НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Сравнение моделей волатильности на данных российских биржевых индексов

Аганин А.Д. Пересецкий А.А.

8 ноября 2017 г.

Цель работы

Волатильность - один из важнейших показателей на финансовых рынках. Он широко используется многими трейдерами, фондами, центрабанками. Поскольку наибольший интерес представляет будущая волатильность, широкое развитие получили модели предсказания волатильности, такие как GARCH, ARFIMA, MIDAS, HAR-RV семейства.

Цель работы: *сравнение большого набора моделей из GARCH, HAR-RV и ARFIMA семейств по качеству прогноза реализованной волатильности на один день вперед на данных 10 российских биржевых активов*

Оценки волатильности

Выделяют два основных подхода:

- Оценка волатильности по дневным данным цен закрытия:

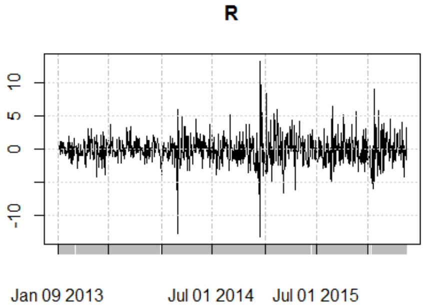
$$\sigma_t^2 = (\log(p_t) - \log(p_{t-1}))^2$$

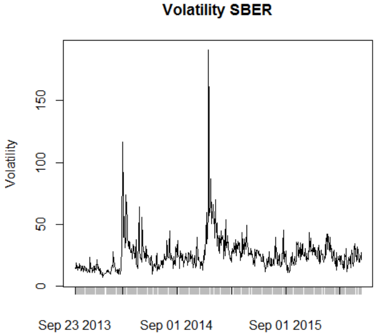
- Оценка волатильности по внутридневным данным - RV.

Пусть $p_{t,j}$ - цена актива в день t и j - номер внутридневного интервала длины Δ, где $j = 1,..N$, а N - общее число таких интервалов. Тогда $r_{t,j} = \log(p_{t,j}) - \log(p_{t,j-1})$ доходность актива на интервале $[j-1;j]$. Отсюда мы можем получить оценку волатильности RV:

$$RV_t = \sum_{j=1}^{N} r_{t,j}^2.$$

Характеристики волатильности





Первая ARCH модель была предложена Engle в 1982 году для учета гетероскедастичности в рядах волатильности.
Обозначим через p_t цену актива в момент t и доходность актива через $r_t = \log(p_t) - \log(p_{t-1})$. Тогда можно записать

$$r_t = \mu + \epsilon_t,$$

где μ детерминированная компонента. Если обозначить волатильность актива в момент t как σ_t и инновации как $\epsilon_t = \sigma_t \varepsilon_t$, то ARCH(1) модель можно записать как

$${\sigma_t}^2 = \omega + \alpha_1 \epsilon_{t-1}^2.$$

Инновации ε_t обычно считаются нормальными с параметрами $N(0{,}1)$. Bollerslev в 1986 предложил обобщенную GARCH(p,q) модель

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2.$$

Основные плюсы GARCH моделей:

- Учет кластеризации волатильности
- Толстые хвосты распределения
- Для инноваций помимо предположения о нормальном распределении возможно использование предположений о других распределениях, что позволяет учесть особенности данных.
- Учет эмпирических эффектов на рынке, таких как эффект "рычага".

Основные минусы GARCH моделей:

- Экспоненциально затухающая память
- Сложность оценивания

GARCH модели

Множество моделей

1. GARCH: $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$;
2. IGARCH: $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^2 - \varepsilon_{t-1}^2)$;
3. NAGARCH: $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$;
4. Thr.-GARCH: $\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i [(1-\gamma_i)\varepsilon_{t-i}^{+} - (1+\gamma_i)\sigma_{t-i}^{-}]^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$;
5. GJR-GARCH: $\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i + \gamma_i I_{(\varepsilon_{t-j}>0)}]\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$;
6. Taylor/Schwert: $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$;
7. EGARCH: $log(\sigma_t^2) = \omega + \sum_{i=1}^{q} [\alpha_i \varepsilon_{t-i} + \gamma_i(|\varepsilon_{t-j}| - E|\varepsilon_{t-i}|)] + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2)$;
8. NGARCH:$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}|^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$;
9. A-PARCH: $\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i [|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}]^\gamma + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$;
10. CSGARCH: $q_t = \omega + \rho q_t - 1 + \phi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$
    $\sigma_t^2 = q_t + \sum_{i=1}^{n} \alpha_i (\varepsilon_{t-i}^2 - q_{t-i}) + \sum_{i=1}^{p} \beta_j (\sigma_{t-j}^2 - q_{t-j})^2$;
11. H-GARCH or ALLGARCH: $\sigma_t^\delta = \omega + \sum_{i=1}^{n} \alpha_i \delta \sigma_{t-j}^\delta [|\varepsilon_t - k| - \tau|\varepsilon_t - k|]^\nu + \sum_{i=1}^{n} \beta_j \sigma_{t-j}^2$.

В 2003 году Andersen предложил использовать ARFIMA(p,d,q) модель для прогнозированию RV с параметрами (1,d,0) и (5,d,0).

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d(RV_t-\mu)=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t,$$

где

$$(1-L)^d=\sum_{k=0}^{\infty}\frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$$

В 2009 году Corsi предложил новый класс моделей волатильности, основываясь на гипотезе о гетерегенности рынка, сформулированной Müller.

Müller выделил несколько типов агентов на рынке в зависимости от их временных горизонтов инвестирования:

- краткосрочные инвесторы(торговля внутри дня)
- среднесрочные инвесторы(торговля в течении месяца)
- долгосрочные инвесторы(центрабанки, определенные фонды)

HAR-RV модели
Определение

Сама модель может быть записана следующим образом:

$$RV_{t+1} = \beta_0 + \beta_d RV_t^d + \beta_w RV_t^w + \beta_m RV_t^m + \epsilon_{t+1},$$

где

$$RV_t^d = RV_t,$$
$$RV_t^w = \sum_{i=1}^{5} RV_{t-i},$$
$$RV_t^m = \sum_{i=6}^{21} RV_{t-i}.$$

Мы получаем HAR-RV модель с параметрами(1,5,21).

RV можно представить как сумму устойчивой компоненты и "прыжков"

$$RV_{t+1}(\Delta) \to \int_t^{t+1} \sigma^2(s)ds + J_{t+1},$$

В качестве устойчивой компоненты в литературе по стохастической волатильности широко используется Bipower Variation, определяемая как:

$$C_{t+1} = BV_{t+1}(\Delta) = \frac{1}{E(|Z|)^2} \sum_{j=2}^{1/\Delta} |r_{i+j\Delta,\Delta}||r_{t+(j-1)\Delta,\delta}|,$$

где Z — случайная величина, нормально распределенная с параметрами N(0,1). Тогда прыжки можно определить как:

$$J_{t+1} \equiv \max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta),0].$$

Для учета эмпирических эффектов, замеченных на рынке, были предложены модификации обычной HAR-RV(1,5,21) модели:

- HAR-RV-J модель

$$RV_{t+1} = \beta_0 + \beta_d RV_t^d + \beta_w RV_t^w + \beta_m RV_t^m + \beta_j J_t + \epsilon_{t+1}.$$

- HAR-RV-CJ модель

$$RV_{t+1} = \beta_0 + \beta_{CD} C_t^d + \beta_{CW} C_t^w + \beta_{CM} C_t^m + \beta_{JD} J_t^d + \beta_{JW} J_t^w + \beta_{JM} J_t^m + \epsilon_{t+1},$$

- HAR-RV-CJ модель с "рычагом"

$$RV_{t+1} = \beta_0 + \beta_{CD} C_t^d + \beta_{CW} C_t^w + \beta_{CM} C_t^m + \beta_{JD} J_t^d + \beta_{JW} J_t^w + \beta_{JM} J_t^m$$
$$+ \gamma_d r_t^{(d)-} + \gamma_w r_t^{(w)-} + \gamma_m r_t^{(m)-} \epsilon_{t+1},$$

где $r_t^{(d)-} = \min(r_t^{(d)}, 0)$.

Model Confidence Set test 1/2

MCS тест (Hansen et al., 2011) заключается в множественном сравнении всех моделей.
Подготовка к тестированию:

1. Оценивание всех модели на движущемся окне из n наблюдений и прогнозирование волатильности на 1 день вперед для горизонта прогнозирования R.
2. Задав функцию потерь $L(\sigma_t, h_{k,t})$, где σ_t -волатильность в день t, $h_{k,t}$ - прогноз волатильности модели k на день t, можно посчитать относительные потери модели k относительно бенчмарка для всех альтернативных моделей как разность $L(\sigma_t, h_{0,t})$ - $L(\sigma_t, h_{k,t})$.

Метод сравнения
Model Confidence Set test 2/2

Нулевая гипотеза MCS теста формулируется так: "Мат. ожидание разности $L(\sigma_t, h_{0,t})$ - $L(\sigma_t, h_{k,t}) = 0$ для любых двух моделей из оптимального множества"
Итеративный поиск оптимального множества моделей выполняется так:

1. Тестируется нулевая гипотеза, рассчитывая соответствующую статистику для доверительного уровня α
2. Если нулевая гипотеза не отвергается, то процедура окончена, получено множество равных по производительности моделей. В противном случае применяется исключающее правило и модель с наихудшей производительностью исключается из множества, после чего мы переходим обратно к тестированию нулевой гипотезы для нового множества.

Hansen P., Lunde A., Nason J. (2011). The Model Confidence Set. Econometrica, 79(2), 453–497.

Были рассчитаны 88 GARCH моделей, 10 HAR-RV моделей и 4 ARFIMA модели.
Сравнение выполнялось на данных 10 активов - ALRS, GAZP, GNWK, LKOH, MICEX, MTSS, ROSN, RTSI, SBER, VTBR с 23 сентября 2013 по 12 мая 2016 (655 торговых дней).

В MCS тесте использовались функции потерь MAE, MSE и R2LOG:

$$MSE_1 \equiv \frac{1}{n} \sum_{i=1}^{n} (RV_t - h_t)^2$$

$$MAE_1 \equiv \frac{1}{n} \sum_{i=1}^{n} |RV_t - h_t|$$

$$R2LOG \equiv \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{RV_t}{h_t} \right)^2$$

Результаты сравнения

	ALROSA			GAZP			GMKN			LKOH			MICEX			MTSS			ROSN			RTSI			SBER			VTBR		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
HAR-RV	*			*			*	*	*	*	*	*							*		*	*	*					*	*	*
LOG-HAR-RV	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
HAR-RV-J							*	*																					*	
LOG-HAR-RV-J	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*		*	*	*	*	*	*	*	*	*
HAR-RV-CJ																														
LOG-HAR-RV-CJ	*			*			*	*	*	*	*	*	*	*	*	*						*	*	*	*	*	*	*	*	*
LHAR-RV-J													*	*	*															
LOG-LHAR-RV-J	*	*	*	*	*	*			*				*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*
LHAR-RV-CJ									*				*	*	*															
LOG-LHAR-RV-CJ				*	*	*	*	*	*				*	*	*	*		*	*	*	*	*	*	*		*		*	*	*

Выводы

В работе выполнено сравнение 102 моделей волатильности из трех разных семейств на данных 10 российских активов. В результате сравнения на основе MCS теста были исключены все модели семейств GARCH и ARFIMA и для всех активов конечное множество состояло исключительно из моделей HAR-RV семейства.

Полученные результаты подтверждают преимущество HAR-RV моделей на российских данных и важность учета волатильности разных рыночных компонент.