Conditionally Heteroscedastic Binary Time Series

Tim Ginker

Economics M.A. Workshop, BIU

4 Jan 2019

Outline

Background and Motivation

Opportunities

Challenges

Main MA thesis results

Probit model with GARCH effects

Robustness of binary choice models to conditional heteroscedasticity

SPY daily returns distributional properties

Mean	Min	Max	Std. dev.	Skew	Kurt	J.B
0.035	-6.73	4.93	0.94	-0.48	4.93	1946.39

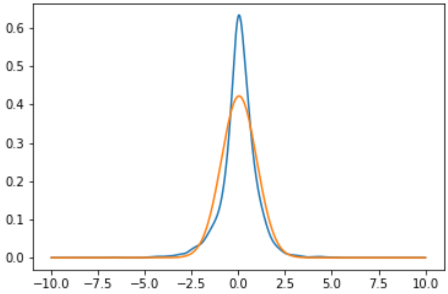
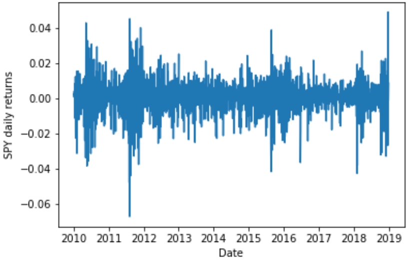


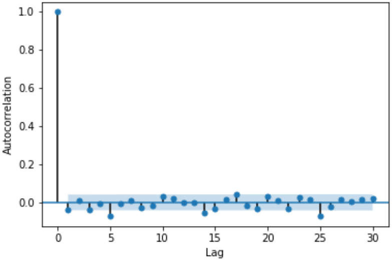
Figure: SPY daily returns KDE (blue) against estimated Normal density (brown).

The results correspond to the period from Jan 2010 to Jan 2019.

SPY daily returns time series properties

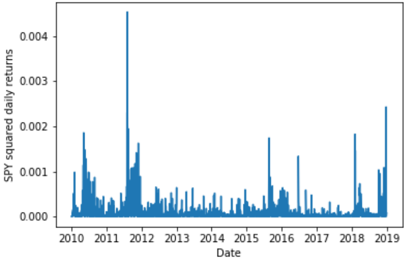


(a) SPY daily returns

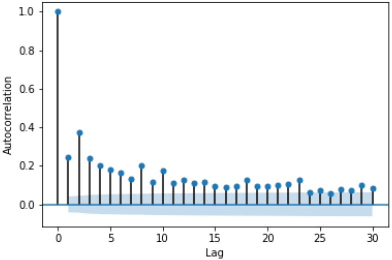


(b) SPY daily returns ACF

SPY daily returns time series properties cont'd



(a) SPY squared daily returns



(b) SPY squared daily returns ACF

SPY daily returns time series properties cont'd

We have observed two sets of phenomena:

- Returns approximate conditional mean independence.
  This finding corroborates the well known Efficient market hypothesis which suggests that returns should not be easy to forecast based on readily available information. It can be stated that also last four decades of empirical work majorly show that the data support the theory(e.g., Fama 1970, 1991).
- Dependence (and hence forcastability) in asset returns volatility
  There is a huge literature that documents the notable dependence, and hence forecastability, of asset return volatility, with important implications not only for asset allocation, but also for asset pricing and risk management. For review see Bollerslev et al. (1992), Ghysels et al. (1996) Franses and van Dijk (2000), and Andersen et al. (2005).

Sign predictability

The above findings, among other things, show that even though the returns are **uncorrelated** over time, they are not **independent**.

Thus, could there be other (than linear) functions of returns that can be forecastable? Indeed, we have already seen that in the example of squared returns.

What about index functions?

Sign predictability

Correct forecasts of a price change direction can be useful because they may lead to profitable speculative positions and correct hedging decisions. Thus, although binary time series models, for instance, seem to be less informative, compared with continuous response models, they are still economically meaningful. There is a number

of studies showing the superiority of the qualitative response models over the continuous ones. Both, in terms of predictive performance and the ability to generate higher trading profits (see for example Leung et. al. (2000) and Nyberg (2011)).

Sign predictability

Binary models were proven to be useful as a part of decomposition models.

Rydberg and Shephard (2003) specified the stock return as a product of two binary variables, defining the returns direction and market activity, multiplied by a process of positive integers which defines the size of a price change.

Anatolyev and Gospodinov (2010) expressed the financial asset return as a product of its sign and its absolute value. These components were modeled separately before the joint forecast was constructed.

Sign predictability

We denote by r_t the stock return at time t, and assume that it can be represented as

$$r_t = \mu_t + \varepsilon_t,$$

where $\mu_t = E[r_t \mid \mathfrak{F}_{t-1}]$, ε_t is conditionally normally distributed

$$\varepsilon_t \mid \mathfrak{F}_{t-1} \sim N(0, \sigma_t^2),$$

σ_t^2 is the variance of ε_t conditional on $\mathfrak{F}_{t-1}$, and $\mathfrak{F}_t$ represents an increasing family of $\sigma -$ *fields* generated by ε_t.

Sign predictability

Let

$$y_t = 1\{r_t > 0\}, \;\; t = 1, 2, \ldots n$$

be an indicator function taking the value of 1 if the condition in the brackets is satisfied and 0 otherwise. Under these assumptions the conditional probability of a positive return is given by

$$E[y_t \mid \mathfrak{F}_{t-1}] = \Pr[r_t > 0 \mid \mathfrak{F}_{t-1}] = \Pr[\mu_t + \varepsilon_t > 0 \mid \mathfrak{F}_{t-1}]$$
$$= \Pr[\varepsilon_t > -\mu_t \mid \mathfrak{F}_{t-1}]$$
$$= \Pr[\frac{\varepsilon_t}{\sigma_t} > -\frac{\mu_t}{\sigma_t} \mid \mathfrak{F}_{t-1}] = \Phi[\frac{\mu_t}{\sigma_t}]$$

Sign predictability

Thus, r_t exhibits sign dependence (or sign forecastability) if y_t has a conditionally varying mean $(E[y_t \mid \mathfrak{F}_{t-1}])$(Christoffersen and Diebold, 2006). Under these assumptions, the sign of return can be predictable as long as $\mu_t \neq 0$ and if μ_t or σ_t vary over time.

Specification error

Let

$$y_t = 1\{x_t'\gamma + \varepsilon_t > 0\},\ \ t = 1, ..., n,$$

where $1\{\cdot\}$ is the indicator function which takes the value of unity if the condition in the brackets is satisfied and zero, otherwise, x_t is a $K \times 1$ vector of explanatory variables, γ is a $K \times 1$ vector of unknown parameters, and for all t and s, conditional on $\mathcal{F}_{t-1}$ and x_s, $\varepsilon_t \sim F(0, \sigma_t^2)$, where $\mathcal{F}_t$ is the increasing sequence of σ-fields generated by $\{\varepsilon_j\}_{j=1}^{t}$ and F is a symmetric CDF on which conditions are given in Assumption A below. When F is a Standard Normal CDF and $\sigma_t = 1$ for every t. The log-likelihood function of a standard probit model is given by

$$l_n^P(\gamma) = \sum_{t=1}^{n} l_t^P(\gamma) = \sum_{t=1}^{n} (y_t \ln[\Phi(x_t'\gamma)] + (1 - y_t)\ln[1 - \Phi(x_t'\gamma)]).$$

Specification error

If $\sigma_t \neq 1 \ \forall t$, we will mistakenly estimate the (misspecified) model with

$$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$$

instead of

$$X^* = \begin{bmatrix} \frac{x_1'}{\sigma_1} \\ \vdots \\ \frac{x_n'}{\sigma_n} \end{bmatrix}.$$

Specification error

In linear models heteroscedasticity affects the efficiency of the least squares estimator (OLS). OLS-based predictions remain consistent but inefficient (see Green (2012), pp. 257-258 for general forms of heteroscedasticity and Engle (1982) for ARCH effects). In contrast, the probit MLE is inconsistent in the presence of any form of heteroscedasticity (Green (2012), p. 693). This implies that even if we use a technique that provides an efficient asymptotic covariance matrix for the misspecified probit model, while its estimators remain biased in the **unknown** direction, they will be meaningless in statistical inference and forecasting (Freedman, 2006). Yatchew and Griliches (1985) developed an approximation for the probit MLE bias in the presence of heteroscedasticity and showed that it is inconsistent and inefficient.

Probit model with GARCH effects

One of the most popular approaches to modeling heteroscedasticity in financial time series is the GARCH family of models (Bollerslev, 1986). Let v_t be a sequence of independently and identically distributed random variables with zero mean and unit variance. The GARCH models are of the form

$$\varepsilon_t = v_t \sqrt{h_t}, \qquad ((1))$$

where

$$h_t = E(\varepsilon_t^2 | \mathfrak{F}_{t-1}) = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \ \ t \in \mathbb{Z}.$$

The non-negativity of h_t is guaranteed if

$$\omega > 0, \alpha_i \geq 0, \beta_j \geq 0. \qquad ((2))$$

Probit model with GARCH effects

The process is second order stationary (Bollerslev (1986), Theorem 1) if, in addition to (2),

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \qquad ((3))$$

While the conditional variance of ε_t, vary with the information set $\mathfrak{F}_{t-1}$, under the restrictions (2) and (3), its unconditional variance is $\sigma_\varepsilon^2 = \frac{\omega}{1-(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j)}$. The process in (1) is called $GARCH(p, q)$ process, where p is the number of squares of the past innovations and q is the number of lags of h_t included in the volatility function.

Probit model with GARCH effects

Define

$$w_t(\theta) = \frac{x_t'\gamma}{\sigma_t}.$$

The correctly specified conditional likelihood and log-likelihood functions of $\{y_t\}$ will be given

$$L_n(\theta) = \prod_{t=1}^{n} \Phi[w_t(\theta)]^{y_t} \Phi[-w_t(\theta)]^{1-y_t},$$

$$l_n(\theta) = \sum_{t=1}^{n} (y_t \ln(\Phi[w_t(\theta)]) + (1 - y_t) \ln(\Phi[-w_t(\theta)])),$$

Probit model with GARCH effects

Assumption C

1. x_t is a sequence of IID random variables with up to order 4 finite even moments and non-singular $E_{\theta_0}[x_t x_t']$.

2. The r_t and y_t processes are as defined in above, with $p = P$ and $q = 0$ $(ARCH(P))$ or $p = 1$ and $q = 1$ $(GARCH(1,1))$.

3. θ is an element of the interior of a convex set Θ.

Assumption D

1. $A(\theta_0) = p\lim E[\frac{1}{n}\frac{\partial^2 l_n(\theta)}{\partial\theta\partial\theta'}|_{\theta=\hat{\theta}_n}]$ is invertible.

Probit model with GARCH effects

Lemma 1. *Under Assumption C*, $E[l_n(\theta)]$ *has a unique maximum at the true parameters value* $\theta_0 \in \Theta$.

Theorem 3

Under Assumption C, $\hat{\theta}_n \xrightarrow{p} \theta_0$, *as* $n \longrightarrow \infty$.

Theorem 4

Under Assumptions C and D, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B(\theta_0)^{-1})$.

By $\mathcal{F}_{t-1}$ we denote the σ-field generated by $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$. We shall make the following assumption.

Assumption A (Ginker and Lieberman, 2017)

1. The data generating function is given by $y_t = 1\{x_t'\gamma + \varepsilon_t > 0\}$.

2. For all t and s, conditional on $\mathcal{F}_{t-1}$ and x_s, ε_t has a zero mean, conditional variance σ_t^2, $0 < \sigma_t < \infty$ and cumulative distribution function (CDF) F. The CDF F is smooth and strictly monotonic with a bounded density f which has $\mathbb{R}$ as its support and which is symmetric. In addition, $F(\nu)$ is concave for $\nu > 0$.

3. The true parameter vector γ_0 is an element of the interior of a convex parameter space $\Gamma \subset \mathbb{R}^K$.

4. The $K \times 1$ vector x_t is finite, strictly stationary and ergodic, and is not contained in any linear subspace of $\mathbb{R}^K$, $\forall t$.

5. The process $\{\sigma_t\}$ is strictly stationary and ergodic and independent of x_s, for all t and s.

We remark that Assumption A(2) holds for the normal and logistic distributions. The misspecified log-likelihood function in which conditional heteroscedasticity is ignored and σ_t is set to unity $\forall t$, is given by

$$\tilde{l}_n(\gamma) = \sum_{t=1}^{n} \tilde{l}_t(\gamma),$$

where

$$\tilde{l}_t(\gamma) = y_t \log\left(F^*(x_t'\gamma)\right) + (1 - y_t)\log\left(1 - F^*(x_t'\gamma)\right) \qquad (1)$$

and F^* is the CDF of a random variable with a CDF F after it has been normalized to have mean zero and unit variance. Similarly, by f^* we denote the density corresponding to F^*. Let $\tilde{\gamma}_n = \arg\max_\Gamma \tilde{l}_n(\gamma)$. To emphasize, $\tilde{\gamma}_n$ is the maximizer of a misspecified log-likelihood function. By E_{γ_0} we denote an expectation taken under the true parameter value. The main result follows below.

Theorem 1 (Ginker Lieberman,2017)

Under Assumption A, there exists a finite and positive ρ which satisfies

$$0 < \frac{1}{E_{\gamma_0}\left(\sigma_t\right)} < \rho < E_{\gamma_0}\left(\frac{1}{\sigma_t}\right) < \infty,$$

such that $\tilde{\gamma}_n \xrightarrow{p} \rho\gamma_0$.

The MLE of the correct likelihood, say, $\hat{\gamma}_n$, is consistent. If, given $x_t = x$, the researcher wishes to base the predictive value, $\hat{y}_t$, of y_t, according to the rule $\hat{y}_t = 1\{F^*(x'\hat{\gamma}_n/\sigma_t) > 0.5\}$, then for large n, the rule is equivalent to $1\left\{x'\gamma_0 > 0\right\}$. Basing the prediction on $\tilde{\gamma}_n$ instead does not affect the result because for large n it is tantamount to

$$1\{\rho x'\gamma_0 > 0\} = 1\left\{x'\gamma_0 > 0\right\}.$$

In other words, the misspecified MLE-based prediction remains unaltered even though $\tilde{\gamma}_n$ is inconsistent. This result corroborates some of the simulation results of Munizaga *et. al.* (2000).

When the classical assumptions including homoscedasticity hold and the usual normalization, $\sigma_t = 1\ \forall t$, is imposed, it follows from Theorem 1 that $\rho = 1$, i.e., that the MLE based on the correct specification is consistent. Therefore, Theorem 1 is a generalization of the standard result. Moreover, for any predetermined $\mu \in \mathbb{R}$, if x_t includes an intercept, the model $y_t = 1\{x_t'\gamma + \varepsilon_t > \mu\}$ is equivalent to

$$y_t = 1\{x_t'\gamma^* + \varepsilon_t > 0\},$$

where $\gamma_1^* = \gamma_1 - \mu$ and $\gamma_j^* = \gamma_j$, $j = 2, ..., K$, in which case, with the obvious use of notation, we shall obtain $\tilde{\gamma}_n^* \xrightarrow{p} \rho\gamma_0^*$. Therefore, we can set without loss of generality $\mu = 0$.

Theorem 2

Under Assumption A: (i) $A_n(\tilde{\gamma}_n)$ converges in probability to a finite and nonsingular matrix

$$A(\gamma^*) = -E_{\gamma_0}\left(\left.\frac{\partial^2 \tilde{l}_t(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\gamma^*}\right).$$

(ii) The limit

$$B(\gamma^*) = E_{\gamma_0}\left(\left.\frac{1}{n}\frac{\partial \tilde{l}_n(\gamma)}{\partial\gamma}\frac{\partial \tilde{l}_n(\gamma)}{\partial\gamma'}\right|_{\gamma=\gamma^*}\right)$$

of $B_n(\gamma^*)$ exists and is equal to $A(\gamma^*)$ and

$$\sqrt{n}\left.\frac{\partial \tilde{l}_n(\gamma)}{\partial\gamma}\right|_{\gamma=\gamma^*} \xrightarrow{d} N(0, B(\gamma^*)).$$

(iii) $\sqrt{n}(\tilde{\gamma}_n - \gamma^*) \xrightarrow{d} N(0, A^{-1}(\gamma^*))$.