecasting and Other Aggregation Techniques for Time Series Forecasting

Preliminary Exam

David Lundquist[1], Daniel Eck[2] (advisor)

March 27, 2024

[1]davidl11@ilinois.edu

[2]dje13@illinois.edu

A seemingly unprecedented event might make one ask

1. What does it resemble from the past?
2. What past events are most relevant?
3. Can we incorporate past events in a systematic, principled manner?

When would we ever have to do this?

- Event-driven investing strategies (unscheduled news shock)
- Pairs trading strategies
- Structural shock to macroeconomic conditions (scheduled news possibly pre-empted by news shock)
- Biomedical panel data subject to exogenous shock or interference

Example (Weekend of March 6th - 8th, 2020)

Coronavirus Health Executive Order

Public Safety

MARCH 7, 2020 | Albany, NY

At Novel Coronavirus Briefing, Governor Cuomo Declares State of Emergency to Contain Spread of Virus

Oil nose-dives as Saudi Arabia and Russia set off ‘scorched earth’ price war

PUBLISHED SUN, MAR 8 2020•9:01 AM EDT | UPDATED MON, MAR 9 2020•5:33 PM EDT

Oil crashes by most since 1991 as Saudi Arabia launches price war

By Matt Egan, CNN Business

3 minute read · Updated 3:21 PM EDT, Mon March 9, 2020

Punchline of the paper

Forecasting is possible under news shocks, so long as we incorporate external information to account for the nonzero errors.

Background and related methods

Volatility Modeling

- GARCH is slow to react (Andersen et al. 2003)
- Asymmetric GARCH models catch up faster but need post-shock data
- Realized GARCH (Hansen, Huang, and Shek 2012), in our setting, would require post-shock information and/or high-frequency data in order to outperform, and Realized GARCH is highly parameterized

Background and related methods

Forecast Augmentation

- Clements and Hendry 1998; Clements and Hendry 1996 laid the groundwork for modeling nonzero errors in time series forecasting
- Guerrón-Quintana and Zhong 2017 use a series' own errors to correct the forecast for that series
- Dendramis, Kapetanios, and Marcellino 2020 use a similarity-based procedure to correct linear parameters in time series forecasts
- Foroni, Marcellino, and Stevanovic 2022 adjust pandemic-era forecasts using intercept correction techniques and data from Great Financial Crisis
- Lin and Eck 2021 use distanced-based weighting (a similarity approach) to aggregate and weight fixed effects from a donor pool

Outline

The news has broken but markets are closed

- After-hours trading provides a poor forum in which to digest news
- The news constitutes public, material information relevant to one or more traded assets
- The qualitative aspects of the news provide basis upon which to match to past events

A Primer on GARCH

Let $\{a_t\}$ denote an observable, real-valued discrete-time stochastic process. We say $\{a_t\}$ is a strong GARCH process with respect to $\{\epsilon_t\}$ iff

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

$$a_t = \sigma_t \epsilon_t$$

$$\epsilon_t \overset{iid}{\sim} E[\epsilon_t] = 0, Var[\epsilon_t] = 1$$

$$\forall k, j, \alpha_k, \beta_j \geq 0$$

$$\forall t, \omega, \sigma_t > 0$$

Our Model is Nested Within GARCH-X

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

Model Preliminaries

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

Model Setup

For $t = 1, \ldots, T_i$ and $i = 1, \ldots, n+1$, the model $\mathcal{M}_1$ is defined as

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

with error structure

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

where $D^{return}_{i,t} = I(t \in \{T^*_i + 1, \ldots, T^*_i + L_{i,return}\})$ and $D^{vol}_{i,t} = I(t \in \{T^*_i + 1, \ldots, T^*_i + L_{i,vol}\})$ and $L_{i,return}, L_{i,vol}$ denote the lengths of the log return and volatility shocks, respectively. Let $\mathcal{M}_0$ denote the subclass of $\mathcal{M}_1$ models such that $\delta \equiv 0$. Note

Volatility Profile of a Time Series

$$
\mathbf{v}_{t} = \begin{pmatrix}
\hat{\alpha}_{1,t} & \hat{\alpha}_{t,2} & \cdots & \hat{\alpha}_{t,n} \\
\hat{\beta}_{1,t} & \hat{\beta}_{t,2} & \cdots & \hat{\beta}_{t,n} \\
\vdots & \vdots & \ddots & \vdots \\
{}^{RV}1,t & {}^{RV}2,t & \cdots & {}^{RV}{}_{n},t \\
{}^{RV}1,t-1 & {}^{RV}2,t-1 & \cdots & {}^{RV}{}_{n},t-1 \\
\vdots & \vdots & \ddots & \vdots \\
{}^{IV}1,t & {}^{IV}2,t & \cdots & {}^{IV}{}_{n},t \\
{}^{IV}1,t-1 & {}^{IV}2,t-1 & \cdots & {}^{IV}{}_{n},t-1 \\
\vdots & \vdots & \ddots & \vdots \\
|r_{1,t}| & |r_{2,t}| & \cdots & |r_{n,t}| \\
|r_{1,t-1}| & |r_{2,t-1}| & \cdots & |r_{n,t-1}| \\
\vdots & \vdots & \ddots & \vdots
\end{pmatrix},
$$

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

Significance of the Volatility Profile

Covariates chosen for inclusion in a given volatility profile may be levels, log differences in levels, percentage changes in levels, or absolute values thereof, among many choices.

Forecasting

We present two forecasts:

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

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

Excess Volatility Estimators

essentially observe the pair $(\{\hat{\omega}_i^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$.

We wish to recover weights $\{\pi_i\}_{i=2}^{n+1} \in \Delta^n$ leading to favorable forecasting properties.

These weights are used to compute $\hat{\omega}^* := \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$, our forecast adjustment term.

the set $\{\pi_i\}_{i=2}^{n+1}$ is deterministic, *modulo* any stochastic ingredient in the numerical methods employed to approximate $\mathbf{x}_{1,T^*}$ using a convex combination of donor covariates. We will say more about the properties of ω_i^* in section 4.

Following Abadie, Diamond, and Hainmueller 2010; Abadie and Gardeazabal 2003, let $\|\cdot\|_{\mathbf{S}}$ denote any semi-norm on $\mathbb{R}^p$, and define

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

Ground Truth Estimators

The time-varying parameter σ_t^2 is a quantity for which even identifying an observable effect in the real world is far more challenging. In this work, we use a common estimator of the variance called realized volatility (RV), one which has the virtue of being “model-free” in the sense that it requires no modeling assumptions (Andersen and Benzoni 2010). The realized variance itself can be decomposed into the sum of a continous component and a jump component, with the former being less predictable and less persistent (Andersen, Bollerslev, and Diebold 2007), cited in De Luca et al. 2006, two factors that further motivate the method employed herein. Suppose we examine K units of of time, where each unit is divided into m intervals of length $\frac{1}{m}$. We follow the notation of (Andersen and Teräsvirta 2009). Let $p_t = \log P_t$, and let $\tilde{r}(t, \frac{1}{m}) = p_t - p_{t-\frac{1}{m}}$. We estimate the variance of ith log return series using Realized Volatility of the K consecutive trading days that conclude with day t, denoted $RV_{i,t}^{K,m}$, using

$$RV^{K,m} \quad \frac{1}{} \sum^{Km} \tilde{r}^2(v/m, 1/m)$$

Loss Functions

We are interested in point forecasts for $\sigma^2_{1,T^*+h}|\mathcal{F}_{T^*}$, $h = 1, 2, ..., H$, the h-step ahead conditional variance for the time series under study, up to a forecast length of H. 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:

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

In more generality, for a volatility forecast with forecast length H, the MSE is

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

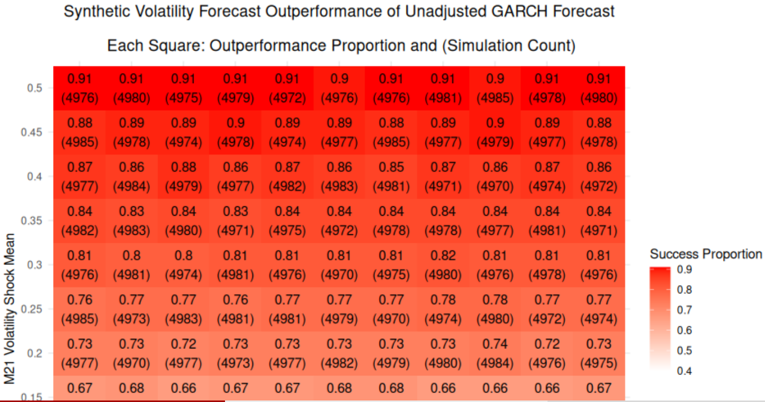
Also of interest in mean absolute percentage error for an h-step-ahead

Simplest Simulation Setup

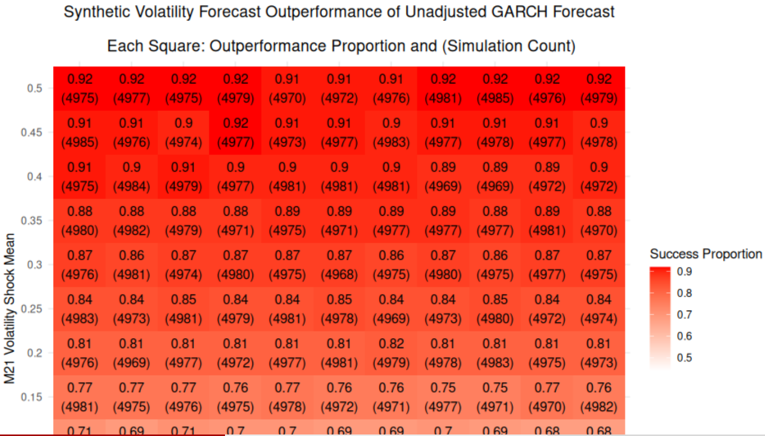
In order to investigate the Synthetic Volatility Forecasting method, our most elementary simulation uses $\mathcal{M}_1$. We vary only two parameters. Recall an $\mathcal{M}_1$ model on the volatility, which is characterized by an exogenous shock to the volatility equation generated by an affine function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

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

In Figure 1, when only two parameters are varied, the volatility shock signal and the volatility shock noise, we observe several encouraging phenomena. First, for any column selected, an increasing trend exists as the shock signal increases. Second, for almost all small values of the shock signal, the outperformance rate hovers around .5, supporting the hypothesis that in the absence of a signal, any level of noise renders the method no better at GARCH than a flip of a coin.



If we switch the values of α and β, we see similar behavior, as in 2. Here we cite Francq and Zakoian 2019, p. 22 in motivating simulations for large α values and then large β values. However, for small values of the shock mean, increasing noise does lead to fewer converged simulations, likely due to large negative realizations of the noise term, which in turn lead to near-zero and even negative estimates of the terms ω_i^*.



Additional Simulations

Alternative Data-Generating Processes

- Could we do all of the above with high-frequency data?
- Realized GARCH with High-Frequency Data
- Stochastic Volatility

Alternative Estimators and Estimands in Volatility Modeling

- Realized GARCH with High-Frequency Data
- Overnight returns instead of open-to-close
- Value-at-Risk using SVF-based $\hat{\sigma}_t^2$
- Signal Recovery Perspective (Ferwana and Varshney 2022)
- Stochastic Volatility: Correlation between errors

New Frontiers in Aggregation Methods

- Integrate lessons from literature on under/over reactions to information shocks (Jiang and Zhu 2017)
- Synthetic Impulse Response Functions

Synthetic Impulse Response Functions: A Proposal

- Suppose we have a multivariate time series of dimension $ptimesT$ subject to shocks from a common shock distribution
- Using an IRF estimate aggregated from the first n shocks of interest, we predict the response of variable i from variable j, $1 \leq i \leq j \leq p$.

We analyze the real-world example with Brexit included.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

Abadie, Alberto and Javier Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country". In: *American Economic Review* 93.1, pp. 113–132.

Andersen, Torben G and Luca Benzoni (2010). "Stochastic volatility". In: *CREATES Research Paper* 2010-10.

Andersen, Torben G, Tim Bollerslev, and Francis X Diebold (2007). "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility". In: *The review of economics and statistics* 89.4, pp. 701–720.

Andersen, Torben G and Timo Teräsvirta (2009). "Realized volatility". In: *Handbook of financial time series*. Springer, pp. 555–575.

Andersen, Torben G et al. (2003). "Modeling and forecasting realized