

Post-shock Volatility Forecasting Using Aggregated Shock Information Notes

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 shock event. Adapting the synthetic prediction framework of ?, we exploit series that have experienced similar shock events. 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 1-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

To be renamed

2 Main Notes

2.1 Hypotheses

1. The smaller the shock length k , the greater the need for a large shock.
2. For the donors with large weights, smaller p-values for the donor shocks will lead to smaller MSE for prediction done using convex hull.
3. The level shock ϵ^* will require a similar (α, β) structure more so than a volatility shock alone. Henceforth, let's refer to this as "homogeneity in the autoregressive structure" across donors. The hypothesis in mind here is that homogeneity matters far more for the level shock.

*davidl11@ilinois.edu

4. In the case of a pure level shock, if the shock length is just one, a convex combination of the $a_{T^*+1}^2$ may outperform all other methods. This would make sense because a GARCH model can be reparamaterized as an ARMA on the sequence a_t^2 .
5. Sensitivity analysis: take only the convex hull approach and look at all $2^p - 1$ non-empty subsets of the covariates. For the larger cardinality subsets, do we see big differences in 1) convex combs, 2) predictions?
6. The variance of the level shock needs to be small, i.e. the variation of the level shocks across the donors needs to be small, in order for the estimation of ω_i^* to work.
7. A volatility shock longer than 1 trading session should we much less necessary is $\alpha + \beta \approx 1$, i.e. if the process is slowly decaying.
8. A level shock at $T^* + 1$ may hinder estimation of w^* because its expected moments will be much larger than those of $\epsilon_{i,t}$, i.e. it will be an outlier.
9. For the prediction step, why not just take a convex combination of the $\hat{\sigma}_i^2$ in the donor pool? Reply is (a) that would require something like homogeneity in the paramaters (α, β) between the donors and the time series under study, and (b) because we have assumed knowledge of an outsize shock, we want to somehow use that information to *intervene in, or make a correction to* the GARCH model, not just use a GARCH model.
10. For \mathcal{M}_2 1 especially, the shock estimator given by the orthogonal projection of the estimated shocks into the 1-dimensional subspace spanned by X_{T^*+1} may be best.
11. If donors are independent, a non-sparse linear combination could have lowest risk.
12. If donors are highly correlated, the s-sparse linear combination eg convex hull implies $\binom{s}{2}$ covariances.
13. For the convex linear combination method to outperform competitors consistently in the context of the mcM_{21} and mcM_{22} models, it may be necessary for the variance of δ to be high compared to the variance of ϵ , or alternatively, the ratio of the two signal to noise ratios might need to be high. The idea there is that the convex linear combination method performs best when the bulk of the noise is coming from δ .

2.2 To do

1. Use a more realistic shock distribution, e.g. skewed t or skewed Laplace, HOWEVER [?] find "no evidence that the Student t likelihood improves forecasting ability despite its potentially more realistic description of return tails"
2. Use bivariate distribution for the level and volatility shocks.
3. Add an intercept to optimization process.
4. Clean up docstring in simulation file as well as signal-to-noise ratio, i.e. output.
5. Do I need a decaying shock? For vol, for level?
6. Do I need shocks in covariates?
7. At the money options less biased.
8. Use Δ^2 , i.e. second difference (of volume, RV, etc) in order to capture the change in the change

9. Emphasize that $\epsilon_{i,t}^*$ is a nuisance parameter and the length of the volatility shock and the length of the level shock are also volatility parameters.
10. What metrics will be used to compare forecasts? In simulations, we may need some way to compare forecasts of series that are not on the same scale, i.e. because some series have larger shocks. Or it could be enough to use rankings, e.g. for each simulation, just rank all K methods. In real-world applications, if the number of applications is small, ranking may not be too informative, and we may need to find a metric suitable for volatility forecast comparison.
11. Use mean absolute percentage error to judge methods. absolute percentage error allows comparisons between volatility series on different scale.
12. Use GoogleTrends data (on Δ GoogleTrends) in volatility profile
13. Look at LOOCV in Lin Eck 2021
14. Read Abadie 2010 again and look at proof of unbiasedness
15. Incorporate lessons and (a) forecasting loss functions and (b) rolling forecasts from <https://arxiv.org/abs/2203.10>
16. Read rubin on matching
17. Fix variance given that the var has correlated entries
18. Diebold Mariano
19. Dan: extended abstract ?
20. I need to use the matrix V in Abadie. it's an unused hyperparameter
21. LOOCV - find in L+E 2021
22. Must use Mahalanobis distance because of covariance
23. $2 * \text{sum}(\text{mat}[\text{upper.tri}(\text{mat})]) + \text{sum}(\text{diag}(\text{mat}))$
24. Canonical correlation analysis
25. <https://cran.r-project.org/web/packages/MultiHorizonSPA/MultiHorizonSPA.pdf>
26. Recall that my omega hats are correlated
27. Develop some theory beyond confidence intervals for adjusted predictions. Under independence (pairwise independence between donors and also between time series under study) the variance is just the sum of the variances.

2.3 Simulation notes

1. Since GARCH parameters many trading periods to stabilize [?] p.172, we simulate series with a minimum of 252 trading days (one year) and a maximum of $3*252$ trading days
2. We can just use squared return a_{i,T^*+1}^2 of the series under study as a ground truth estimator, since it is unbiased and over many examples, it should reveal the ordinal comparison between forecasting methods [?] p.181
3. For each pair of variables that I vary in the simulations, construct an MSE heat map like the one found here <https://www.r-bloggers.com/2013/06/grid-search-for-free-parameters-with-parallel-computing/>

2.4 Procedure for building a donor pool

1. Determine time series of under study (TSUS).
2. Nominate set of donors D for the TSUS, $|D| = n$.
3. Let C_i denote the set of covariates for the shock distribution of the i th time series, $1 \leq i \leq n + 1$.
4. Define $C = \cap_{i=1}^{n+1} C_i$, i.e. the intersection of the covariate sets. (Dan asks: could we also take the union and then just impute structural zeroes?)
5. Let X_i denote the vector of length C_i containing the entries of the covariate of the shock distribution of the i th time series, and let X_i^C denote the subvector of length $|C|$ with the covariates common to all $n + 1$ series.
6. For each $i \geq 2$, compute $\widehat{corr}(X_1^C, X_i^C)$.
7. Cull the donor set D down to only the donors that satisfy $|\widehat{corr}(X_1^C, X_i^C)| \geq \xi$, for some suitably chosen $\xi \in (0, 1)$

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

$$\mathcal{M}_2: \begin{aligned} \sigma_{i,t}^2 &= \omega_i + \omega_i^* D_{i,t} + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{x}_{i,t}, a_{i,t} = \sigma_{i,t} \epsilon_{i,t} \\ \omega_i^* &= \mu_{\omega^*} + \delta_i' \mathbf{x}_{i,T_i^*+1} + \tilde{\epsilon}_i, \end{aligned} \quad (1)$$

where the added random effects are

$$\begin{aligned} \omega_i^* &\stackrel{iid}{\sim} \mathcal{F}_{\omega^*} \text{ with } \mathbb{E}_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \mu_{\omega^*}, \text{Var}_{\mathcal{F}_{\omega^*}}(\omega_i^*) = \sigma_{\omega^*}^2 \\ \epsilon_i^* &\stackrel{iid}{\sim} \mathcal{F}_{\epsilon^*} \text{ with } \mathbb{E}_{\mathcal{F}_{\epsilon^*}}(\epsilon_i^*) = \mu_{\epsilon^*}, \text{Var}_{\mathcal{F}_{\epsilon^*}}(\epsilon_i^*) = \sigma_{\epsilon^*}^2 \\ \gamma_i &\stackrel{iid}{\sim} \mathcal{F}_{\gamma} \text{ with } \mathbb{E}_{\mathcal{F}_{\gamma}}(\gamma_i) = \mu_{\gamma}, \text{Var}_{\mathcal{F}_{\gamma}}(\gamma_i) = \Sigma_{\gamma}^2 \\ \delta_i &\stackrel{iid}{\sim} \mathcal{F}_{\delta} \text{ with } \mathbb{E}_{\mathcal{F}_{\delta}}(\delta_i) = \mu_{\delta}, \text{Var}_{\mathcal{F}_{\delta}}(\delta_i) = \sigma_{\delta}^2 \\ \epsilon_{i,t} &\stackrel{iid}{\sim} \mathcal{F}_{\epsilon} \text{ with } \mathbb{E}_{\mathcal{F}_{\epsilon}}(\epsilon_{i,t}) = 0, \text{Var}_{\mathcal{F}_{\epsilon}}(\epsilon_{i,t}) = 1 \\ \omega_i^* &\perp (\alpha, \beta)_i \perp \gamma_i \perp \epsilon_{i,t}. \end{aligned}$$

We further define $\widetilde{\omega}_i^* = \mu_{\omega^*} + \delta_i' \mathbf{x}_{i,T_i^*+1}$. We will investigate the post-shock aggregated estimators in \mathcal{M}_2 in settings where δ_i is either fixed or random. We let \mathcal{M}_{21} denote model \mathcal{M}_2 with $\delta_i = \delta$ for $i = 1, \dots, n + 1$, where δ is a fixed unknown parameter. We let \mathcal{M}_{22} denote model \mathcal{M}_2 with the following random effects structure for δ_i :

$$\delta_i \stackrel{iid}{\sim} \mathcal{F}_{\delta} \text{ with } \mathbb{E}_{\mathcal{F}_{\delta}}(\delta_i) = \mu_{\delta}, \text{Var}_{\mathcal{F}_{\delta}}(\delta_i) = \Sigma_{\delta} \quad \text{with} \quad \delta_i \perp \tilde{\epsilon}_i.$$

We further define the parameter sets

$$\begin{aligned} \Theta &= \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i): t = 1, \dots, T_i, i = 2, \dots, n + 1\}, \\ \Theta_1 &= \{(\eta_i, \phi_i, \theta_i, \alpha_i, \mathbf{x}_{i,t}, y_{i,t-1}, \delta_i): t = 1, \dots, T_i, i = 1\}, \end{aligned} \quad (2)$$

where Θ and Θ_1 can adapt to \mathcal{M}_1 by dropping δ_i . We assume this for notational simplicity.