Return predictability and risk management

Nour Meddahi* Mamiko Yamashita†

February 14, 2018

Preliminary and incomplete draft

Abstract

When various risk measures are computed, it is often assumed that the conditional mean of an asset return is constant. However, it is well documented that the predictability of returns increases as the horizon of prediction increases. This paper assesses the impact of ignoring such possible predictability of returns on computing risk measures, especially Value-at-Risk(VaR). For this purpose, we study the term structure of VaR when the conditional mean of returns is actually time-varying, and when one assumes it to be time-varying and constant. First we compute VaR analytically when one knows parameter values, and show that the impact of ignoring time-variability of the conditional mean is non-negligible. Simulation studies show that, when one has a parameter uncertainty, estimating a model with time-varying conditional mean yields VaR that is closer to the true VaR, even though a model with constant conditional mean is often times not statistically rejected. In the empirical studies, we estimate a GARCH-in-Mean model which has a time-varying conditional mean and a GARCH model with constant conditional mean. We compare their predictive ability by Diebold-Mariano test and show that the GARCH-in-Mean model outperforms GARCH model for horizons over 10 days.

1 Introduction

Good risk management is nowadays a major concern for regulators and financial institutions. Many financial institutions compute every day the Value-at-Risk(VaR) of their units of trading in order to assess their risk exposures. In the framework of Basel II Accord, the regulator decides the capital requirement based on their 10-day VaR at 1% level, and now Expected Shortfall (ES) draws more attention in Basel III Accord. There are other risk measures, such as CoVaR (Adrian and Brunnermeier, 2011) and SRISK (Brownlees and Engle, 2017) which are computed.

The traditional way to compute these measures is to model the distribution of the return of time $t+1$ given the information available at time t (Christoffersen, 2012), and to compute one-step and multi-step ahead VaR by resampling the residual with an assumption that they are i.i.d.[1]. In modeling the conditional distribution of a return, its conditional variance is commonly assumed to be time-varying using GARCH or other specifications. On the other hand, its conditional mean is often assumed to be constant because of the difficulty of detecting its variability at the daily level.

*Toulouse School of Economics, e-mail:nour.meddahi@tse-fr.eu

†Corresponding author: Toulouse School of Economics, e-mail: yamashita.mamiko@tse-fr.eu

[1] See for instance Engle (2011).

However, it is well known that when one increases the horizon of prediction, the predictability of asset returns increases; a popular predictor is the dividend yield (dividend-price ratio).[2] Consequently, computing VaR at long horizons by assuming the constant mean could be problematic. The main goal of this paper is to assess whether ignoring possible long horizon predictability of the returns has an impact on computing VaR.

For this purpose, we start the analysis by considering a return model where the expected return is driven by a persistent autoregressive process of order one. The daily (realized) return equals the expected return plus a noise, which is assumed to have a large variance relative to the variance of the expected return, homeskedastic and being Normally distributed. The large variance assumption is realistic and makes that the return follows an ARMA(1,1) process, whose autoregressive root is very close to the moving-average one, making the return look like a white noise. Such approach is popular in the asset pricing literature at the monthly or quarterly frequencies.[3] The homoscedasticity and Normality assumptions are made in order to be able to compute VaR analytically at any horizon. The homoscedasticity assumption is relaxed later.

For this model, we first compute VaR by knowing the true model and by observing the state variable. These assumptions are unrealistic but we use these VaR as the benchmark and call the "oracle VaR". We then consider two other models. The first one corresponds to the i.i.d. case with normal shocks, where an agent has a misspecified model in mind. The second one considers the ARMA (1,1) representation implied by the model where an agent has the true model but does not observe the state variable. ARMA(1,1) representation allows an agent to get a proxy of the state variable using the past value of returns. In a first step, we take the parameters as known and compute analytically VaR. At very short horizons, we find that the three specifications lead to similar VaR. However, when we increase the horizon, including the 10-days used by regulators, we find differences. In particular, the VaR implied by the ARMA is quite close to the oracle one. However, the VaR implied by the i.i.d. model is quite different from the oracle one when the state variable is far from its unconditional mean. In addition, the difference increases with the horizon. The intuition is simple. When the state is bad, the i.i.d. model will underestimate the VaR because it will ignore that the future states are also bad due to the high persistence and it is bad from the perspective of the regulator. On the other hand, when the state is good, the i.i.d. model will over-estimate the VaR, which is bad from the perspective of the financial institution. In addition, when the state variable is far from its unconditional mean, the future values of the state variable will mean-revert. However the i.i.d. model will ignore this mean reversion and will provide the same VaR whatever the state variable.

The above analyses assume that the parameters are known, which is unrealistic. We therefore make a simulations-based analysis to assess the impact of parameter uncertainty. We simulate a realistic model and then estimated an i.i.d. and ARMA (1,1) model. An interesting result is that in the simulations, the confidence interval of the autoregressive coefficient is often times so wide as to include the value zero, which means the i.i.d. assumption is not rejected. However, when one computes the VaR implied by the two estimated models, they are close to the oracle one when the horizon is short, but the i.i.d. starts to be different from the oracle one when the horizon increases, including the 10-days-ahead case. In contrast, the VaR computed with the ARMA (1,1) model is closed to the oracle one, even if the ARMA(1,1) parameters are not significantly different from those of the i.i.d. model. In addition, the difference between the two models increases with horizon.

[2] See Fama and French (1988). Other predictors include interest rates and other macro variables. Ludvigson and Ng (2007) extracts the information from a large set of macro and financial variables using the principal components following Stock and Watson (2002).

[3] See for instance Campbell (2001), Barberis (2000) and Pástor and Stambaugh (2012).

Then we consider a model with time-varying volatility. We consider a Heston and Nandi (2000) model because it is affine and allows us to compute the VaR at any horizon by a numerical method as in Duffie, Pan, and Singleton (2000) and Duffie and Pan (2001). We do the same analysis as for the first model, that is, we consider a model with a constant mean (which we call GARCH model) and another model with time-varying mean (GARCH-in-Mean model). We make the analysis when one knows the parameter values, and then conduct simulation analysis to take into account the parameter uncertainty. The results obtained are similar to the constant volatility model. It is better to assume that the mean is time-varying, even if it is not statistically different from zero, when one computes the VaR at a long horizon.

Finally we consider a model which takes into account Realized Variance (RV). We follow the model proposed by Christoffersen, Feunou, Jacobs, and Meddahi (2014) which is an extension of Heston-Nandi GARCH model. It is also affine and thus we compute VaR analytically. We again confirm the result that assuming the time-varying conditional mean is better to compute VaR especially at long horizon.

Then we provide an empirical analysis with series of SPDR S&P 500 Index of about 2500 observations. We estimate a model with time-varying conditional mean (GARCH-in-Mean model) and another model with constant conditional mean (GARCH model) proposed by Heston and Nandi (2000) and compute VaR of horizons of 10 days, 22 days and 66 days. By Diebold-Mariano test, we reject the null hypothesis that the two models have equal predictive ability for horizons over 10 days, and the result is in favor of GARCH-in-Mean model.

The paper is incomplete. We would like to make a robust control analysis (Sargent and Hansen, 2001) in order to take into account statistically parameter uncertainty.

2 Normal and constant volatility model

Suppose an agent is interested in VaR of his portfolio. For simplicity, he possesses a diversified portfolio that can be approximated as aggregate assets such as S&P 500 Index. Let P_t be the price of the asset at day t and $R_{t+\tau|t} = \frac{P_{t+\tau}-P_t}{P_t}$ be the return of holding this asset from day t until day $t+\tau$. The τ day ahead α-level VaR is defined as

$$Pr(R_{t+\tau|t} \leq -VaR|I_t) = \alpha$$

and we focus on the 1% VaR throughout this paper.[4]

In order to compute the VaR, we introduce a model of its log returns, denoted by y_t.

2.1 Data Generating Process

2.1.1 Model

Let us define $y_t = \log P_t - \log P_{t-1}$. Let us assume that y_t and its conditional mean, μ_t follow a VAR(1) process:

Correct model ($\mathbf{M_0}$)

$$y_t = \mu_{t-1} + u_t \quad (1)$$

$$\mu_t = c + \theta\mu_{t-1} + w_t \quad (2)$$

$$\begin{pmatrix} u_t \\ w_t \end{pmatrix} \sim iid.\ N\left(0, \begin{pmatrix} \sigma_u^2, & \sigma_{uw} \\ \sigma_{uw}, & \sigma_w^2 \end{pmatrix}\right) \quad (3)$$

[4]Since in Basel II, the banks are required to compute 1% VaR.

This is a state space model and widely applied to the dynamics of the asset returns, such as Campbell (2001), Barberis (2000) and Pástor and Stambaugh (2012). The state variable μ_t is usually assumed to be unobservable to the agent. We denote by $\overline{\mu}$ and σ_μ as the mean and variance of μ_t.

The τ-day ahead VaR is determined by the distribution of the future log returns denoted by $y_{t+1:t+\tau}$:

$$y_{t+1:t+\tau} = \log P_{t+\tau} - \log P_t = \sum_{j=1}^{\tau} y_{t+j}.$$

In the following, we consider three different value of VaR computed by three different agents: (i) an agent equipped with a misspecified model and assumes a constant conditional mean, (ii) an agent with the correct model and observes μ_t, and (iii) an agent with the correct model but does not observe μ_t. For all agents, the conditional distribution of $y_{t+1:t+\tau}$ becomes Normal, but the conditional mean and the conditional variance are not always the same due to the difference in their beliefs in the model and their information set. In any case, because of the normality, they can compute VaR by the formula below:

$$VaR(R_{t+\tau|t}|M, I_t) = 1 - \exp\left(E(y_{t+1:t+\tau}|M, I_t) - 2.33\sqrt{Var(y_{t+1:t+\tau}|M, I_t)}\right)$$

where M denotes the model which the agent assumes, and I_t denotes the information set of the agent at date t.

2.1.2 Parameter values

We calibrate the model by matching the first and second moments of y_t with the data as well as the estimation results of Pástor and Stambaugh (2012) (PS(2012) henceforth). The data are log returns of S&P 500 daily index, 1990/01/02-2015/12/31 with 6553 observations. The key parameters are (i) θ, (ii) $\frac{\sigma_\mu^2}{\sigma_y^2}$ which we denote by R^2, and (iii) σ_{uw}. Firstly θ governs how persistent μ_t is. Secondly R^2 decides the impact of μ_t on the variation of y_t. Finally σ_{uw} governs the correlation between the two error terms, which is likely to be negative according to Pástor and Stambaugh (2009).

PS(2012) shows the estimation results in Figure 4 in their paper. According to it, the posterior distribution of $Corr(u, w)$ and θ are at its peak at around $Corr(u, w) = -0.9$ and $\theta = 0.9$.[5] The posterior distribution of R^2 is sensitive to the prior. Different three prior leads to different modes at $R^2 = 0.06$ or around 0.10.

It is straightforward to connect the daily model that we consider and the annual model in PS(2012). Let us consider the $h-$day aggregated process, $y_t^{(h)}$, i.e., $y_i^{(h)} \equiv y_{(i-1)h+1} + y_{(i-1)h+2} + \cdots + y_{ih} = \sum_{j=0}^{h-1} y_{ih-j}$. Then, there exists $\mu_i^{(h)}$ such that $(y_i^{(h)}, \mu_i^{(h)})$ follows a VAR(1) process:

$$\begin{aligned} y_i^{(h)} &= \mu_{i-1}^{(h)} + u_i^{(h)} \\ \mu_i^{(h)} &= c^{(h)} + \theta^h \mu_{i-1}^{(h)} + w_i^{(h)} \\ \begin{pmatrix} u_i^{(h)} \\ w_i^{(h)} \end{pmatrix} &\sim iid.\ N\left(0, \begin{pmatrix} \sigma_u^{2(h)}, & \sigma_{uw}^{(h)} \\ \sigma_{uw}^{(h)}, & \sigma_w^{2(h)} \end{pmatrix}\right) \end{aligned}$$

where all the parameters and variables are expressed as those in the daily model.[6]

[5] See Pástor and Stambaugh (2009) for the discussion why $Corr(u, w)$ is likely to be negative.

[6] $c^{(h)} = (1-\theta^h)\overline{\mu}h$. $u_i^{(h)} = \frac{1}{1-\theta}\sum_{j=1}^{h-1}(1-\theta^j)w_{ih-j} + \sum_{j=0}^{h-1} u_{ih-j}$ and $w_i^{(t)} = \frac{1-\theta^t}{1-\theta}\sum_{j=0}^{i-1}\theta^j w_{ih-j}$. Also, the other components of variance-covariance matrix are the following: $\sigma_w^{2(h)} = \left(\frac{1-\theta^h}{1-\theta}\right)^2\left(\frac{1-\theta^{2h}}{1-\theta^2}\right)\sigma_w^2$ and $\sigma_{uw}^{(h)} = \left(\frac{1-\theta^h}{1-\theta}\right)^2((\theta-\theta^h)\sigma_\mu^2 + \sigma_{uw})$. The variance of u_t is $\sigma_u^{2(h)} = V_{\tau,full}(h)$

Especially, the AR coefficient of aggregated μ_t is θ^h and the variance of aggregated conditional mean is

$$\sigma_\mu^{2(h)} = \left(\frac{1-\theta^h}{1-\theta}\right)^2 \sigma_\mu^2$$

which enables us to connect the estimation of PS(2012) and our calibration. We first pick $\theta = 0.999$ (so that its counterpart of annually aggregated return is $\theta^{250} = 0.78$) and the annual $R^2 = 0.06$ (one of the posterior mode of PS(2012)).

The calibrated parameter values are summarized in Table 1.

Table 1: Calibrated parameter values

Parameter	c	θ	σ_μ^2	σ_u^2	σ_{uw}
	2.678e-7	0.999	1.290e-4	1.937e-7	-4.000e-6

2.2 Case 1: Agent with misspecified i.i.d. model

First, as a benchmark, let us consider when an agent *wrongly* assumes that y_t is i.i.d..

Misspecified iid model ($\mathbf{M_{iid}}$)

$$y_t = \overline{\mu} + \sigma_y \widetilde{u}_t, \tag{4}$$

$$\widetilde{u}_t \sim i.i.d.\ N(0,1) \tag{5}$$

The key difference between the correct model and the misspecified iid model is that the agent believes that the conditional mean does not vary over time. Other features of the correct model, such as the Normality of the innovation and the constant volatility remain the same. In addition, the agent knows correctly the unconditional mean and variance of y_t, i.e., $\overline{\mu}$ and σ_y^2. Therefore, the agent is wrong only by missing the time-variation of the conditional mean.

Based on M_{iid}, the agent derives the "wrong" conditional distribution of $y_{t+1:t+\tau}$ as follows:

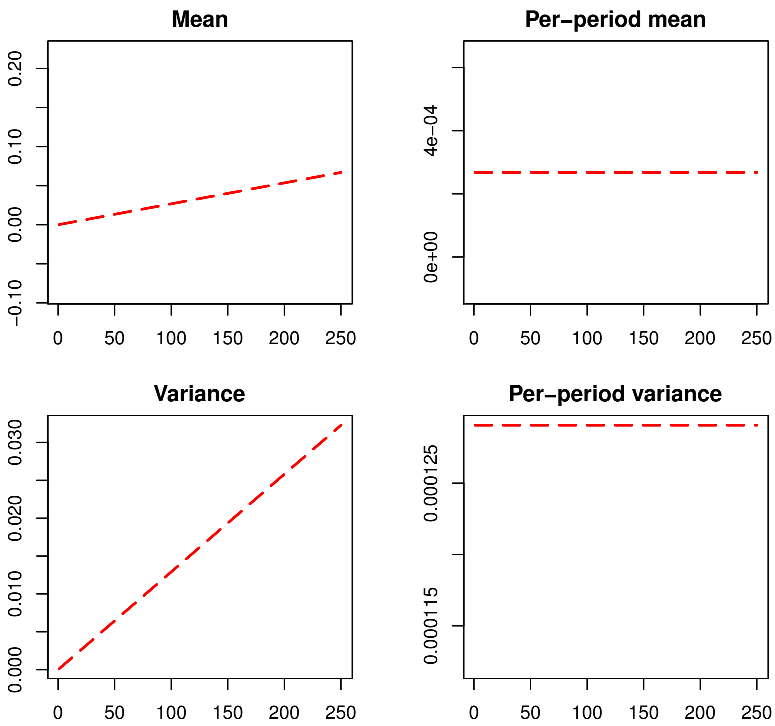
$$y_{t+1:t+\tau}|M_{iid} \sim N\left(\overline{\mu}\tau,\ \sigma_y^2\tau\right) \tag{6}$$

Here, the conditional mean and the conditional variance of $y_{t+1:t+\tau}$ are both linear in τ, and they are shown in the left panels of Figure 1. The right panels show the per-period mean and variance, i.e., we divide the conditional mean and variance by τ. The per-period mean and variance are both constant.

The "iid VaR" is

$$VaR(R_{t+\tau|t}|M_{iid}) = 1 - \exp\left(\mu_t\tau - 2.33\sqrt{\sigma_y^2\tau}\right)$$

Figure 1: Conditional mean and variance of $y_{t+1:t+\tau}$ assuming M_{iid}



2.3 Case 2: Agent with correct model with observable μ_t (Oracle)

Now let us consider an agent who has the correct model in mind (M_0) and who observes the conditional expected return or who has a perfect predictor of it. We call this as the "oracle" case and denote by $I_t^{Full} \equiv \{\underline{\mu_t}, \underline{y_t}\}$. Then

$$y_{t+1:t+\tau}|I_t^{Full} \sim N(E_{\tau,oracle},\ V_{\tau,oracle}) \tag{7}$$

where

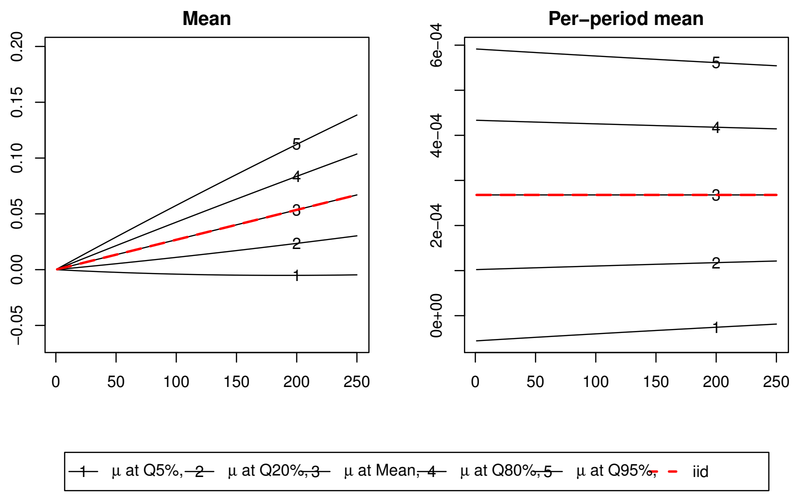
$$\begin{aligned}
E_{\tau,oracle} &= \overline{\mu}\tau + (\mu_t - \overline{\mu})P_\tau \\
P_\tau &= \frac{1-\theta^\tau}{1-\theta} \\
V_{\tau,oracle} &= \tau\left(\sigma_u^2 + \frac{2\sigma_{uw}}{1-\theta}A_\tau + \frac{\sigma_w^2}{(1-\theta)^2}B_\tau\right) \\
\text{where } A_\tau &= 1 - \frac{1-\theta^\tau}{\tau(1-\theta)}, \\
B_\tau &= 1 - \frac{2(1-\theta^\tau)}{\tau(1-\theta)} + \frac{1-\theta^{2\tau}}{\tau(1-\theta^2)}
\end{aligned}$$

The mean is decomposed into two parts: the first term is $\overline{\mu}\tau$ which is the unconditional mean of $y_{t+1:t+\tau}$ and it is identical to the mean of $y_{t+1:t+\tau}$ with M_{iid}. The second term, $(\mu_t - \overline{\mu})P_\tau$ depends on the realization of μ_t. Since $P_\tau > 0$ for all τ, $E(y_{t+1:t+\tau}|I_t^{Full}) > \overline{\mu}\tau$ if and only if $\mu_t > \overline{\mu}$, and thus the deviation of

$E(y_{t+1:t+\tau}|I_t^{Full})$ from $\overline{\mu}\tau$ is caused only by a deviation of μ_t from $\overline{\mu}$. Also, from $\frac{dP_\tau}{d\tau} > 0$, the deviation of $E(y_{t+1:t+\tau}|I_t^{Full})$ increases as τ increases.

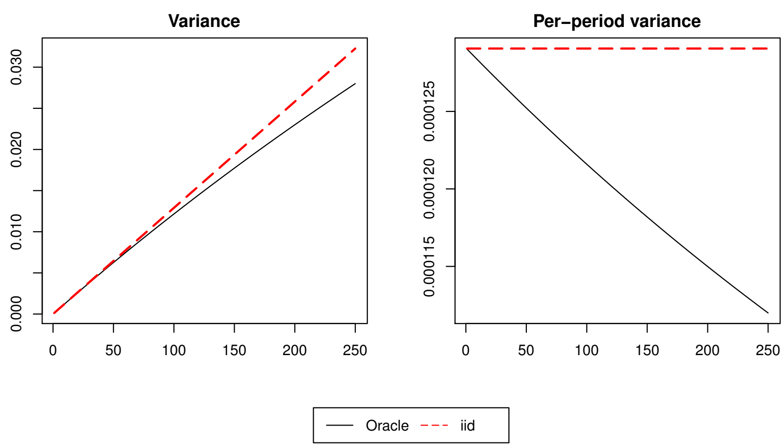
Figure 2 shows the conditional mean of $y_{t+1:t+\tau}$ given realization of μ_t.

Figure 2: Conditional mean of $y_{t+1:t+\tau}$ for Oracle case ($E_{\tau,oracle}$)



Now let us look at the variance. First, the three terms in $V_{\tau,oracle}(\tau)$ are called "i.i.d. uncertainty", "mean reversion", and "future $\mu_{t+j}(j = 1, \cdots \tau - 1)$ uncertainty" by Pástor and Stambaugh (2012).

Figure 3: Conditional variance of $y_{t+1:t+\tau}$ for oracle: $V_{\tau,oracle}$



Finally the "oracle VaR" is given by

$$VaR(R_{t+\tau}|I_t^{Full}) = 1 - \exp\left(E_{\tau,oracle} - 2.33\sqrt{V_{\tau,oracle}}\right)$$

2.4 Case 3: Agent with correct model with unobserved μ_t: ARMA(1,1) approach

Now let us consider an agent who has the correct model (M_0) but who does not observe $\{\mu_t\}$. One of the ways to compute VaR is to use a proxy of μ_t using some information available. Here, we consider the case where an agent extracts the information from the past log returns denoted by $\underline{y_t} = \{y_t, y_{t-1}, y_{t-2}, \cdots\}$. Meddahi (2002) shows how this can be done by introducing an ARMA(1,1) representation.

Meddahi (2002) shows that, if y_t follows the correct model M_0, as shown in equations 1 to 3, there exists a real number γ and a white noise process $\{\eta_t\}$ with variance σ_η^2 that satisfies the below equations:[7]

ARMA(1, 1) representation

$$y_t = c + \theta y_{t-1} + \eta_t - \gamma\eta_{t-1} \tag{8}$$

$$\eta_t \sim i.i.d.\ N(0, \sigma_\eta^2) \tag{9}$$

where the conditional mean $c + \theta y_t - \gamma\eta_t$ is in the information set $\underline{y_t}$. [8] Defining it as m_t, we have

$$\begin{aligned} y_t &= m_{t-1} + \eta_t \\ m_t &= \frac{c}{1-\gamma} + (\theta - \gamma)\sum_{j=0}^{\infty}\gamma^j y_{t-j} \end{aligned}$$

At time t, the agent has the information $\underline{y_t}$ and thus can compute m_t as a proxy of μ_t. It is straightforward to show that m_t is Normally distributed with mean $\overline{\mu}$ and variance σ_m^2, with $0 \le \sigma_m^2 \le \sigma_\mu^2$.[9] Let π_m be the fraction of the variance, $\pi_m = \frac{\sigma_m^2}{\sigma_\mu^2}$ with $0 \le \pi_m \le 1$. π_m indicates a degree of information content of $\underline{y_t}$ on μ_t since the conditional distribution of m_t given the realization of μ_t is given by

$$m_t|\mu_t \sim N((1-\pi_m)\overline{\mu} + \pi_m\mu_t, \quad \sigma_\mu^2\pi_m(1-\pi_m))$$

The mean term implies that, m_t is in expectation a weighted average of $\overline{\mu}$ and reaslized μ_t. When π_m is larger, m_t becomes closer to μ_t in expectation. When $\pi_m = 0$, we have no information on μ_t, so we have $m_t = \overline{\mu}$. $\pi_m = 1$ means that μ_t is observed. The conditional distribution of μ_t given the realization of m_t is given by

$$\mu_t|m_t \sim N(m_t, \quad \sigma_\mu^2(1-\pi_m)).$$

The variance term implies that, as π_m becomes larger, we have more precise information on μ_t. With the calibrated parameter values, $\pi_m = 0.267$. It is possible, though we do not consider in this paper, to increase π_m by considering additional information, called predictor variables in the literature, such as dividend-price ratio (Fama and French (1988)), interest rates and other macro variables. Ludvigson and Ng (2007) extracts the information from a large set of macro and financial variables using the principal components following Stock and Watson (2002).

[7] The new parameters, γ and σ_η has explicit forms: $\gamma = \frac{-(1-\sqrt{1-4\phi^2})}{2\phi}$ with ϕ being the ACF(1) of $(1-\theta L)y_t$, and $\sigma_\eta^2 = \frac{\theta\sigma_u^2 - \sigma_{uw}}{\gamma}$.

[8] $m_t = c + \theta y_t - \gamma\eta_t = c + \theta y_t - \gamma(y_t - m_{t-1}) = c + (\theta-\gamma)y_t + \gamma m_{t-1}$. Therefore $(1-\gamma L)m_t = c + (\theta - \gamma)y_t$

[9] Actually $\sigma_m^2 = \sigma_y^2 - \sigma_\eta^2$.

Note that $\underline{y_t}$ is the infinite past history. When the past history is finite, one can use Kalman filter to obtain the proxy of μ_t.

The conditional distribution of $y_{t+1:t+\tau}$ given $\underline{y_t}$ is below:[10]

$$y_{t+1:t+\tau}|\underline{y_t} \sim N(E_{\tau,ARMA}, \quad V_{\tau,ARMA}) \tag{10}$$

where

$$\begin{aligned} E_{\tau,ARMA} &= \overline{\mu}\tau + (m_t - \overline{\mu})P_\tau \\ V_{\tau,ARMA} &= V_{\tau,oracle} + P_\tau^2\sigma_\mu^2(1-\pi_m) \end{aligned}$$

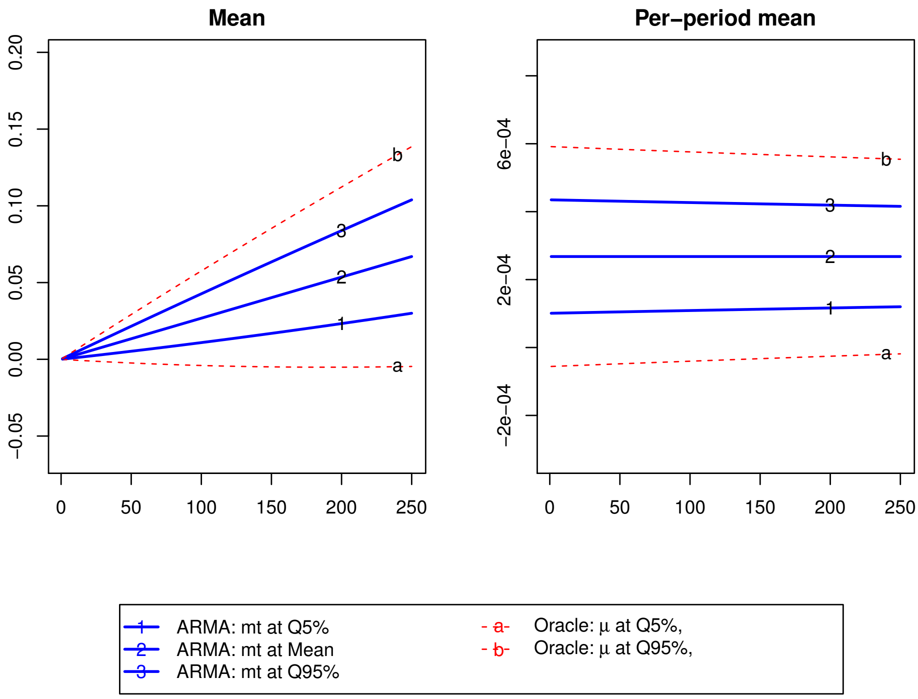
The conditional expectation of $y_{t+1:t+\tau}$ has a similar form to that in the oracle case. The only difference is that we have m_t instead of μ_t. The conditional variance has also a link with the oracle case; unobservability of μ_t means uncertainty, and it gives the additional term in the conditional variance $(P_\tau^2\sigma_\mu^2(1-\pi_m))$. Naturally, with larger π_m, this additional term is smaller.

Figure 4 shows the conditional mean of $y_{t+1:t+\tau}$. The solid lines correspond to the mean and per-period mean of $y_{t+1:t+\tau}$ when one uses m_t instead of μ_t. The line 2 is when m_t is at the mean ($\overline{\mu}$). The line 1 (3) is when m_t is at the quantile 5% (95%) value. The dotted lines are the conditional mean of $y_{t+1:t+\tau}$ when one can observe μ_t, and the line a (b) is when μ_t is at the quantile 5% (95%). Since the variance of m_t is lower than μ_t, the variations of the mean of $y_{t+1:t+\tau}$ is also smaller when one cannot observe μ_t. This lack of variation shows the lack of information.

[10] The mean and variance are derived from

$$\begin{aligned} E(y_{t+1:t+\tau}|\underline{y_t}) &= E[E(y_{t+1:t+\tau}|I_t^{Full})|\underline{y_t}] \\ Var(y_{t+1:t+\tau}|\underline{y_t}) &= Var[E(y_{t+1:t+\tau}|I_t^{Full})|\underline{y_t}] + E[Var(y_{t+1:t+\tau}|I_t^{Full})|\underline{y_t}] \end{aligned}$$

Figure 4: $E_{\tau,ARMA}$ and $E_{\tau,oracle}$



Note of the figure: The left panel shows $E_{\tau,ARMA}$ and the right panel shows $E_{\tau,ARMA}/\tau$. The solid lines correspond to the case where one uses m_t according to the realization of m_t. The dotted lines corresponds to the oracle case: the upper line is when μ_t is at quantile 95%, whereas the lower line is when μ_t is at 5% quantile.

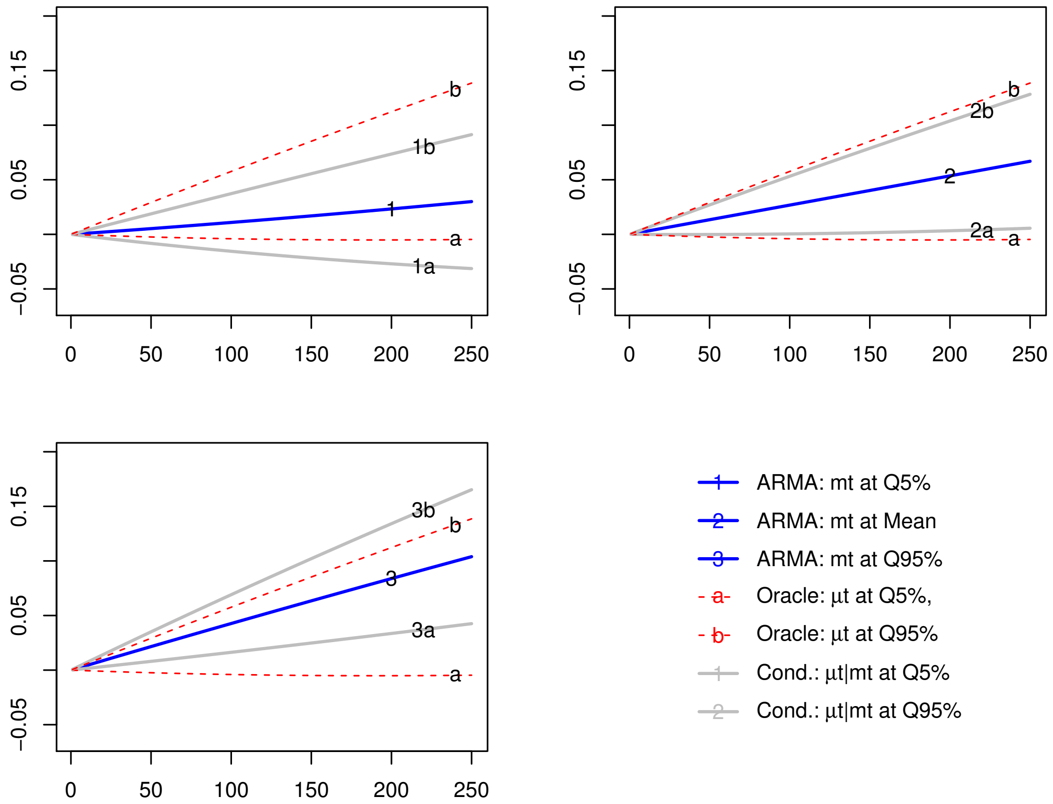
Next we study the lower and upper bound of the 95% confidence interval of $E_{\tau,oracle}$ given the information $\underline{y_t}$. Since $E_{\tau,oracle}$ is a function of μ_t which is the only random variable,

$$E_{\tau,oracle}|\underline{y_t} \sim N(\overline{\mu}\tau + (m_t - \overline{\mu})P_\tau, \quad P_\tau^2 \sigma_\mu^2 (1-\pi_m))$$

because of the conditional distribution of μ_t given $\underline{y_t}$. Let us denote by $E_{\tau,oracle|\underline{y_t}}$ and $V_{\tau,oracle|\underline{y_t}}$ the mean term and variance term above.

The three panels of Figure 5 shows them according to three different realizations of m_t: 5% quantile (upper-left panel), mean (upper-right panel)and 95% quantile (lower panel). The dotted lines corresponds to $E_{\tau,oracle}$ when μ_t is at 5% and 95% quantiles and the mean, and therefore they are the bounds when agents have no information on μ_t. The solid line corresponds to $E_{\tau,ARMA}$ and the two gray lines show the upper and lower bound of the 95 % confidence interval given the realization of m_t derived from $E_{\tau,oracle|\underline{y_t}}$ and $V_{\tau,oracle|\underline{y_t}}$. For example, the lines 1a and 1b shows the upper and lower bound of the confidence set of $E_{\tau,oracle}$ when m_t realizes at 5% quantile. Comparing the dotted lines and the gray lines, the confidence interval shifts downwards because of m_t is low. Moreover, the width of the confidence set decreases because the information from $\underline{y_t}$ resolves some uncertainty.

Figure 5: $E_{\tau,oracle}|\underline{y_t}$ with different realization of m_t



Note of the figure: The solid line are $E_{\tau,ARMA}$ with three different realizations of m_t in three panels. The dotted lines are $E_{\tau,oracle}$ when μ_t is at quantiles 5% and 95%. Finally the gray lines are the upper and lower bounds of the 95% confidence interval of $E_{\tau,oracle}$ given the realization of m_t.

Figure 6 shows the distribution of $E_{\tau,arma}$ conditional on the realization of μ_t.

Figure 6: Distribution of $E_{\tau,ARMA}$ conditional on realization of μ_t

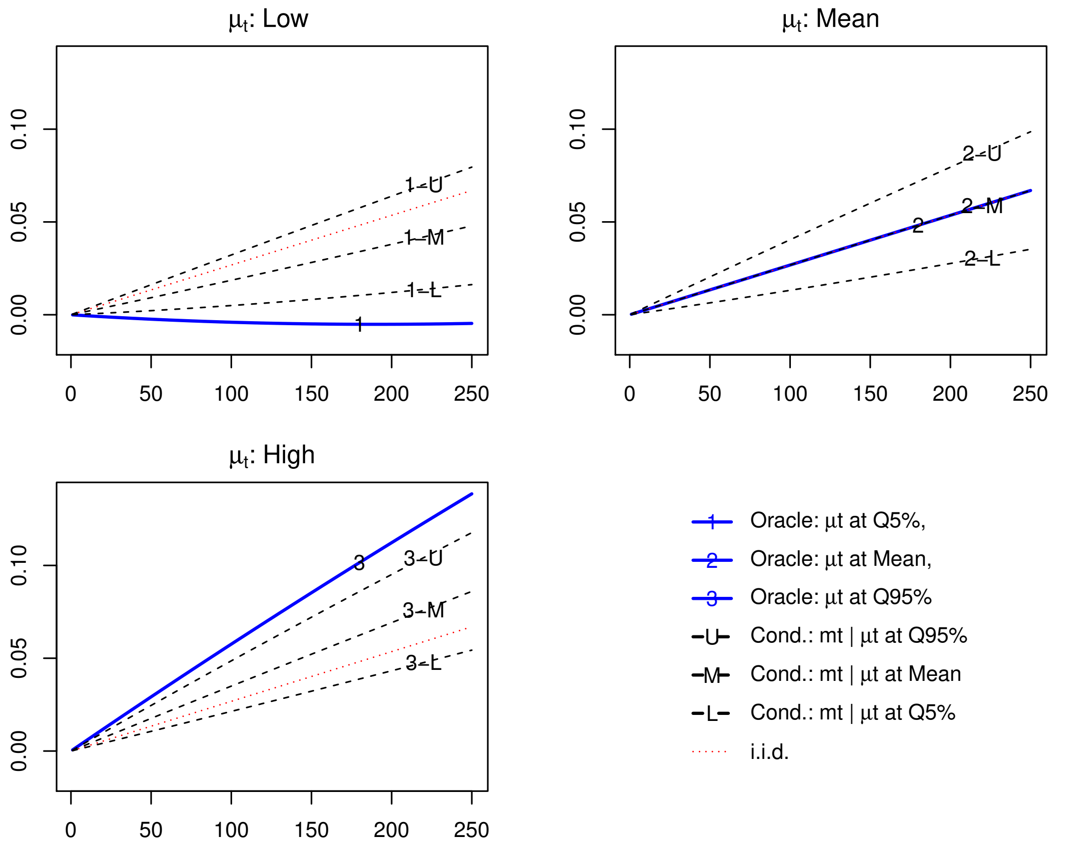
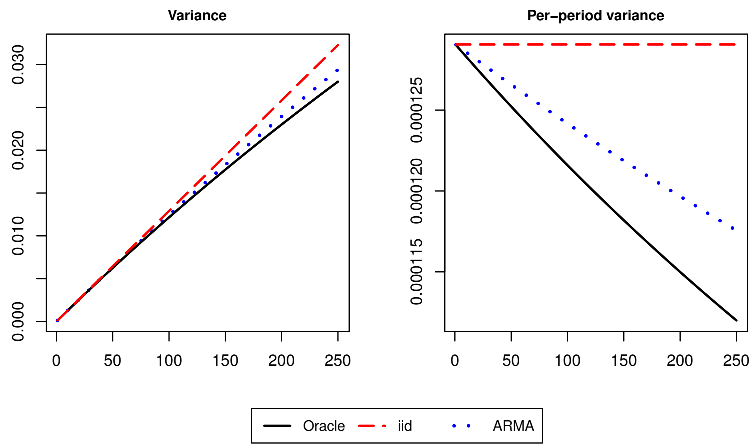


Figure 7 shows the conditional variance of $y_{t+1:t+\tau}$. As a reference, those of oracle case, i.i.d. misspecified case are also plotted. When μ_t is not observed, the conditional variance is between the two cases.

Figure 7: Conditional variance of $y_{t+1:t+\tau}$



Note of the figure: the left panel shows $V_{\tau,oracle}$, $V_{\tau,iid}$ and $V_{\tau,ARMA}$. The right panel shows those divided by τ. The x-axis is the forecast horizon τ.

Finally, the VaR based on the correct model and unobservable μ_t which is updated using $\underline{y_t}$ is

$$VaR(R_{t+\tau}|\underline{y_t}) = 1 - \exp\left(E_{\tau,ARMA} - 2.33\sqrt{V_{\tau,ARMA}}\right)$$

2.5 Term structure of VaR

In this section, we study the "term structure of VaR", which means that we plot VaR as a function of the horizon. This enables us to study VaR when we increase the horizon, from 1 day to 10 days, or 22 days. In the following figures, we let the horizon up to 1 year (250 days). In practice, not many financial institutions care about VaR of such a long horizon, but there are some papers which emphasizes the importance of longer-horizon risk management, such as Engle (2011).

We will draw VaR with the three cases discussed above. We call them as "iid VaR", "oracle VaR" and "ARMA-based VaR". Because Basell II Accord uses 1% VaR of 10-day horizon, we consider 1% VaR of horizons of 1 day up to 1 year.

VaR are shown in the dollar term, i.e., it tells "how much money the agent might lose". It is obtained by multiplying a fixed invest amount in the dollar term and the return rate. We fix the invest amount to 2052.6 million dollars having in mind that 1-day VaR of Bank of America in 2015 was 53 million dollars (reported in page 8 of the annual report, 2015) and 1-day 1% VaR in the return is at about 2.5% with the calibrated parameter values. Therefore, the dollar-terms VaR is computed as

$$\textbf{Dollar term VaR} = 2052.6 \times VaR(R_{t+\tau|t}|M, I_t)$$

First, we study the oracle VaR and see how different realizations of the state variable μ_t affects the computed VaR in order to see the impact of the time-variability of μ_t on oracle VaR. We choose several values of μ_t and draw the term structure of VaR. By comparing the oracle VaR with different μ_t, we study the *mean effect*. We show that mean effect is more important for a longer horizon, but at as short as 10 day horizon, there exists a non-negligible mean effect.

Second, we study iid VaR and see how "well" these VaR are, i.e., how close iid VaR can get to the oracle VaR. We look at the *variance effect* by comparing iid VaR with oracle VaR with μ_t at the mean value. The two VaR values are computed with the same conditional mean, but different conditional variance. We show that the variance effect is not as large as the mean effect that we studied with the oracle VaR. Therefore iid VaR performs well when μ_t is close to its mean but when μ_t takes a value far away from its mean, it performs very badly.

Finally we study ARMA-based VaR. Since ARMA-based VaR is dependent on the realization of m_t, we consider a sort of confidence interval of oracle VaR given the realization of m_t. We show that the agent can adjust their VaR according to the realization of m_t, and therefore it performs relatively well especially when m_t takes value far away from its mean.

2.5.1 Oracle VaR

The "oracle VaR" is given by

$$VaR(R_{t+\tau}|I_t^{Full}) = 1 - \exp\left((E_{\tau,oracle} - 2.33\sqrt{V_{\tau,oracle}}\right)$$

Figure 8 and Table 2 show the oracle VaR with 5 different values of μ_t: (1) 5% quantile, (2) 20% quantile, (3) mean, (4) 80% quantile and (5) 95% quantile.

Figure 8 shows that the VaR values are fanning out. As the horizon increases, the oracle VaR takes the value more and more different with each other as the horizon increases. It means that VaR becomes more and more sensitive to the realization of μ_t.

Table 2 shows that, at 10-day horizon, the difference of VaR when μ_t takes quantiles 95% and 5% is 12.2 million dollars, which is about 7.6% of the Oracle VaR when μ_t takes its mean value. Under the Basel II framework, the difference of VaR implies 36.6 million dollars difference in the capital held by a bank.

Figure 8: Term structure of VaR: Oracle VaR

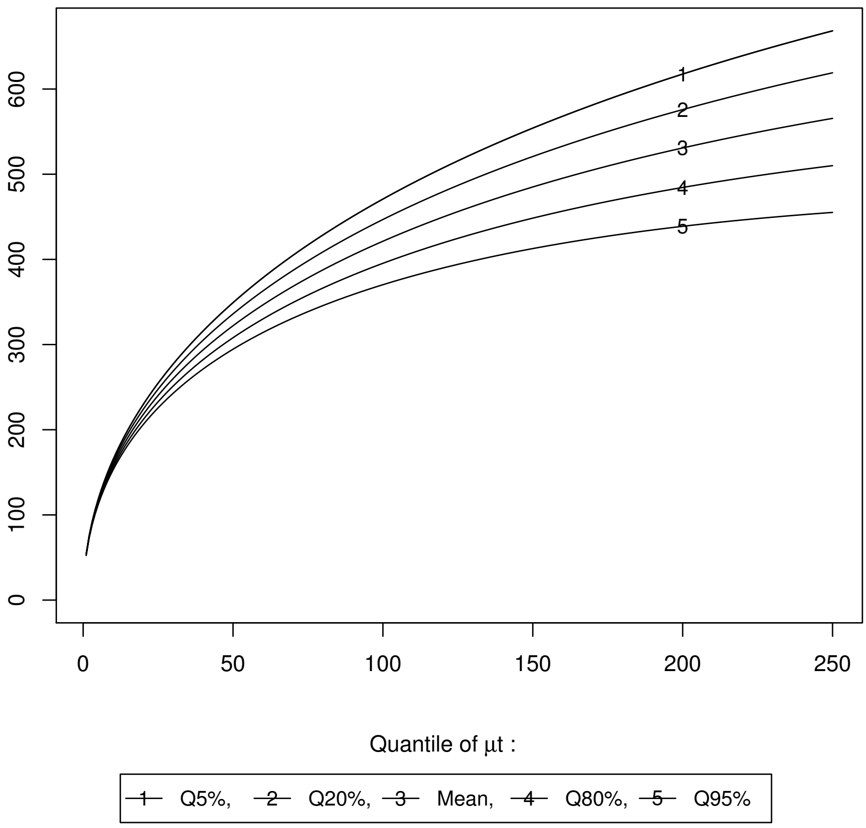


Table 2: Term structure of VaR: Oracle and iid VaR

	Oracle VaR					iid VaR
Horizon	Q5%	Q20%	Mean	Q80%	Q95%	
1	53.64	53.32	52.99	52.66	52.34	53.00
10	165.13	162.16	159.04	155.92	152.93	159.51
22	239.91	233.66	227.09	220.50	214.18	228.58
66	394.59	377.76	359.94	341.94	324.57	367.07
125	515.68	486.85	456.07	424.69	394.16	473.10
250	668.33	619.06	565.55	510.04	455.13	607.47

(Million Dollars)

2.5.2 iid VaR

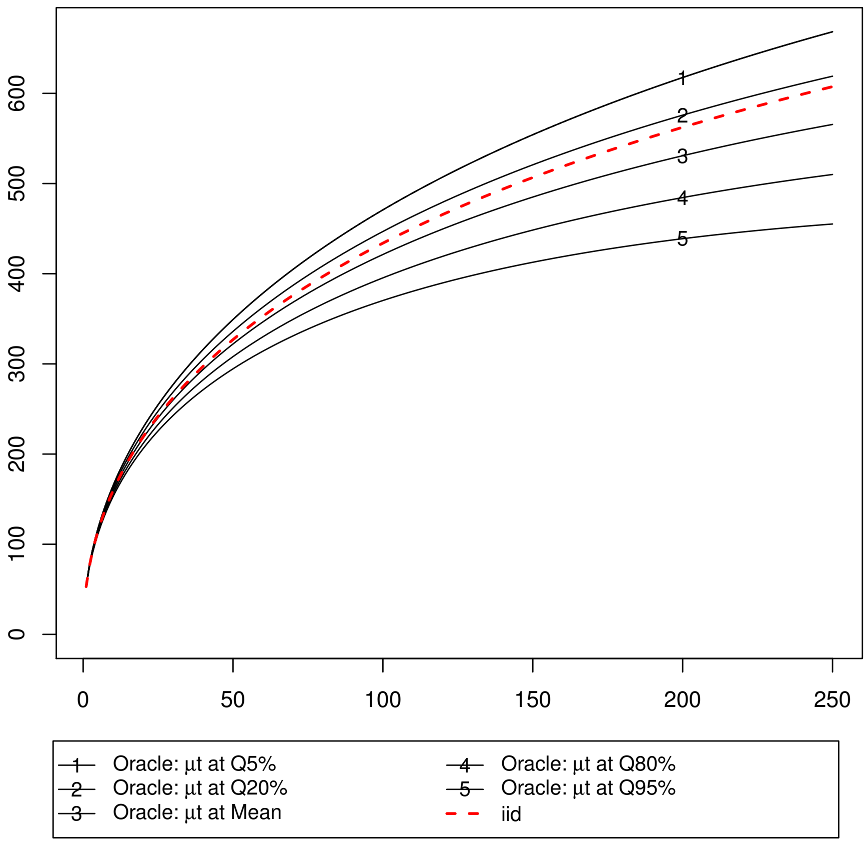
The "iid VaR" is given by

$$VaR(R_{t+\tau|t}|M_{iid}) = 1 - \exp\left(\mu_t\tau - 2.33\sqrt{\sigma_y^2\tau}\right)$$

Note that iid VaR is not vary regardless of the realization of μ_t. Figure 9 shows the term structure of iid VaR with the oracle VaR surimposed, and Table 2 shows the iid VaR on the right column. When μ_t takes the mean value (line 3 in the graph), the conditional mean in the oracle case coincides with the i.i.d. case and therefore, the difference between the oracle VaR (line 3) and iid VaR comes only from the difference of the conditional variance. This difference does not seem large: even at 1 year horizon, the difference is 42 million dollars which accounts 7.4% of the average oracle VaR at this horizon. Therefore, the mean term seems to have more effect on VaR than the variance term.

Comparing iid VaR with other oracle VaR, there is smaller difference if μ_t takes lower value, i.e., in the bad state. This is because the conditional variance is higher with i.i.d. because of the lack of the information. On the other hand, there is larger difference when μ_t takes higher value, i.e., in the good state. In this case, an agent who has the misspecified model wrongly overestimates the VaR which result in having too much capital in the framework of Basel II.

Figure 9: Term structure of VaR: iid VaR



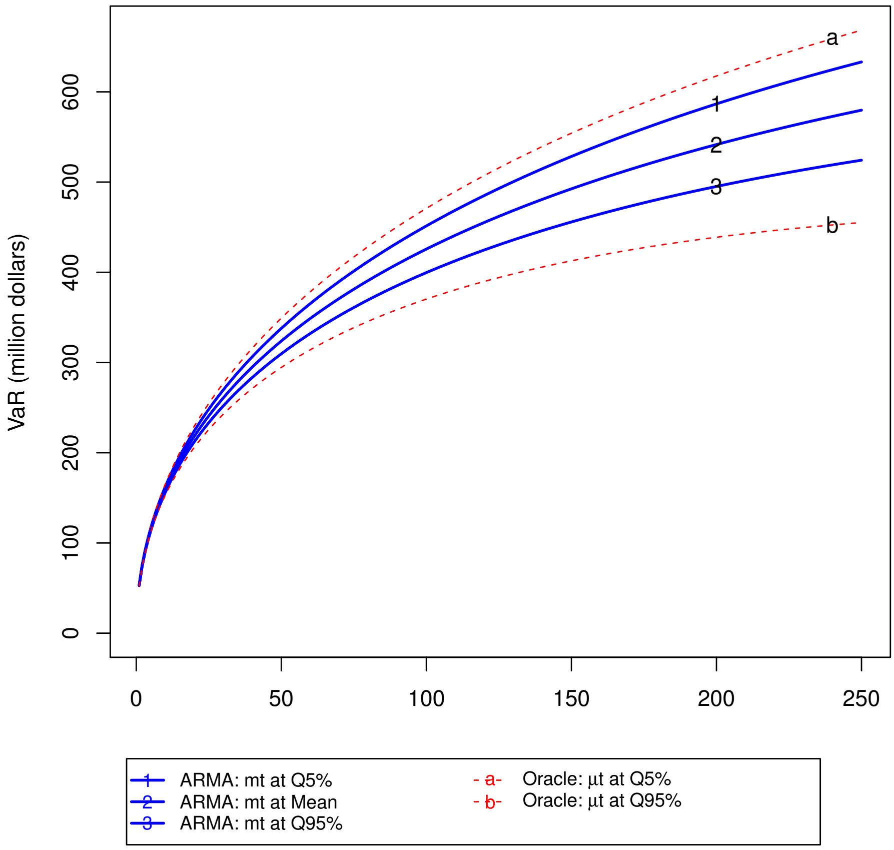
2.5.3 ARMA-based VaR

The ARMA-based VaR is given by

$$VaR(R_{t+\tau}|\underline{y_t}) = 1 - \exp\left(E_{\tau,ARMA} - 2.33\sqrt{V_{\tau,ARMA}}\right)$$

and it is subject to the realization of the proxy of μ_t, m_t. We compare ARMA-based and oracle VaR as we do to compare the conditional mean in the above section. Firstly, we compare the unconditional distribution of VaR in Figure 10 shows ARMA-based VaR in the solid line and oracle VaR in the dotted line. The 3 solid line corresponds to the case where m_t takes its mean value, and 5% and 95% quantile values. The two dotted line shows the oracle VaR where μ_t takes these its mean and 5% and 95% quantile values. As can be seen, the ARMA-based VaR has less variation than the oracle VaR.

Figure 10: ARMA-based VaR compared to oracle VaR



Now let us study the distribution of the Oracle VaR given $\underline{m_t}$. Oracle VaR is written as

$$VaR(R_{t+1:t+\tau}) = 1 - \exp\left(E_{\tau,oracle} - 2.33\sqrt{V_{\tau,oracle}}\right)$$

where the term inside the exponential is, given $\underline{m_t}$, Normally distributed with mean $E_{\tau,oracle|\underline{m_t}} - 2.33\sqrt{V_{\tau,oracle}}$ and variance $V_{\tau,oracle|\underline{m_t}}$. Therefore the exponential term is log-Normally distributed. From this, we can derive the 5% and 95% quantile of $VaR(R_{t+1:t+\tau})$ given $\underline{m_t}$.

Figure 11 shows a predictive confidence interval of oracle VaR when one uses ARMA(1,1) representation. The gray lines show the upper and lower bound of the confidence interval at 95% when m_t takes values of 5% and 95% quantiles and the mean. The dotted lines are the oracle VaR when μ_t is at 5% and 95% quantile of its unconditional distribution, so it may be seen as the confidence interval when one does not use any information on μ_t. If m_t takes a low value, it is probable that μ_t also take a low value, so ARMA-based VaR becomes larger.

Figure 11: ARMA-based VaR and 5%, 95% Quantile of Oracle VaR given $\underline{m_t}$

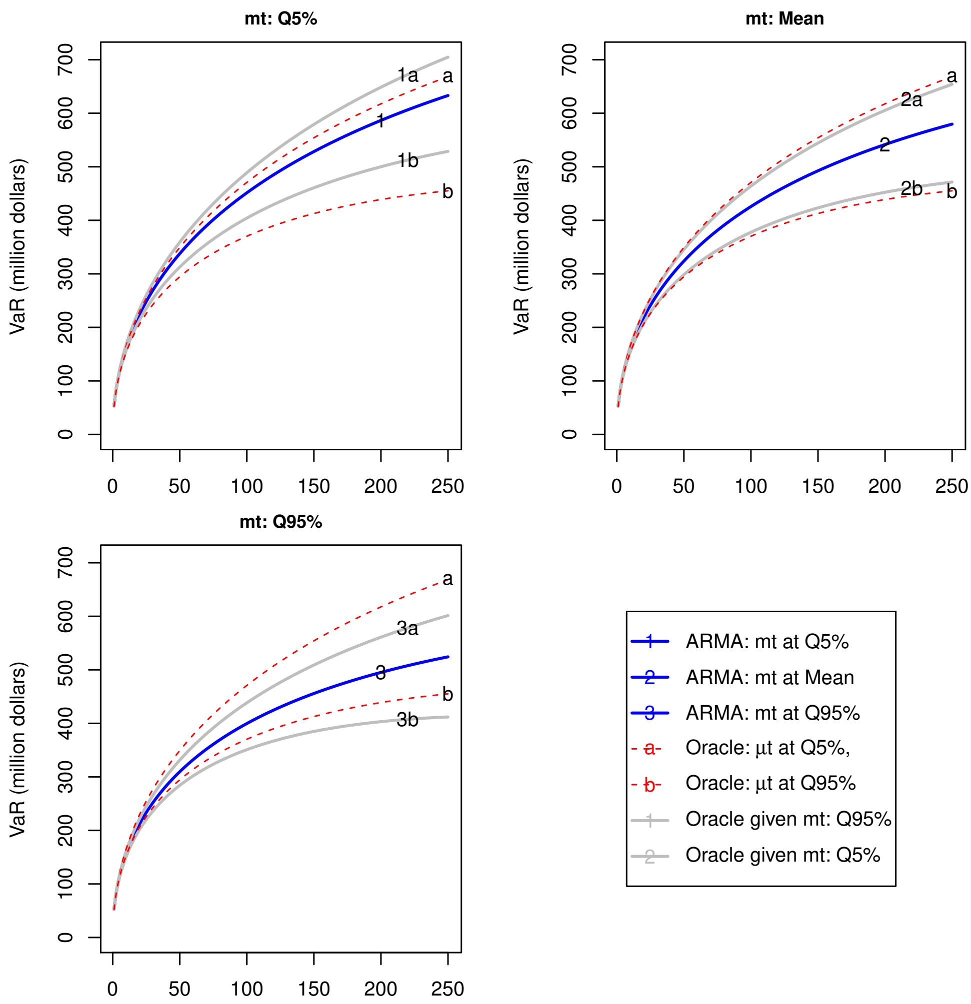


Table 3: Oracle VaR conditional on realization of m_t

	Horizon	ARMA	Oracle given m_t		
m_t			5%	Mean	95%
Q5%	1	53.33	53.88	53.33	52.77
	10	162.36	167.40	162.18	156.96
	22	234.25	244.66	233.70	222.70
	66	380.39	407.31	377.81	347.97
	125	492.90	537.31	486.80	435.24
	250	633.12	704.78	618.49	528.81
Mean	1	53.00	53.55	52.99	52.44
	10	159.22	164.26	159.04	153.81
	22	227.63	238.08	227.08	216.04
	66	362.46	389.67	359.84	329.69
	125	461.98	507.27	455.76	403.18
	250	579.69	654.04	564.51	471.45
Q95%	1	52.66	53.21	52.66	52.10
	10	156.07	161.12	155.89	150.65
	22	220.99	231.47	220.43	209.35
	66	344.33	371.83	341.69	311.21
	125	430.45	476.64	424.11	370.48
	250	524.25	601.40	508.50	411.93

(Million dollars)

Now let us study the opposite case, i.e., the distribution of ARMA-based VaR conditional on the realization of μ_t.

Figure 12: ARMA-basd VaR given μ_t

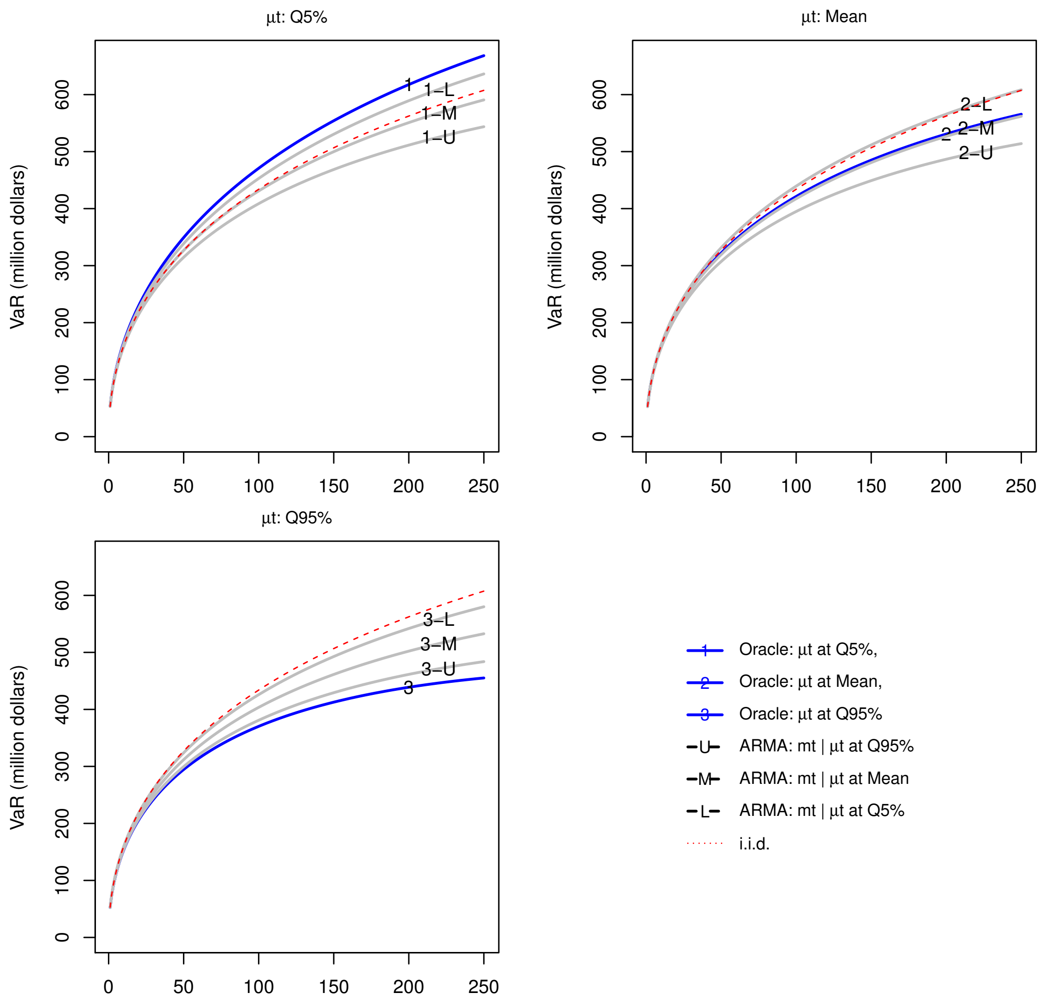


Table 4: ARMA-based VaR conditional on realization of μ_t

	Horizon	Oracle	ARMA-based VaR given μ_t		
μ_t			5%	Mean	95%
Q5%	1	53.64	53.46	53.17	52.88
	10	165.13	163.53	160.84	158.14
	22	239.91	236.71	231.05	225.37
	66	394.59	387.03	371.73	356.30
	125	515.68	504.26	478.01	451.31
	250	668.33	652.53	607.50	561.03
Mean	1	52.99	53.28	53.00	52.71
	10	159.04	161.91	159.22	156.52
	22	227.09	233.31	227.63	221.94
	66	359.94	377.83	362.46	346.94
	125	456.07	488.50	461.98	435.01
	250	565.55	625.58	579.69	532.32
Q95%	1	52.34	53.11	52.83	52.54
	10	152.93	160.29	157.59	154.89
	22	214.18	229.89	224.21	218.50
	66	324.57	368.59	353.13	337.52
	125	394.16	472.58	445.79	418.55
	250	455.13	598.12	551.34	503.06

(Million dollars)

2.5.4 Other parameter values

Appendix shows the oracle and iid VaR computed with different parameter values. When we increase the variance of μ_t keeping other parameters constant, VaR becomes more sensitive to the realization of μ_t, i.e., the mean effect becomes larger.

On the other hand, when θ is decreased, i.e., the state variables is assumed to be less persistent, then the variance effect becomes more important. This is because the state variable mean revert more quickly and thus there is not many uncertainty as to the future value of μ_t and thus the conditional variance in the oracle case decreases. Therefore the iid VaR performs worse and worse as θ becomes smaller.

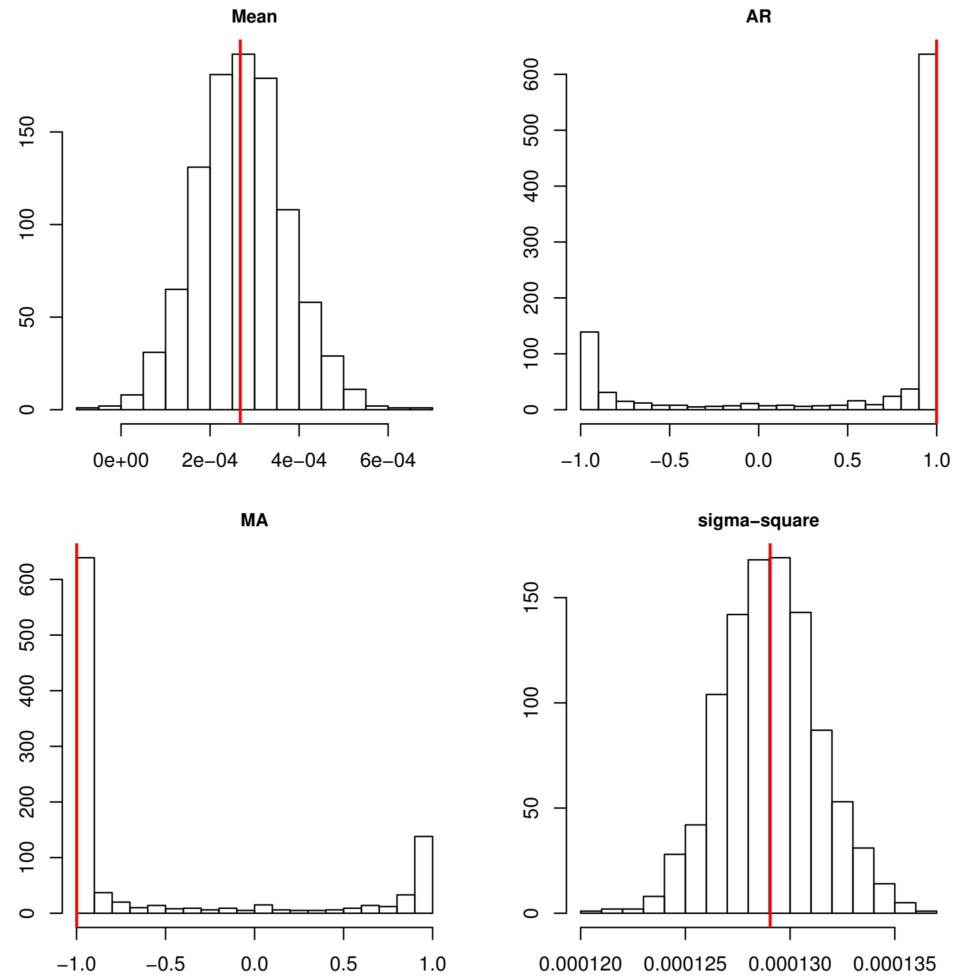
2.6 Parameter uncertainty: Simulation results

In this section, we simulate 1000 replications of path of size 6250. Then we estimate the i.i.d. model (M_{iid}) and ARMA(1,1) model. First we discuss the identification issue in estimating the ARMA(1,1) model when the process is highly persistent. Second we study the properties of VaR computed with estimated parameter values.

2.6.1 Estimating ARMA(1,1) model

Figure 13 shows histograms of estimated parameter values based on maximum likelihood. The red and solid lines show the true values. Overall, the distribution of estimators are peaked at the true parameter values.

Figure 13: Monte Carlo estimation results: ARMA(1,1) estimation

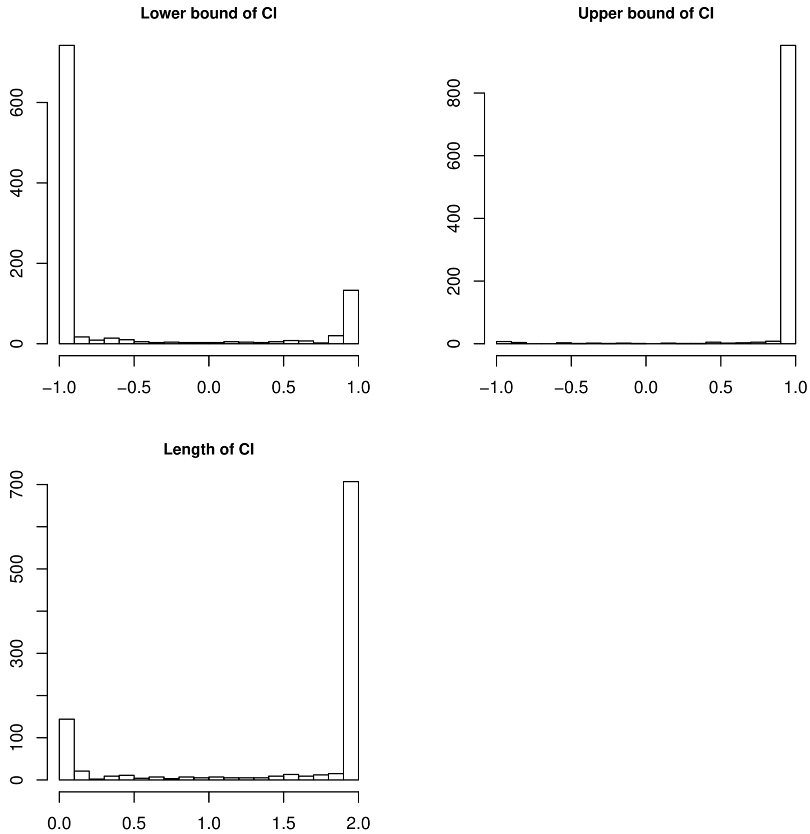


2.6.2 Identification issue: is $H_0 : M_{iid}$ rejected?

We examine the rate at which the hypothesis that both AR and MA coefficient are equal to zero, which means that y_t is iid. This is to test if M_{iid} is true.

We build the confidence interval for the value of the AR coefficient for each simulation. We do so by inverting the Likelihood-Ratio test. We take a grid of $[-1, 1]$ with 0.01 interval. Figure 14 demonstrates the properties of confidence intervals. Out of 1000 simulations, the lower bound of the confidence interval becomes -0.99 for 869 times, the upper bound becomes 0.99 for 919 times, and the length becomes 0.99 for 828 times. Therefore, it is extremely difficult to distinguish between M_0 and M_{iid} by looking at the data.

Figure 14: Estimator and confidence interval of AR coefficient in ARMA(1,1) model



Finally, the null hypothesis that both the AR and MA coefficients are equal to zero is rejected at 5 % level for 182 times out of 1000. In other words, we fail to reject it for 818 times out of 1000.

2.6.3 VaR

We compute the VaR using estimated parameters for the iid model and ARMA(1,1) model. In each case, we use the estimated parameter values whereas we assume that the value of the state variable, μ_t is known at t. The sample mean of ARMA-estimated and iid-estimated VaR are shown in Figures 15. First, ARMA-estimated VaR changes according to the value of μ_t. But their variation is much smaller than the oracle VaR. Second, the iid-estimated VaR is invariant to the value of μ_t and it is close to the misspecified VaR without parameter uncertainty. Table 5 summarizes.

Figure 15: ARMA- and iid- estimated VaR (sample mean)

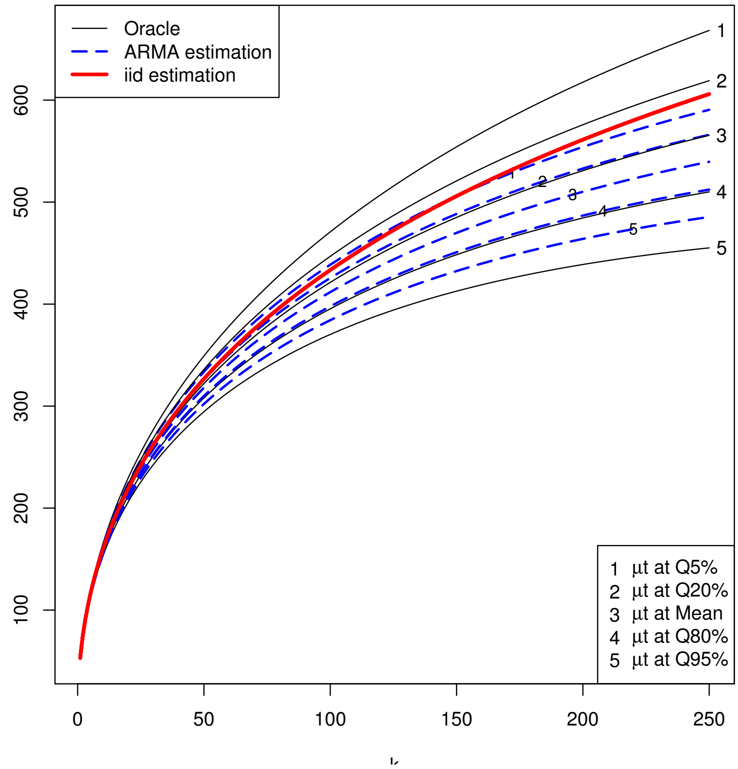


Table 5: VaR based on estimation (ARMA and M_{iid})

μ_t	Horizon (Days)	Oracle	ARMA-estimate Mean	ARMA-estimate Q-5%	ARMA-estimate Q-95%	iid-estimate Mean	iid-estimate Q-5%	iid-estimate Q-95%
Q5%	1	53.6	**53.6**	52.9	54.5	53.0	52.2	53.9
	10	165.1	**162.7**	156.1	168.0	159.5	155.5	163.2
	22	239.9	**233.7**	220.4	245.4	228.5	220.8	235.3
	66	394.6	**373.7**	341.6	408.2	366.8	346.9	384.7
	125	515.7	**475.3**	415.4	530.3	472.4	437.9	503.8
	250	668.3	590.5	478.5	695.5	**606.0**	543.7	662.9
Q20%	1	53.3	**53.3**	52.6	54.1	53.0	52.2	53.9
	10	162.2	**160.7**	155.4	166.2	159.5	155.5	163.2
	22	233.7	**229.8**	218.3	242.8	228.5	220.8	235.3
	66	377.8	364.2	333.5	399.8	**366.8**	346.9	384.7
	125	486.9	460.0	399.8	525.2	**472.4**	437.9	503.8
	250	619.1	566.0	449.0	690.0	**606.0**	543.7	662.9
Mean	1	53.0	**53.0**	52.2	53.8	53.0	52.2	53.9
	10	159.0	**158.6**	153.6	164.8	159.5	155.5	163.2
	22	227.1	**225.7**	215.4	240.3	228.5	220.8	235.3
	66	359.9	**354.2**	321.4	396.7	366.8	346.9	384.7
	125	456.1	**443.7**	382.2	519.5	472.4	437.9	503.8
	250	565.5	**539.6**	414.4	683.8	606.0	543.7	662.9
Q80%	1	52.7	**52.7**	51.9	53.5	53.0	52.2	53.9
	10	155.9	**156.5**	151.6	163.8	159.5	155.5	163.2
	22	220.5	**221.6**	210.9	237.6	228.5	220.8	235.3
	66	341.9	**344.1**	308.7	393.0	366.8	346.9	384.7
	125	424.7	**427.1**	360.1	515.9	472.4	437.9	503.8
	250	510.0	**512.3**	369.4	677.6	606.0	543.7	662.9
Q95%	1	52.3	**52.3**	51.6	53.2	53.0	52.2	53.9
	10	152.9	**154.6**	149.0	163.1	159.5	155.5	163.2
	22	214.2	**217.7**	205.4	236.1	228.5	220.8	235.3
	66	324.6	**334.3**	294.5	390.3	366.8	346.9	384.7
	125	394.2	**411.1**	332.6	512.5	472.4	437.9	503.8
	250	455.1	**485.4**	322.5	673.7	606.0	543.7	662.9

(Million dollars)

Note of the table: The bold number indicates the closer value to the oracle VaR.

Figure 16 shows the distribution of ARMA-estimation-based VaR with various values of μ_t. When μ_t is lower than the mean, the agent tends to underestimate the risk. When μ_t is higher than the mean, the agent tends to overestimate the risk.

Figure 17 shows the distribution of iid-estimation-based VaR. The pattern observed with the ARMA-estimation-based VaR is more prominent here. When μ_t is smaller than the mean, the agent almost always underestimate the risk. When μ_t is larger than the mean, the agent almost always overestimate the risk.

Figure 16: ARMA-estimated VaR: distribution

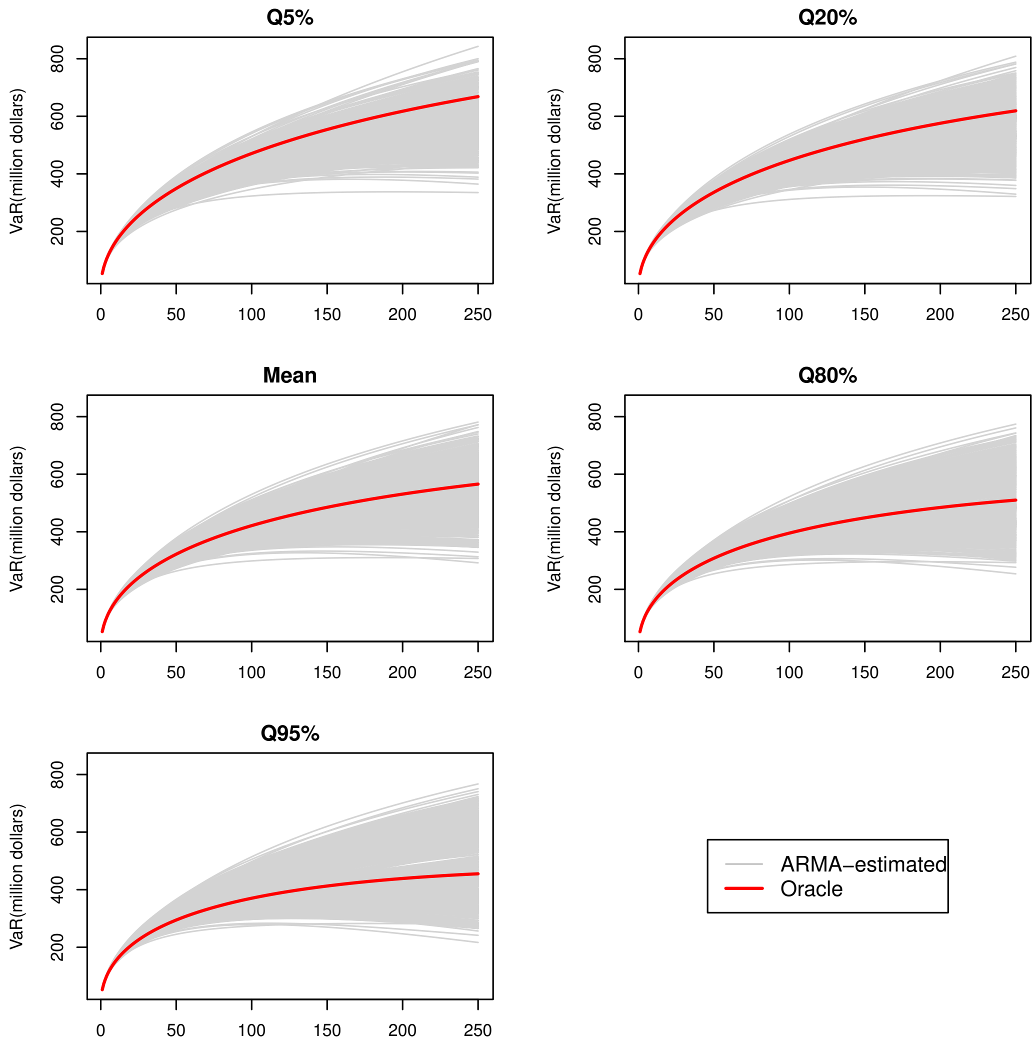
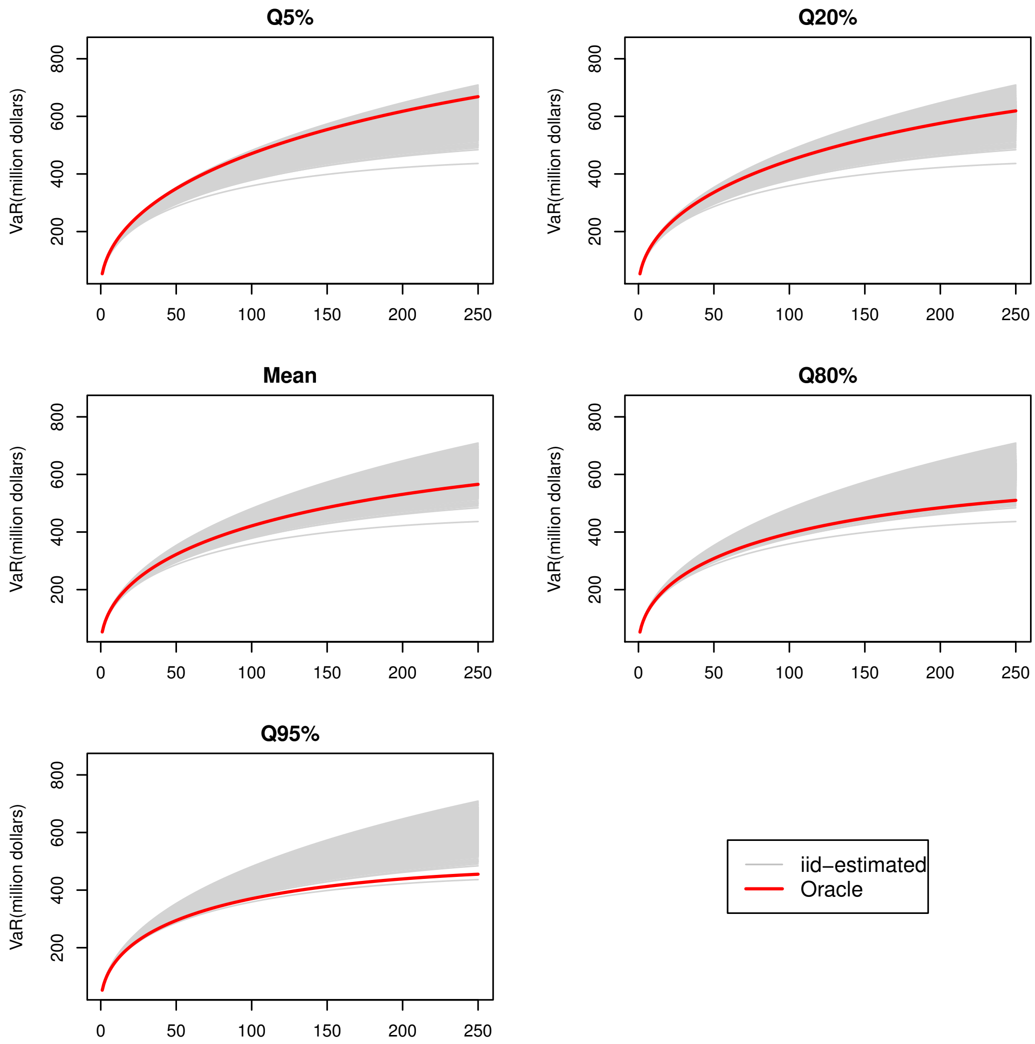


Figure 17: iid-estimated VaR: distribution



3 Conditional heteroscedasticity model

3.1 Model

Now let us consider the case where the volatility is time-varying. We consider a model where the characteristic function of future returns are affine in the state variable at t, so that the analytical computation is possible. Especially we consider the model proposed by Heston and Nandi (2000) as the following:

Correct ($\mathbf{M_1}$)

$$y_t = r + \lambda h_{t-1} + \sqrt{h_{t-1}} z_t \tag{11}$$

$$h_t = \omega + \beta h_{t-1} + \alpha (z_t - \gamma \sqrt{h_{t-1}})^2 \tag{12}$$

$$z_t \sim i.i.d. N(0,1) \tag{13}$$

with $\beta + \alpha\gamma^2 < 1$.[11] r is the risk-free rate.

Then the characteristic function of $y_{t+1:t+\tau}$, which we denote by $C(\tau, u)$ is affine in h_t, i.e., $C(\tau, u) = \exp\left(a(\tau, u) + b(\tau, u)h_t\right)$ with some function $a(\cdot)$ and $b(\cdot)$. It is possible to invert $C(\tau, u)$ to obtain the distribution function of $y_{t+1:t+\tau}$.[12] and the 1% VaR of horizon τ, VaR_τ is such that

$$Pr(y_{t+1:t+\tau} \leq VaR_\tau) = 0.01$$

3.2 Misspecification in the mean

Suppose the true DGP is given as M_1 but an agent has a misspecified model as follows:

Misspecified constant mean model ($\mathbf{M_{cons}}$)

$$y_t = \widetilde{r} + \sqrt{h_{t-1}} z_t \tag{14}$$

$$\widetilde{r} = r + \lambda E(h_t) = r + \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2} \tag{15}$$

$$h_t = \omega + \beta h_{t-1} + \alpha(z_t - \gamma\sqrt{h_{t-1}})^2 \tag{16}$$

$$z_t \sim i.i.d.N(0, 1) \tag{17}$$

Here the agent wrongly assumes that the conditional mean is constant. For now we suppose that the investor knows the "true" parameter values and h_t at time t as well as the parameter values.

3.3 Illustration by estimated parameters

3.3.1 Estimation procedure

We use the estimated results to calibrate the model. The estimation uses S&P 500 daily Index obtained by CRSP, from 02/01/1990 until 31/31/2015 with 6553 observations. The maximum likelihood estimator is shown in Table 6:

Table 6: Estimation results

	GARCH-in-Mean			GARCH		
Parameter	Estimates	Std. error	t-stat	Estimates	Std. error	t-stat
λ	2.295	2.179	1.052	0 (fixed)		
$r(*)$	4.81e-09	1.71e-04	2.80e-05	4.81e-09	1.34e-04	3.56e-05
ω	-7.51e-07	4.01e-07	-1.872	-7.73e-07	3.62e-07	-2.132
α	3.46e-06	3.87e-07	8.930	3.47e-06	2.57e-07	13.52
β	0.846	2.01e-02	41.974	0.849	2.69e-02	31.534
γ	192.871	21.809	8.843	192.871	23.988	8.040
Log likelihood	21416.684			21414.706		
Obs	6553			6553		

(*) with restriction $r \geq 0$

[11] The first and second moment of h_t is given by $E(h_t) = \frac{\omega+\alpha}{1-\beta-\alpha\gamma^2}$ and $Var(h_t) = \frac{2\alpha^2(1+2\gamma^2 E(h_t))}{1-(\beta+\alpha\gamma^2)^2}$.

[12] See Duffie, Pan, and Singleton (2000) and the Appendix.

where we put a restriction that $r \geq 0$ since it has a theoretical meaning of risk-free rate.

The persistence of h_t, $\beta + \alpha\gamma^2$ is estimated to be 0.963. The confidence interval based on the likelihood ratio test for λ is $[0.04, 4.56]$.

λ determines how much the conditional mean varies, i.e., if λ is larger, the importance is higher. Table 7 shows the estimation results from other papers. They use S&P500 Index for different periods of time, and their estimates vary from 0.205 to 2.899. Given these results in the literature, the estimation we obtained $\widehat{\lambda} = 2.295$ is consistent with the literature.

Table 7: Estimation results from literature

Parameter	HN(2000)	CJO(2009)	CJO(2011)
λ	0.205	2.899	1.661
r	fixed	fixed	fixed
ω	5.02e-6	-7.756e-7	-1.269e-6
α	1.32e-6	4.546e-6	2.807e-6
β	0.589	0.9041	0.9451
γ	421.39	115.9	116.0
Obs	755	2547	
	(01/08/92-12/30/94)	(02/01/95-31/12/04)	(06/62-12/09)

CJO(2009): Christoffersen, Jacobs, and ornthanalai (2009)
CJO(2011): Christoffersen, Jacobs, and ornthanalai (2011)

In the next section, we study the VaR treating the estimated parameter values as the "true" values that generates the data. We also compute the VaR with misspecified model (M_{cons}) and study the difference between the VaR with the correcly specified model (M_1).

3.3.2 Value of the state variable: h_t

In order to compute the VaR based on the information at t, we need the value of the state variable, which is h_t in our case. We chose the sample quantile of 5%, 20%, 80% and 95% as well as the sample mean of the estimated h_t. The sample quantiles and the sample mean are shown in Table 8.

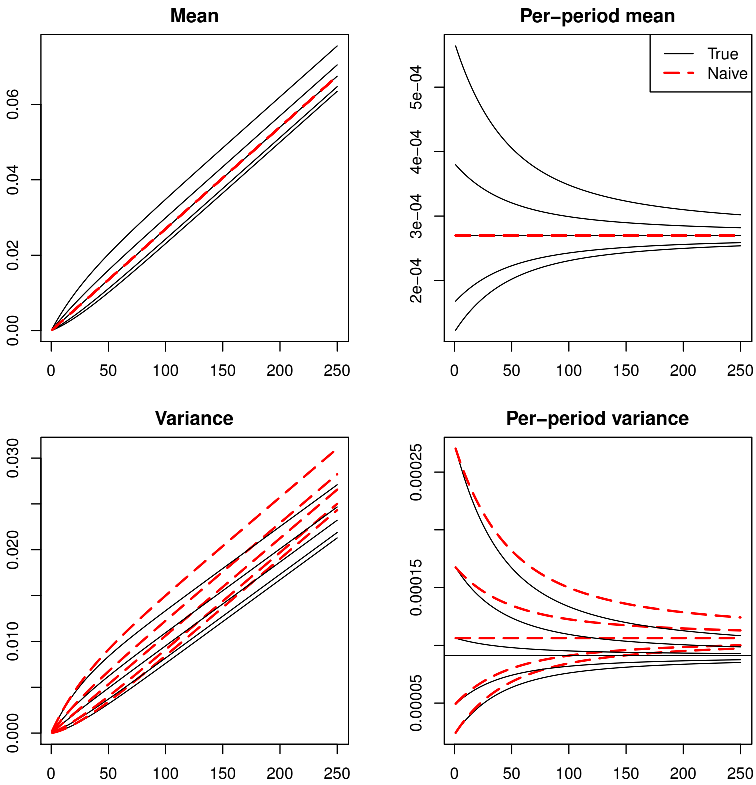
Table 8: Sample quantile of h_t

Quantile	5%	20%	Mean	80%	95%
Value	2.59e-05	4.54e-05	1.09e-04	1.56e-04	2.65e-04

3.3.3 Mean and variance effect

Figure 18 shows the mean and variance of $y_{t+1:t+\tau}$ with h_t taking 5 different values. The right panels depicts the per-period mean and variance of $y_{t+1:t+\tau}$, i.e., $E_t(y_{t+1:t+\tau})/k$ and $Var_t(y_{t+1:t+\tau})/k$. The per-period mean converges to the unconditional mean, but the per-period variance converges differently if we have the correct model and the misspecified model.

Figure 18: Mean and variance of $y_{t+1:t+\tau}$



3.3.4 VaR with the two cases

In this section, we consider the oracle VaR, when an agent has the correct model (M_1). Here we assume away the parameter uncertainty, so the agent is assumed to know the true parameter values.

Table 9 and Figure 20 show VaR computed with Garch-in-Mean and Garch model with various values of h_t. For each horizon, the higher the state variable, the greater the VaR is. As the horizon increases, it increases at first but reduces due to the mean-reversion of the volatility.

Table 9: VaR with GARCH model in comparison with GARCH-in-Mean model

h_t	Horizon (days)	GARCH-in-Mean A	GARCH B	Difference B-A
Q5%	1	24.0	24.2	0.2
	10	107.0	110.7	3.7
	22	178.5	189.8	11.3
	66	346.6	394.9	48.3
	125	473.1	570.9	97.8
	250	610.0	786.9	176.9
Q20%	1	31.7	32.0	0.3
	10	125.8	130.6	4.8
	22	200.0	213.7	13.7
	66	364.1	417.1	53.0
	125	484.2	587.0	102.8
	250	615.5	795.9	180.5
Mean	1	48.8	49.3	0.5
	10	168.9	177.0	8.2
	22	251.2	271.7	20.4
	66	410.0	476.5	66.5
	125	516.1	633.4	117.3
	250	631.9	823.7	191.8
Q80%	1	58.2	59.0	0.8
	10	192.6	203.0	10.5
	22	280.0	304.9	24.9
	66	437.5	512.7	75.3
	125	536.4	663.5	127.1
	250	643.3	842.9	199.7
Q95%	1	75.3	76.5	1.2
	10	235.1	250.2	15.1
	22	331.0	365.1	34.1
	66	487.8	580.5	92.7
	125	575.5	722.5	147.0
	250	666.5	883.1	216.6

(Million dollars)

Figure 19: VaR with GARCH-in-Mean model and GARCH model (250 days)

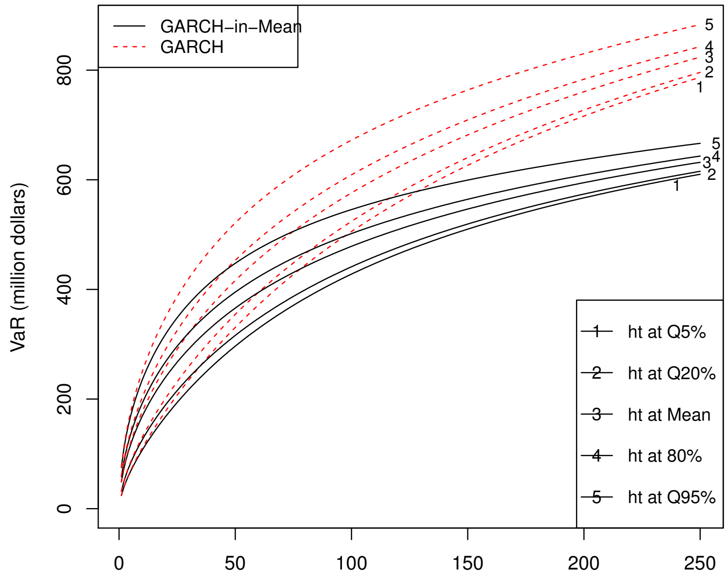
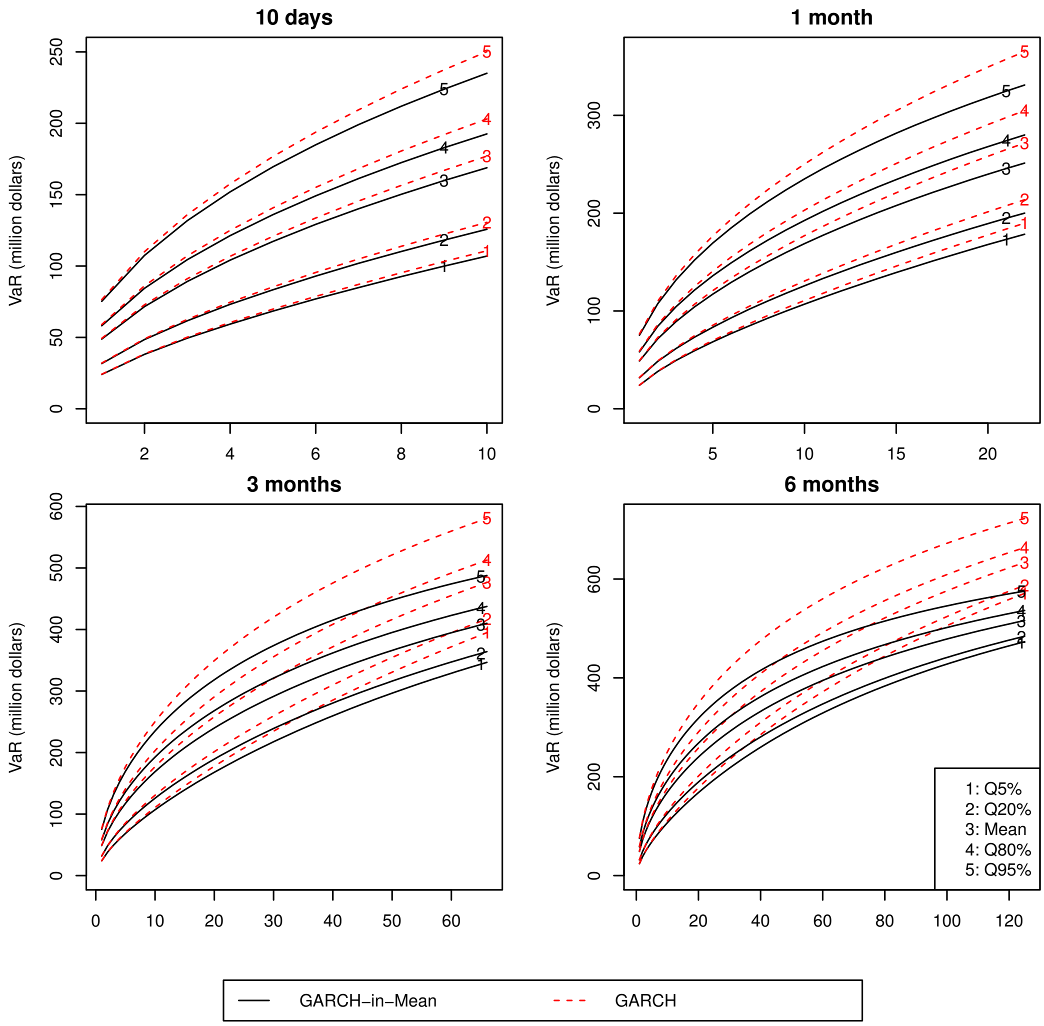


Figure 20: VaR with GARCH-in-Mean model and GARCH model (shorter horizons)



3.4 Parameter uncertainty: Simulation results

In this section, we show the simulation results where we estimate the two models: the correct model (M_1) and the misspecified model (M_{cons}). The estimation is done by the Maximum Likelihood. The former case is referred to as "GARCH-in-Mean" and the latter is referred to as "GARCH" since $\lambda = 0$ is imposed. First we discuss the identification issue for λ and then we study the VaR with estimated parameters with two models.

Simulation procedure is as follows. We first simulate 1000 paths of sample size 2500 with the data-generating process as in M_1 with parameter values the same as the previous section, i.e., shown in the column (a) of Table 6. Then, we estimate the two models for each simulation.

3.4.1 Estimation of GARCH-in-Mean and GARCH models

Figure 21 shows the Monte Carlo results of the GARCH-in-Mean estimation result. Each panel shows the histogram of estimated parameter values. The red and dashed lines show the true parameter values. Overall, the histograms are centered at the true parameter values. Figure 22 shows the counterpart of GARCH. When imposing $\lambda = 0$, ω and γ are influenced and the estimations become far from the true values.

The bottom-right panel shows the histogram of estimated $E(h_t)$ with the two models. With the GARCH-in-Mean, it is centered at the true value, but with GARCH, $E(h_t)$ tends to be estimated to be larger than the true values.

Table 10: Estimation results for GARCH-in-Mean

	Sample mean	Sample std.dev	min	max	True
r	7.25e-05	1.04e-04	4.74e-09	6.23e-04	4.81e-09
ω	-7.44e-07	3.03e-07	-1.66e-06	5.960e-07	-7.51e-07
α	3.37e-06	2.89e-07	2.53e-06	4.30e-06	3.46e-06
β	0.845	1.23e-02	0.803	0.879	0.846
γ	196.95	5.08	192.85	223.15	192.871
λ	1.715	1.949	-3.591	9.317	2.295

Table 11: Estimation results for GARCH

	Sample mean	Sample std.dev	min	max
r	1.33e-04	9.97e-05	4.77e-09	5.02e-04
ω	-8.10e-07	2.97e-07	-1.68e-06	5.95e-07
α	3.39e-06	2.83e-07	2.58e-06	4.31e-06
β	0.845	1.20e-02	0.804	0.879
γ	197.54	3.76	193.59	219.44

Figure 21: Monte Carlo: GARCH-in-Mean estimation (simulation = 1000)

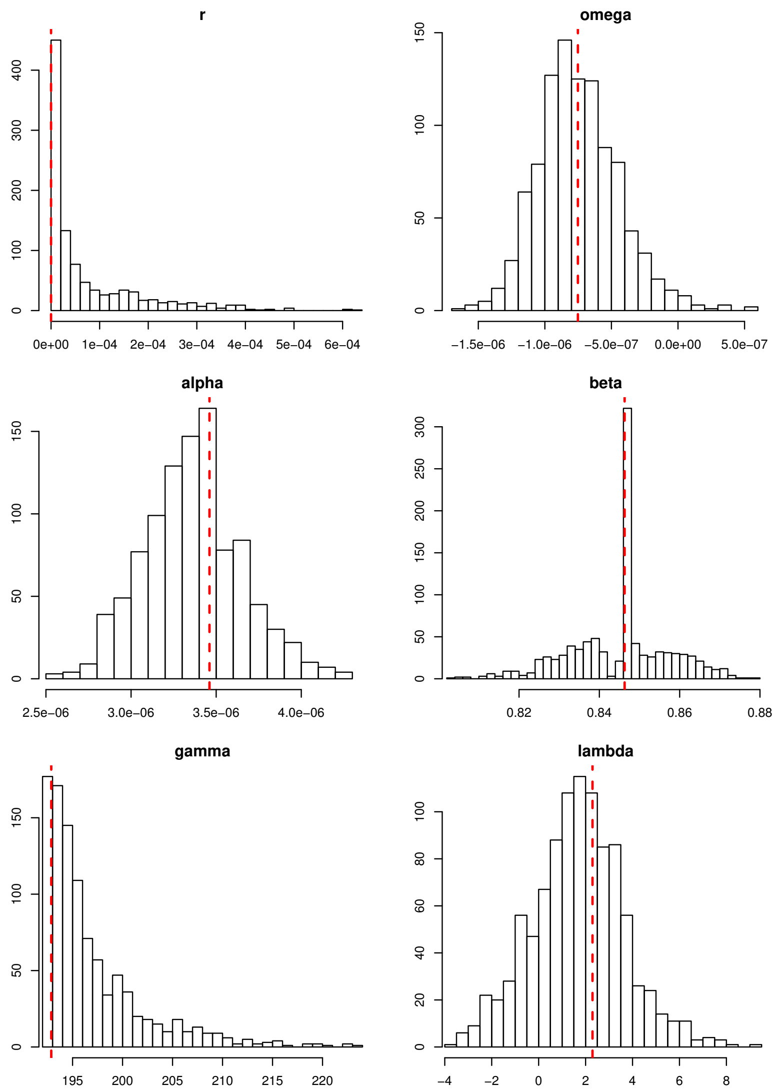
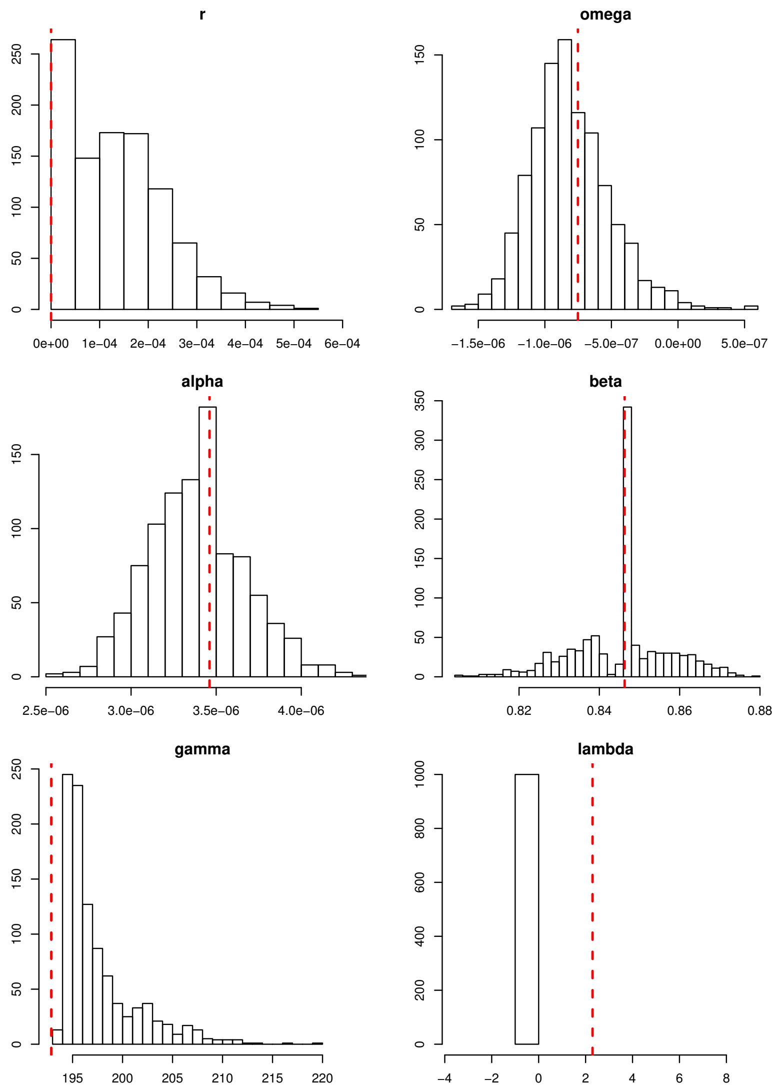


Figure 22: Monte Carlo: GARCH estimation results

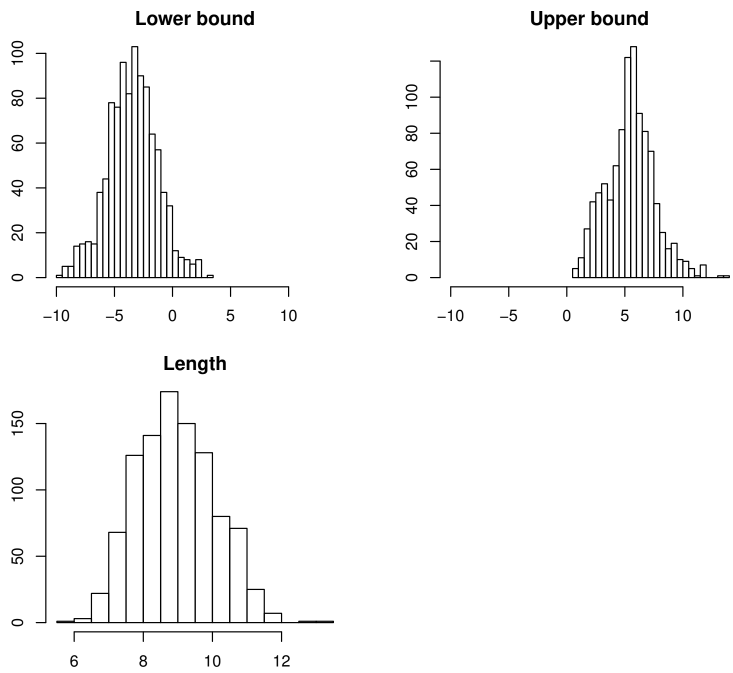


3.4.2 Identification issue for λ: is $H_0 : \lambda = 0$ rejected?

For each simulation, we conduct a Likelihood-Ratio (LR) test, The null hypothesis that $H_0 : \lambda = 0$ is rejected at 5% level for 55 times out of 1000 simulations. In other words, we fail to reject it 945 times out of 1000. It means that, we tend to fail to reject $\lambda = 0$ when the true value is actually $\lambda = 2.295$.

We also construct a confidence interval of λ for each simulation by inverting the Likelihood-Ratio test, by estimating the model imposing $\lambda = -10.0, -9.9, \cdots, 9.9, 10.0$. The histograms of the lower bound, upper bound and the length of the confidence intervals are reported in Figure 23. On average, the confidence interval is from -0.50 until 4.17 and the average length is 4.67. It implies that the confidence interval for λ is not tight and it is difficult to identify it.

Figure 23: Monte Carlo: Confidence interval of λ



3.4.3 VaR

We show the VaR computed with estimated parameter values. Table 12 reports the sample mean of VaR for each horizon. The highlighted two columns are the sample mean of VaR with estimated parameters for GARCH-in-Mean and GARCH models.

First, comparing the highlighted columns with the oracle VaR, the deviation of estimated VaR from the oracle VaR increases as the horizon increases. Also, the deviation is larger when the state variable h_t takes larger values. Second, the deviation of GARCH-in-Mean-based VaR is much smaller than GARCH-based VaR.

Figure 24 show the average value of GARCH-in-Mean-based VaR superimposed on the oracle

VaR. For any horizon, the two lines are close. On the other hand, the sample mean of GARCH-based VaR deviates from the oracle VaR.

Finally, Figures 25 and 26 plot 1000 simulated VaRs with GARCH-in-Mean (Figure 25) and GARCH (Figure 26). With the GARCH-in-Mean, the VaRs are centered at the oracle VaR whereas the GARCH-based VaRs are systematically above the oracle VaR.

Table 12: VaR with estimated parameters

h_t	Horizon	Oracle	Garch-in-Mean			Garch		
			Mean	Diff	SD	Mean	Diff	SD
Q5%	1	24.0	**23.9**	23.5	24.2	23.9	23.5	24.2
	10	107.0	105.9	99.3	112.6	**107.3**	101.4	113.2
	22	178.5	**177.0**	159.9	193.8	182.1	169.3	195.1
	66	346.6	**347.3**	285.4	408.2	372.5	339.5	403.4
	125	473.1	**479.7**	358.9	601.2	532.9	478.5	578.9
	250	610.0	**626.0**	410.2	845.2	724.5	631.1	795.5
Q20%	1	31.7	31.6	31.2	32.0	**31.7**	31.3	32.0
	10	125.8	**125.1**	117.7	132.6	127.4	121.9	132.9
	22	200.0	**199.4**	180.6	217.9	206.4	194.5	218.1
	66	364.1	**366.6**	299.8	433.6	395.5	363.5	424.2
	125	484.2	**492.8**	366.6	619.8	549.8	495.6	595.0
	250	615.5	**632.8**	413.3	855.8	734.3	640.9	804.6
Mean	1	48.8	**48.8**	48.2	49.5	49.0	48.7	49.3
	10	168.9	**169.1**	157.9	179.5	173.9	168.9	178.6
	22	251.2	**252.7**	225.5	278.8	264.8	253.9	274.8
	66	410.0	**417.3**	333.7	498.8	456.3	425.1	482.3
	125	516.1	**530.1**	386.7	679.1	598.3	543.2	642.1
	250	631.9	**653.6**	419.1	891.6	764.3	671.4	835.5
Q80%	1	58.2	**58.3**	57.4	59.1	58.7	58.3	59.0
	10	192.6	**193.5**	179.1	206.9	200.0	195.1	204.5
	22	280.0	**282.7**	249.3	314.0	298.1	287.1	307.7
	66	437.5	**447.5**	351.0	543.0	493.2	460.3	519.7
	125	536.4	**553.8**	397.3	714.9	629.5	572.3	673.9
	250	643.3	**667.7**	421.2	917.1	784.9	691.0	856.7
Q95%	1	75.3	**75.4**	73.8	76.8	76.2	75.7	76.5
	10	235.1	**237.2**	216.5	256.9	247.3	242.3	251.6
	22	331.0	**336.3**	290.3	380.6	358.6	347.4	368.2
	66	487.8	**503.1**	384.4	623.2	562.0	527.9	589.7
	125	575.5	**599.3**	415.2	789.3	690.3	632.3	737.3
	250	666.5	**696.7**	423.4	966.2	827.9	733.4	901.2

(Million dollars)

Figure 24: Mean of value-at-risk with estimated parameters , GARCH-in-Mean vs GARCH models

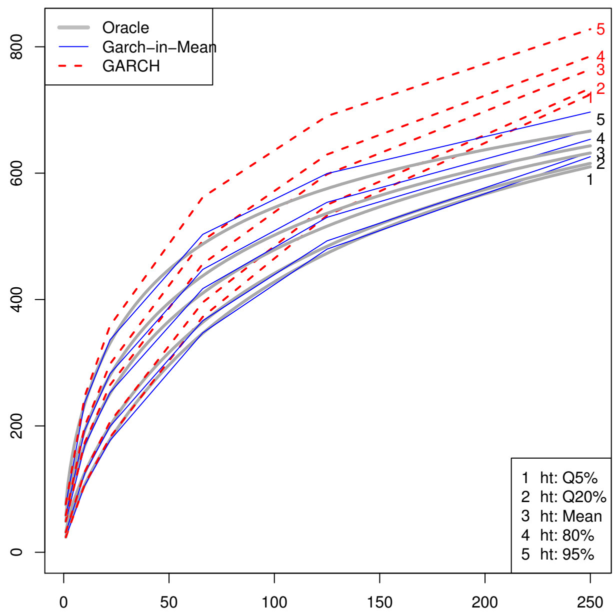
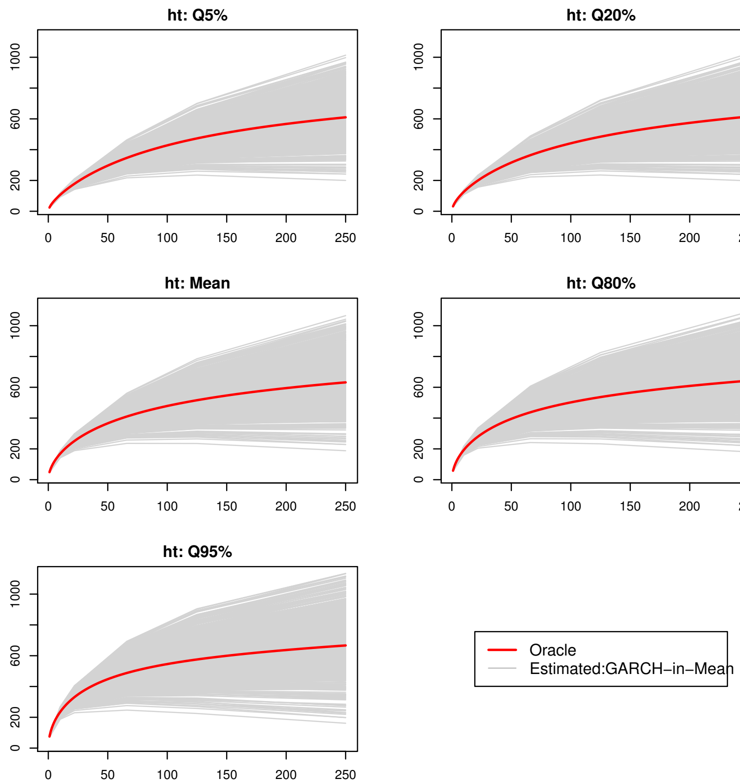
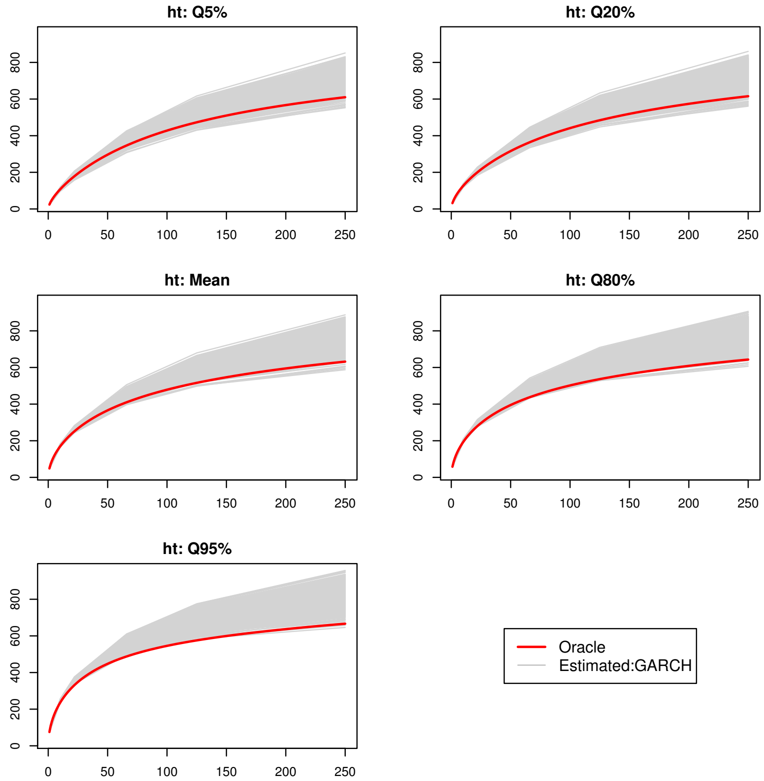


Figure 25: VaR with Garch-in-Mean model: distribution



Note: Each panel plot 1000 simulated VaRs with the GARCH-in-Mean estimates. The red line shows the oracle VaR

Figure 26: VaR with Garch model: distribution



Note: Each panel plot 1000 simulated VaRs with the GARCH estimates. The red line shows the oracle VaR

4 Model with Realized Variance

Christoffersen, Feunou, Jacobs, and Meddahi (2014) introduce an extension of Heston and Nandi (2000) which incorporates the dynamics of the Realized Variance (RV).

$$\begin{aligned} y_t &= r + \lambda \overline{h}_{t-1} + \sqrt{\overline{h}_{t-1}}\varepsilon_{1t} \\ RV_t &= h_{t-1}^{RV} + \sigma v(\varepsilon_{2t}) \end{aligned}$$

$$
\begin{aligned}
\overline{h_t} &= \kappa h_t^R + (1-\kappa) h_t^{RV} \\
h_t^R &= \omega_1 + \beta_1 h_{t-1}^R + \alpha_1 \left(\varepsilon_{1t} - \gamma_1 \sqrt{\overline{h}_{t-1}}\right)^2 \\
h_t^{RV} &= \omega_2 + \beta_2 h_{t-1}^{RV} + \alpha_2 \left(\varepsilon_{2t} - \gamma_2 \sqrt{\overline{h}_{t-1}}\right)^2
\end{aligned}
$$

with

$$
\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim i.i.d.N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)
$$

(h_t^R, h_t^{RV}) follows VAR(1) process:

$$
\begin{pmatrix} h_t^R \\ h_t^{RV} \end{pmatrix} = \begin{pmatrix} \omega_1 + \alpha_1 \\ \omega_2 + \alpha_2 \end{pmatrix} + \phi_1 \begin{pmatrix} h_{t-1}^R \\ h_{t-1}^{RV} \end{pmatrix} + \begin{pmatrix} \alpha_1 v(\epsilon_{1t}) \\ \alpha_2 v(\epsilon_{2t}) \end{pmatrix}
$$

with

$$
\phi_1 = \begin{pmatrix} \beta_1 + \alpha_1 \gamma_1^2 \kappa & \alpha_1 \gamma_1^2 (1-\kappa) \\ \alpha_2 \gamma_2^2 \kappa & \beta_2 + \alpha_2 \gamma_2^2 (1-\kappa) \end{pmatrix}
$$

and for $j = 1, 2,$

$$
\begin{aligned}
v(\epsilon_{jt}) &= \left(\epsilon_{jt} - \gamma_j \sqrt{\overline{h_t}}\right)^2 - 1 - \gamma_j^2 \overline{h_t} \\
E\left[v(\epsilon_{jt})\right] &= 0
\end{aligned}
$$

Therefore, given the largest eigenvalue of ϕ_1 has a modulus smaller than 1,

$$
E \begin{pmatrix} h_t^R \\ h_t^{RV} \end{pmatrix} = (I_2 - \phi_1)^{-1} \begin{pmatrix} \omega_1 + \alpha_1 \\ \omega_2 + \alpha_2 \end{pmatrix}
$$

4.1 Misspecified model

As in the previous chapter, we consider an agent who has a misspecified model without a time-varying conditional mean.

$$
\begin{aligned}
y_t &= \widetilde{r} + \sqrt{\overline{h}_{t-1}} \varepsilon_{1t} \\
\widetilde{r} &= r + E(\overline{h}_t)
\end{aligned}
$$

4.2 Parameter values to compute VaR

We follow the estimation results of Christoffersen, Feunou, Jacobs, and Meddahi (2014). They estimate the GARV model using daily close-to-close returns and realized variance data for the S&P 500 index for the period Jan. 2, 1990 to Dec. 31, 2010.

Parameters	Estimate	Std. Error
κ	0.395	(2.07e-2)
λ	0.97	(1.20)
α_1	4.61e-6	(3.57e-7)
β_1	9.67e-7	(5.56e-2)
γ_1	457	(21.1)
ω_1	5.74e-12	
α_2	2.57e-6	(2.23e-7)
β_2	4.07e-6	(6.60e-2)
γ_2	617	(45.4)
ω_2	5.84e-12	
σ	7.50e-6	(5.67e-7)
ρ	0.103	(9.50e-3)
r	Fixed and unknown	
$E(\overline{h}_t)$	1.19e-4	(1.34e-5)

We compute the VaR with the two cases with different state variables. We pick the same five values for h_t shown in Table 8. Since $E(h_t^R) \approx E(h_t^{RV})$, we set $h_t^R = h_t^{RV} = \overline{h_t}$ for every value of $\overline{h_t}$.

4.3 Term structure of VaR

The VaR below is computed using the parameter values estimated by CFJM(2014) with the risk-free rate 9.81e-05, which is the average of 3-month T-bill rate for this sample. In the figure, the red lines are with the misspecified constant mean model.

Table 13: VaR with GARCH model in comparison with GARCH-in-Mean model

h_t	Horizon (days)	GARCH-in-Mean A	GARCH B	Difference B-A
$Q5\%$	1	23.9	23.7	-0.2
	10	117.3	116.6	-0.7
	22	200.2	200.8	0.6
	66	396.9	406.9	10.0
	125	545.9	567.7	21.9
	250	707.1	745.4	38.3
$Q20\%$	1	31.7	31.5	-0.2
	10	136.6	136.6	0.0
	22	222.9	224.4	1.5
	66	415.1	427.3	12.2
	125	556.6	580.9	24.3
	250	712.9	752.9	40.0
Mean	1	48.9	48.9	-0.0
	10	180.8	182.1	1.3
	22	276.0	280.7	4.7
	66	464.0	481.5	17.5
	125	590.3	620.8	30.5
	250	731.8	775.9	44.0
$Q80\%$	1	58.4	58.5	0.1
	10	205.4	207.7	2.3
	22	306.2	312.9	6.7
	66	494.3	515.8	21.5
	125	613.1	647.6	34.5
	250	744.5	791.7	47.2
$Q95\%$	1	75.7	76.0	0.3
	10	249.3	253.5	4.2
	22	360.8	370.7	9.8
	66	549.4	578.1	28.7
	125	657.1	699.7	42.6
	250	772.2	825.9	53.7

(Million dollars)

5 Empirical studies

We now use the data and conduct an out-of-sample analysis. The data are the series of SPDR S&P 500 ETF (SPY). This is an ETF (exchange traded fund) which tracks the S&P 500 Index. The data spans daily from June 15, 2004 through June 13, 2014, amounting to 2497 observations. We consider the Heston-Nandi model and Realized Variance model, and compare them with the counterpart model without Garch-in-mean effect (denoted by λ). In order to study the forecast ability of the two models, we conduct a Diebold-Mariano test.

5.1 Heston-Nandi Model

First, we consider the model introduced by Heston and Nandi (2000). The "baseline" model refers to the original model:

$$\begin{aligned} y_t &= r + \lambda h_{t-1} + \sqrt{h_{t-1}} z_t \\ h_t &= \omega + \beta h_{t-1} + \alpha (z_t - \gamma \sqrt{h_{t-1}})^2 \\ z_t &\sim i.i.d.N(0,1) \end{aligned}$$

whereas the "model without the mean effect" refers to the simplified model with $\lambda = 0$:

$$\begin{aligned} y_t &= r + \sqrt{h_{t-1}} z_t \\ h_t &= \omega + \beta h_{t-1} + \alpha (z_t - \gamma \sqrt{h_{t-1}})^2 \\ z_t &\sim i.i.d.N(0,1) \end{aligned}$$

Also, we also consider the "model without the leverage effect", i.e., $\gamma = 0$. In this model, we set $Cov_t(y_{t+1}, h_{t+1}) = 0$.

$$\begin{aligned} y_t &= r + \lambda h_{t-1} + \sqrt{h_{t-1}} z_t \\ h_t &= \omega + \beta h_{t-1} + \alpha z_t^2 \\ z_t &\sim i.i.d.N(0,1) \end{aligned}$$

5.1.1 Out-of-sample analysis

First, we split the data and use the first 1997 observations (06/15/2004 - 06/07/2012) to estimate the parameters, and then forecast with the remaining 500 observations. Estimation results are shown in Table 19. We impose $r \geq 0$ with an assumption that the risk-free rate is nonnegative, and $\omega \geq 0$, $\beta \geq 0$ in order to guarantee the positivity of h_t.

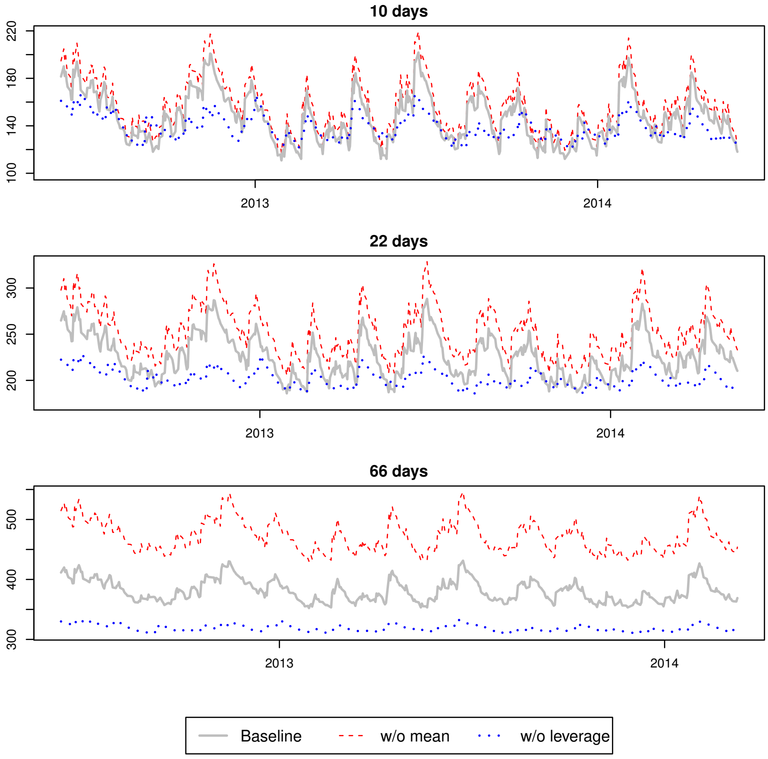
Table 14: estimation with first 1997 obs.

	Baseline			w/o mean		
Parameter	estimates	S.e.	t-stat	estimates	Std. error	t-stat
r	1.21e-15	9.94e-04	0.00	3.17e-16	3.27e-04	0.00
λ	3.17e+00	7.03e+00	0.45			
ω	1.60e-25	1.74e-06	0.00	1.03e-25	9.37e-07	0.00
α	3.63e-06	1.24e-06	2.92	3.81e-06	1.10e-06	3.47
β	8.10e-01	5.98e-02	13.54	8.15e-01	4.98e-02	16.37
γ	2.11e+02	3.97e+01	5.31	2.04e+02	3.63e+01	5.63
Log likelihood	6321.875			6320.288		
Obs	1997			1997		
Persistence	0.971			0.974		

	w/o leverage		
Parameter	Estimates	S.e.	t-stat
r	7.50e-05	5.79e-04	1.30e-01
λ	4.50	6.51e+00	6.91e-01
ω	2.51e-22	2.53e-06	9.94e-17
α	8.40e-06	3.89e-06	2.16e+00
β	9.37e-01	4.53e-02	2.07e+01
γ			
Log likelihood	6241.206		
Obs	1997		
Persistence	0.937		

Figure 27 shows the VaR computed with three different models, and Table 15 shows the average of VaR.

Figure 27: VaR with three different models (Forecast window = last 500 days)



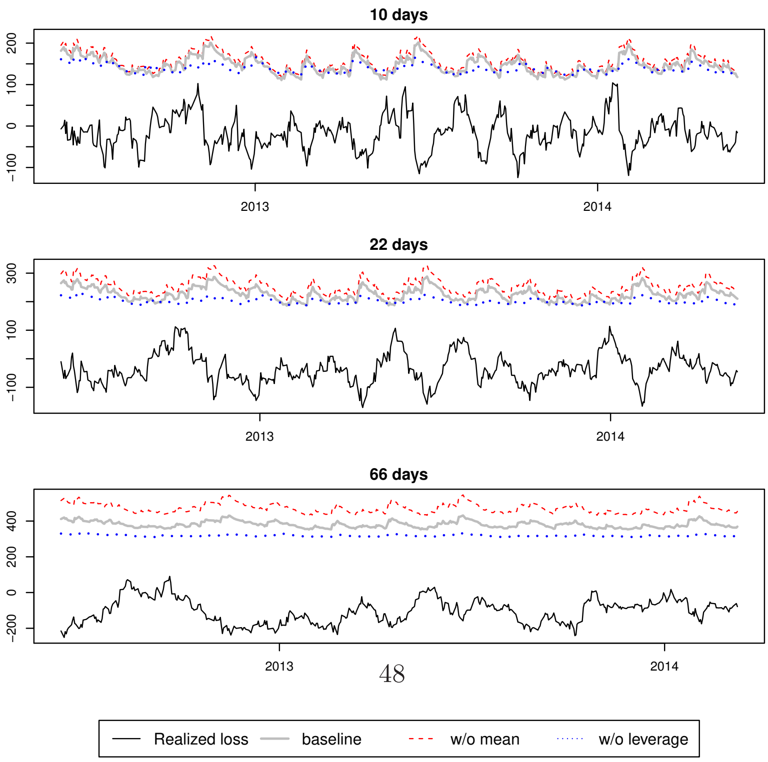


Table 15: Mean of VaR (forecast window = last 500 days)

		Horizon (days)					
		1	10	22	66	125	250
Baseline	(A)	38.79	145.90	224.48	380.21	480.29	577.74
w/o mean	(B)	39.33	155.93	250.68	470.27	643.38	850.36
Difference	(B) - (A)	0.54	10.03	26.19	90.07	163.08	272.63
w/o leverage	(C)	42.73	138.06	201.08	318.27	392.32	459.27
Difference	(C)-(A)	3.94	-7.84	-23.40	-61.94	-87.97	-118.47

(Million dollars)

Table 16: Violation rate (forecast window=last 500 days)

Horizon	Baseline	w/o mean	w/o leverage
1	1.60	1.00	0.80
10	0.00	0.00	0.00
22	0.00	0.00	0.00
66	0.00	0.00	0.00
125	0.00	0.00	0.00
250	0.00	0.00	0.00

(%)

In order to test statistically the forecast performance, we conduct the Diebold-Mariano test. Suppose $f_{1,t+1|t}$ is the 1% quantile forecast based on the baseline model, and $f_{2,t+1}$ is that based on another model. Following Giacomini and Komunjer (2005) we consider the "tick" loss function that is consistent for the 1% quantile, i.e., the 1% quantile (q) is a minimizer of the following loss function:

$$q = \arg\min_f E[L(f, y)] = E[-0.99(y - f)I_{\{y \leq f\}} + 0.01(y - f)I_{\{y > f\}}]$$

We use this loss function and consider the loss differences, denoted by d_t.[13]

$$d_{t+1} \quad = \quad L(f_{2,t+1|t}, y_{t+1}) - L(f_{1,t+1|t}, y_{t+1})$$

[13] As Patton (2016) points out, there are infinitely many loss functions that is consistent for the quantile, and the choice of the loss function may matter to evaluate the forecast performance.

where $I_{\{\}}$ is an indicator function. Note that we are studying the 1% VaR. Diebold and Mariano (1995) suggest a test the null hypothesis $H_0 : E(d_t) = 0$ under an assumption that $\{d_t\}$ is second-order stationary and have a short memory. Later Giacomini and White (2006) prove that the same test statistic can be used when we fix the estimation window that does not go to infinity, under a mild assumption (Theorem 4 of Giacomini and White (2006)).

we have

$$\sqrt{T}\,\overline{d} \xrightarrow{d} N(\mu, \Omega)$$

where $\overline{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$, $\mu = E(d_t)$, and $\Omega = \sum_{t=-\infty}^{\infty} \gamma_d(t)$ with $\gamma_d(t)$ being the auto-covariance function of $\{d_t\}$.

We can then conduct a test using a consistent estimator of Ω. We use the Bartlett kernel as was suggested by Newey and West (1987) and

$$\widehat{\Omega} = \widehat{Var}(d_t) + 2\sum_{l=1}^{L}\left(1 - \frac{l}{L+1}\right)\widehat{Cov}(d_t, d_{t+l})$$

According to the rule of thumb, the lag length L is computed as $0.75 \times T^{1/3} = 7.5$ with $T = 1000$. However The lag should be at least $h-1$ for $h-$step ahead forecast, so we use $L = h-1$ for $h = 10, 22$ and 66.

Table 17 shows the test results.

Table 17: Diebold-Mariano test: Baseline vs w/o mean

Horizon	Mean	s.e.	t-stat	p-value
1	-1.18e-06	1.50e-06	-0.785	0.78
10	5.33e-05	1.48e-06	35.9	0.00
22	1.46e-04	3.84e-06	38.0	0.00
66	5.56e-04	8.63e-06	64.5	0.00
125	1.10e-03	9.39e-06	117	0.00
250	2.05e-03	9.33e-06	220	0.00

Note: $\overline{d}$ is a sample average of loss difference between the two models, i.e., loss from the model without the mean minus the loss from the baseline model.

For horizons over 10 days, We reject the null hypothesis of an equal predictive ability of the two models, and the predictive ability of the baseline model is higher than the model without the mean effect.

Table 18 shows the test statistic to compare the predictive ability of the baseline model and the model without the leverage effect. Here, we fail to reject the null hypothesis.

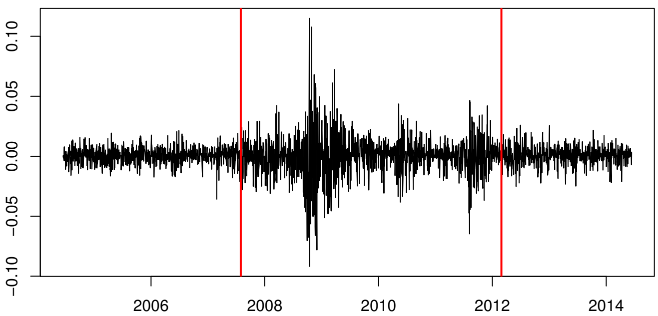
Table 18: Diebold-Mariano test: Baseline vs w/o leverage effect

Horizon	Mean	s.e.	t-stat	p-value
1	5.26E-06	1.06E-05	0.498	0.309
10	-4.33E-05	8.57E-06	-5.05	1.00
22	-1.31E-04	1.37E-05	-9.53	1.00
66	-3.65E-04	1.35E-05	-26.9	1.00
125	-5.46E-04	7.80E-06	-70.0	1.00
250	-7.75E-04	4.27E-06	-181	1.00

5.1.2 Before, during and after the crisis

Now we split the series into three periods, namely before, during and after the crisis. The before crisis period is from 2004-06-15 to 2007-07-31 (T = 782). The during crisis period is from 2007-08-02 to 2012-02-28 (T = 1145). Finally the after crisis period is from 2012-02-29 to 2014-06-13 (T = 570). Figure 28 shows the three periods shown in the graph of log returns.

Figure 28: Divided three periods



Estimation w/ 573 obs. until 2009-11-12, Forecast w/ 572 obs. Estimation w/ 285 obs. until 2013-04-23, Forecast w/ 285 obs.

5.1.3 Before crisis

For the before crisis period, we estimate the models with the first 391 observations until 2006-01-03 and forecast with the remaining 391 observations (from 2004-06-15 to 2006-01-03). Table 19 summarizes the estimation result. λ is estimated to be 9.56 which is larger than the estimation results in the literature, but it is not significant at 5% level (t-statistic is 0.09). If we test $\lambda = 0$ by the likelihood ratio test, we cannot reject it either. The leverage effect parameter γ is estimated to be 855 but it is not significant according to its t-statistic. However, we reject the null that $\gamma = 0$ by the likelihood ratio test.

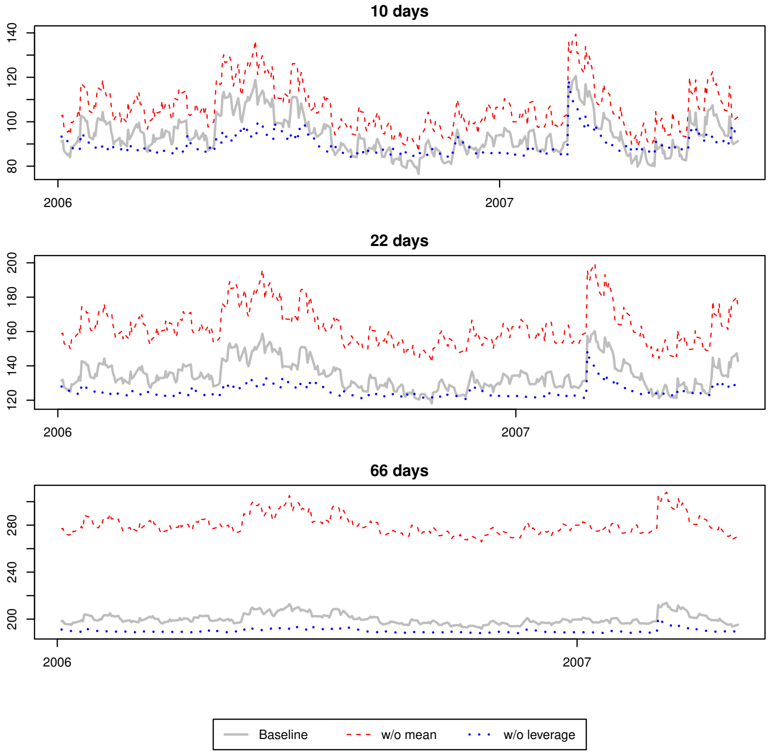
Table 19: estimation with before-crisis period

	Baseline			w/o mean		
Parameter	estimates	S.e.	t-stat	estimates	S.e.	t-stat
r	0.00e+00	4.29e-03	0.00e+00	0.00e+00	3.28e-04	0.00e+00
λ	9.56e+00	1.04e+02	9.22e-02			
ω	1.95e-06	1.68e-06	1.16e+00	2.28e-06	1.52e-06	1.50e+00
α	4.24e-07	2.09e-06	2.03e-01	4.30e-07	6.40e-07	6.72e-01
β	6.32e-01	1.35e+00	4.70e-01	6.12e-01	4.39e-01	1.39e+00
γ	8.55e+02	3.99e+03	2.14e-01	8.75e+02	1.20e+03	7.32e-01
Log likelihood	1420.681			1419.912		
Obs.	391			391		
Persistence	0.942			0.941		

	w/o leverage		
Parameter	estimates	S.e.	t-stat
r	0.00e+00	8.45e-03	0.00e+00
λ	9.42e+00	2.01e+02	4.68e-02
ω	3.46e-06	1.19e-05	2.92e-01
α	2.01e-06	1.44e-06	1.39e+00
β	8.72e-01	2.51e-01	3.47e+00
γ			
Log likelihood	1413.151		
Obs	391		
Persistence	0.872		

Figure 29 shows the VaR computed with three different models for horizons 10 days, 22 days and 66 days. Most of the time, the model without the mean effect yields the highest VaR in million dollars and the model without the leverage effect yields the lowest VaR. vaR from the model without leverage effect fluctuates less than the other VaR. The lower panel shows the same graph with the actual loss superimposed. There are several violations for horizons of 10 days and 22 days, in 2006 and 2007. Table 20 shows the average amount of VaR for the three models. The difference between the baseline model and the model without the mean effect is larger than the difference between the baseline model and the model without the leverage effect.

Figure 29: VaR for before-crisis period



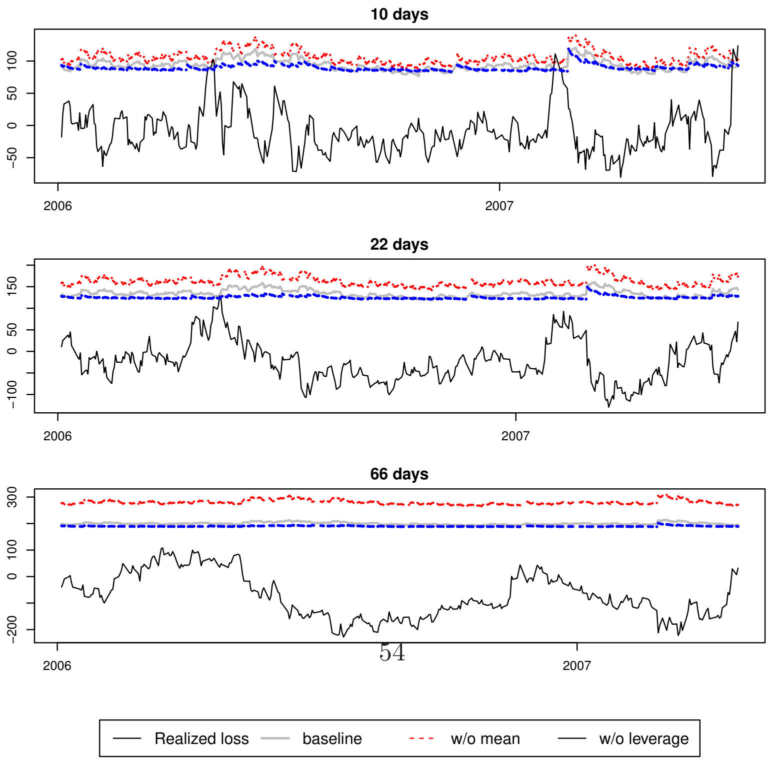


Table 20: Mean of VaR (Before the crisis)

		Horizon (days)					
		1	10	22	66	125	250
Baseline	(A)	28.75	94.33	134.76	199.98	234.49	257.32
w/o mean	(B)	29.64	106.43	162.92	280.08	370.48	492.87
Difference	(B) - (A)	0.89	12.10	28.15	80.10	135.99	235.55
w/o leverage	(C)	30.27	89.95	125.58	190.12	231.51	267.36
Difference	(C)-(A)	1.52	-4.38	-9.18	-9.86	-2.98	10.05

(Million dollars)

Table 21 shows the fraction of days where the actual loss exceeds the VaR. For horizons 1 day and 10 days, the actual loss exceeds the VaR for days more than 1%, which is the target level. However, for horizons above 22 days, there are less violations than 1%.

Table 21: Violation rate (Before the crisis)

Horizon	Baseline	w/o mean	w/o leverage
1	3.08	2.82	2.82
10	2.89	1.05	2.62
22	0.00	0.00	0.27
66	0.00	0.00	0.00
125	0.00	0.00	0.00
250	0.00	0.00	0.00

(%)

Table 22 shows the Diebold-Mariano test results. For horizons over 22 days, we reject the null hypothesis that the two have the equal predictive ability in expectation.

Table 22: Diebold-Mariano test: Baseline vs w/o mean (Before the crisis)

Horizon	Mean	s.e.	t-stat	p-value
1	-9.79e-06	4.11e-06	-2.38e+00	9.91e-01
10	-4.01e-05	5.50e-05	-7.28e-01	7.67e-01
22	1.47e-04	2.90e-06	5.07e+01	0.00e+00
66	4.42e-04	5.62e-06	7.86e+01	0.00e+00
125	7.76e-04	6.07e-06	1.28e+02	0.00e+00
250	1.42e-03	5.76e-06	2.46e+02	0.00e+00

Note: $\overline{d}$ is a sample average of loss difference between the two models, i.e., loss from the model without the mean minus the loss from the baseline model.

Table 23 shows the results to test equal predictive ability of the baseline model versus the model without the leverage effect.

Table 23: Diebold-Mariano test: Baseline vs w/o leverage effect (Before the crisis)

Horizon	Mean	s.e.	t-stat	p-value
1	-1.13e-05	1.01e-05	-1.11e+00	8.67e-01
10	-2.83e-05	1.72e-05	-1.65e+00	9.50e-01
22	-3.85e-05	9.06e-06	-4.25e+00	1.00e+00
66	-5.40e-05	4.70e-06	-1.15e+01	1.00e+00
125	-1.68e-05	2.25e-06	-7.49e+00	1.00e+00
250	5.57e-05	1.10e-07	5.07e+02	0.00e+00

5.1.4 During crisis

For the during crisis period, we estimate the models with the first 573 observations from 2007-08-02 until 2009-11-12 and forecast with the remaining 572 observations (from 2009-11-13 to 2012-02-28). Table 24 summarizes the estimation results. λ is estimated to be 0.56 which is much lower than the estimated λ before the crisis period, but it is still positive value. It means that, when the volatility is high the expected return is higher but not as high as before-crisis period. The leverage effect parameter γ is estimated to be 197 and it is significant by its t-statistic.

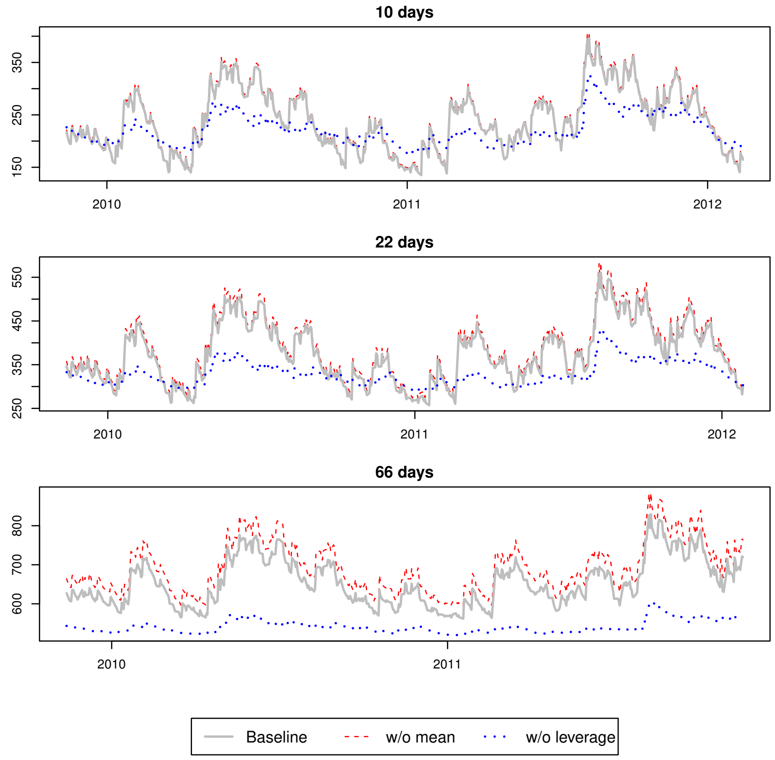
Table 24: Estimation with during-crisis period

Parameter	Baseline			w/o mean		
	Estimates	S.e.	t-stat	Estimates	S.e.	t-stat
r	0.00e+00	3.13e-03	0.00e+00	0.00e+00	1.18e-03	0.00e+00
λ	5.67e-01	8.01e+00	7.07e-02			
ω	4.90e-19	1.41e-05	3.48e-14	2.85e-19	7.79e-06	3.66e-14
α	7.43e-06	6.49e-06	1.14e+00	7.63e-06	5.35e-06	1.43e+00
β	6.86e-01	2.20e-01	3.11e+00	6.91e-01	1.42e-01	4.85e+00
γ	1.97e+02	5.66e+01	3.48e+00	1.93e+02	5.47e+01	3.53e+00
Log likelihood	1540.329			1540.274		
Obs.	573			573		
Persistence	0.974			0.975		

Parameter	w/o leverage		
	Estimates	S.e.	t-stat
r	6.56e-15	2.05e-03	3.19e-12
λ	1.36e+00	8.50e+00	1.60e-01
ω	7.02e-83	8.54e-06	8.22e-78
α	2.09e-05	8.63e-06	2.42e+00
β	9.37e-01	4.50e-02	2.08e+01
γ			
Log likelihood	1513.23		
Obs	573		
Persistence	0.937		

Figure 30 shows the VaR forecast with three models. Because the estimated λ is close to zero, there is not large difference between the baseline and without mean effect. VaR with the model without the leverage effect is more smooth and overall it is below the other two VaR. Table 25 shows the average of VaR. The average VaR is twice as large as VaR computed for before-crisis period.

Figure 30: VaR forecast: during crisis: 2009-11-13 to 2012-02-28 (T=572)



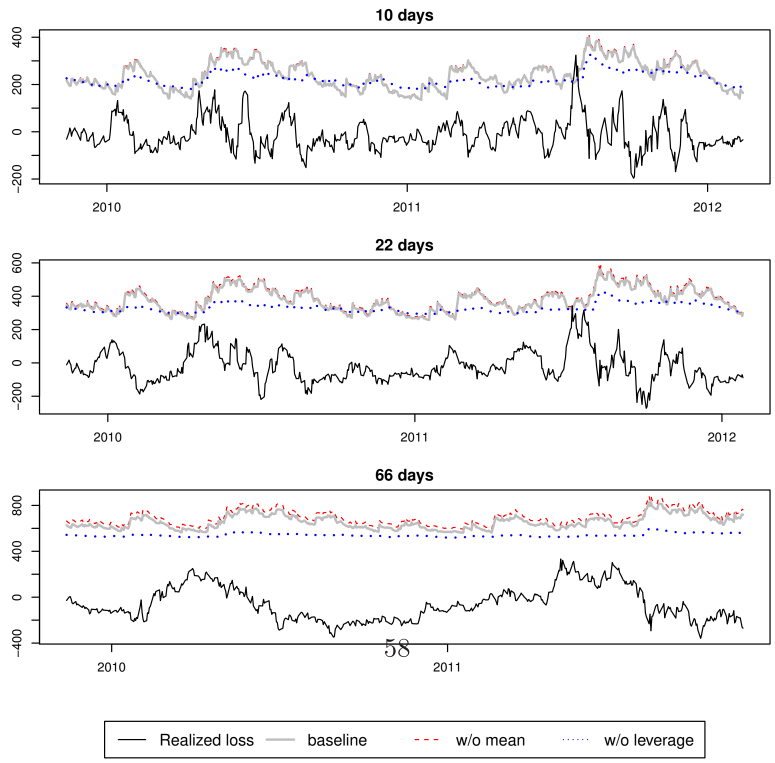


Table 25: Mean of VaR (During the crisis)

	Horizon (days)					
	1	10	22	66	125	250
Baseline	60.60	237.07	371.57	652.15	846.76	1055.71
w/o mean	60.74	241.55	383.34	690.65	912.17	1153.86
Difference	0.14	4.48	11.78	38.50	65.41	98.16
w/o leverage	65.80	220.42	328.05	541.13	692.89	872.55
Difference	5.20	-16.65	-43.52	-111.02	-153.87	-183.16

(Million dollars)

Table 26 shows the violation rate. For 1-day horizon, the violation rate is close to 1% with the baseline model and the model without the mean effect. The violation rate gradually reduces and finally becomes zero at 66 day horizon.

Table 26: Violation rate (during the crisis)

Horizon	Baseline	w/o mean	w/o leverage
1	1.75	1.75	0.35
10	0.89	0.89	0.71
22	0.18	0.18	0.18
66	0.00	0.00	0.00
125	0.00	0.00	0.00
250	0.00	0.00	0.00

(%)

Table 27 summaries the Diebold-Mariano test comparing the predictive ability of the baseline model and the model without the mean effect. At 5% level, we reject the null of equal predictive ability for horizons above 22 days. Table 28 shows the test results comparing the predictive ability of the baseline model and the model without the leverage effect. At 5% level, we reject the null for horizons above 22 days.

Table 27: Diebold-Mariano test: Baseline vs w/o mean

Horizon	Mean	s.e.	t-stat	p-value
1	-1.43e-06	8.76e-07	-1.63e+00	9.49e-01
10	5.68e-06	1.76e-05	3.22e-01	3.74e-01
22	6.32e-05	9.41e-06	6.72e+00	9.12e-12
66	2.83e-04	1.14e-05	2.49e+01	0.00e+00
125	5.55e-04	1.23e-05	4.51e+01	0.00e+00
250	1.03e-03	8.98e-06	1.14e+02	0.00e+00

Note: $\overline{d}$ is a sample average of loss difference between the two models, i.e., loss from the model without the mean minus the loss from the baseline model.

Table 28: Diebold-Mariano test: Baseline vs w/o leverage effect

Horizon	Mean	s.e.	t-stat	p-value
1	-5.13e-06	2.62e-05	-1.96e-01	5.78e-01
10	-6.21e-05	4.49e-05	-1.38e+00	9.17e-01
22	-2.88e-04	4.58e-05	-6.30e+00	1.00e+00
66	-7.86e-04	7.90e-05	-9.95e+00	1.00e+00
125	-1.19e-03	5.95e-05	-1.99e+01	1.00e+00
250	-1.65e-03	3.17e-05	-5.20e+01	1.00e+00

5.1.5 After crisis

For the after crisis period, we estimate the models with the first 285 observations from 2012-02-29 until 2013-04-23 and forecast with the remaining 285 observations (from 2013-04-24 to 2014-06-13). Table 29 summarizes the estimation result. λ is estimated to be 7.53 which is larger than the estimation results in the literature, but it is not significant at 5% level (t-statistic is 0.40). If we test $\lambda = 0$ by the likelihood ratio test, we fail to reject it either. The leverage effect parameter γ is estimated to be 160 but it is not significant according to its t-statistic. However, we reject the null that $\gamma = 0$ by the likelihood ratio test.

Table 29: estimation with after-crisis period

	Baseline			w/o mean		
Parameter	estimates	S.e.	t-stat	estimates	S.e.	t-stat
r	0.00e+00	1.93e-03	0.00e+00	0.00e+00	7.53e-04	0.00e+00
λ	7.53e+00	1.87e+01	4.02e-01			
ω	3.19e-147	3.41e-05	9.33e-143	7.36e-149	1.56e-05	4.72e-144
α	5.97e-06	4.31e-05	1.38e-01	6.20e-06	1.92e-05	3.23e-01
β	7.65e-01	1.60e+00	4.77e-01	7.68e-01	6.92e-01	1.11e+00
γ	1.60e+02	1.43e+03	1.12e-01	1.62e+02	5.96e+02	2.71e-01
Log likelihood	975.094			974.469		
Obs.	285			285		
Persistence	0.919			0.930		

	w/o leverage		
Parameter	estimates	S.e.	t-stat
r	0.00e+00	3.46e-03	0.00e+00
λ	1.33e+01	4.43e+01	3.01e-01
ω	5.81e-44	7.90e-06	7.36e-39
α	9.23e-06	1.30e-05	7.10e-01
β	8.72e-01	5.68e-02	1.54e+01
γ			
Log likelihood	966.331		
Obs	285		
Persistence	0.872		

Figure 31 shows the VaR forecast with three models, and Table 30 shows the average of VaR. The baseline VaR is smaller than the VaR without the mean effect but larger than VaR without the leverage effect except for 1 day horizon. At 10 day horizon, the difference is 21.64 and 23.34 million dollars.

Figure 31: VaR forecast after the crisis: 2013-04-24 - 2014-06-13 (T = 285)

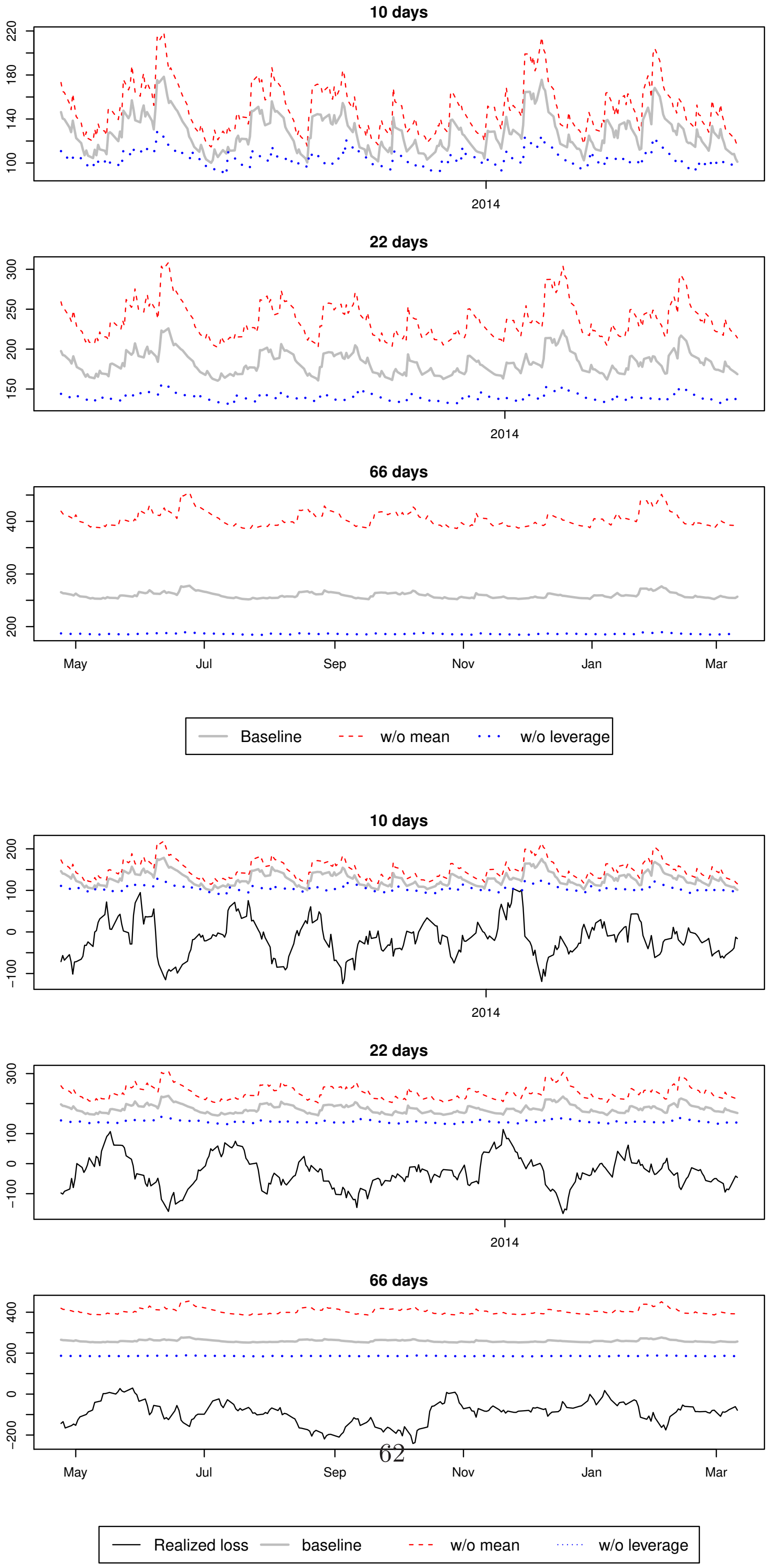


Table 30: Mean of VaR (After the crisis)

	Horizon (days)					
	1	10	22	66	125	250
Baseline	34.05	127.06	181.70	259.32	294.72	309.15
w/o mean	34.93	148.70	234.73	404.34	524.81	679.40
Difference	0.88	21.64	53.04	145.02	230.10	370.25
w/o leverage	34.62	103.71	139.31	186.11	192.87	140.11
Difference	0.57	-23.34	-42.39	-73.21	-101.85	-169.04

(Million dollars)

Table 31 shows the violation rate. It exceeds 1% at 1 day horizon but is is lower than 1% for horizons above 10 days.

Table 31: Violation rate (After the crisis)

Horizon	Baseline	w/o mean	w/o leverage
1	1.41	1.41	2.11
10	0.00	0.00	0.73
22	0.00	0.00	0.00
66	0.00	0.00	0.00
125	0.00	0.00	0.00
250	0.00	0.00	0.00

(%)

Table 32 shows the Diebold-Mariano test comparing the predictive ability of the baseline model and the model without the mean effect, and Table 33 to compare the baseline model and the model without leverage effect. For both test, for horizons above 10 days, we reject the null hypothesis of equal predictive ability.

Table 32: Diebold-Mariano test: Baseline vs w/o mean

Horizon	Mean	s.e.	t-stat	p-value
1	-3.09e-06	3.96e-06	-7.80e-01	7.82e-01
10	1.14e-04	4.16e-06	2.75e+01	0.00e+00
22	2.91e-04	7.67e-06	3.80e+01	0.00e+00
66	8.46e-04	6.54e-06	1.29e+02	0.00e+00
125	1.40e-03	6.69e-06	2.10e+02	0.00e+00
250	2.39e-03	8.78e-06	2.72e+02	0.00e+00

Note: $\overline{d}$ is a sample average of loss difference between the two models, i.e., loss from the model without the mean minus the loss from the baseline model.

Table 33: Diebold-Mariano test: Baseline vs w/o leverage effect

Horizon	Mean	s.e.	t-stat	p-value
1	1.87e-05	1.09e-05	1.71e+00	4.35e-02
10	-1.18e-04	9.52e-06	-1.24e+01	1.00e+00
22	-2.28e-04	8.94e-06	-2.55e+01	1.00e+00
66	-4.01e-04	3.01e-06	-1.33e+02	1.00e+00
125	-5.63e-04	1.85e-06	-3.05e+02	1.00e+00
250	-9.26e-04	8.95e-07	-1.04e+03	1.00e+00

5.1.6 Estimation with all observations

Table 34 shows the estimation result using all the sample. The mean parameter λ is estimated to be 4.11 and it is not significant according to its t-statistic. However, we reject the null that $\lambda = 0$ at 5% by the likelihood ratio test. The leverage effect parameter γ is estimated to be 235 and it is significant according to its t-statistic. Also we reject the null $\gamma = 0$ by the likelihood ratio test.

Table 34: Estimation with all observations

	Baseline			w/o mean		
Parameter	estimates	S.e.	t-stat	estimates	S.e.	t-stat
r	0.00e+00	7.36e-04	0.00e+00	0.00e+00	2.63e-04	0.00e+00
λ	4.11e+00	6.11e+00	6.72e-01			
ω	2.35e-18	1.30e-06	1.81e-12	4.62e-18	7.73e-07	5.98e-12
α	3.18e-06	9.75e-07	3.27e+00	3.37e-06	8.92e-07	3.77e+00
β	7.94e-01	5.34e-02	1.49e+01	8.00e-01	4.82e-02	1.66e+01
γ	2.35e+02	4.17e+01	5.64e+00	2.28e+02	3.78e+01	6.03e+00
Log likelihood	8079.713			8076.810		
Obs.	2497			2497		
Persistence	0.971			0.975		

	w/o leverage		
Parameter	estimates	S.e.	t-stat
r	0.00e+00	5.38e-04	0.00e+00
λ	6.86e+00	6.67e+00	1.03e+00
ω	5.90e-19	2.05e-06	2.87e-13
α	7.62e-06	3.11e-06	2.45e+00
β	9.33e-01	4.22e-02	2.21e+01
γ			
Log likelihood	7972.523		
Obs	2497		
Persistence	0.933		

5.2 Realized Variance Model

We consider the model introduced by Christoffersen, Feunou, Jacobs, and Meddahi (2014). The "baseline" model refers to the original model,

$$\begin{aligned}
\overline{h_t} &= \kappa h_t^R + (1-\kappa) h_t^{RV} \\
h_t^R &= \omega_1 + \beta_1 h_{t-1}^R + \alpha_1 \left(\varepsilon_{1t} - \gamma_1 \sqrt{\overline{h}_{t-1}}\right)^2 \\
h_t^{RV} &= \omega_2 + \beta_2 h_{t-1}^{RV} + \alpha_2 \left(\varepsilon_{2t} - \gamma_2 \sqrt{\overline{h}_{t-1}}\right)^2
\end{aligned}$$

with

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim i.i.d.N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

the "model without the mean effect" sets $\lambda = 0$ and the "model without the levreage effect" set $\gamma_1 = \gamma_2 = 0$.

The estimation is done by QMLE, assuming that y_t and RV_t are, conditional on the information at $t-1$, jointly normally distributed.

$$\left(\begin{array}{c} y_t \\ RV_t \end{array} \middle| I_{t-1} \right) \sim \left(\begin{pmatrix} r + \lambda \overline{h_{t-1}} \\ h_{t-1}^{RV} \end{pmatrix}, \begin{pmatrix} \overline{h_{t-1}} & -2\rho\gamma_2\sigma\overline{h_{t-1}} \\ -2\rho\gamma_2\sigma\overline{h_{t-1}} & 2\sigma^2(1 + 2\gamma_2^2\overline{h_{t-1}}) \end{pmatrix} \right)$$

Note that by setting $\gamma_2 = 0$ for the model without the leverage effect, we assume that the conditional covariance between the y_t and RV_t are zero, and ρ is not identified.

5.2.1 Out-of-sample analysis

We split the data and use the first 1997 observations (06/15/2004 - 06/07/2012) to estimate the parameters, and then forecast with the remaining 500 observations. Estimation results are shown in Table 35. We impose $r \geq 0$ with an assumption that the risk-free rate is nonnegative, and $\omega_j \geq 0$, $\beta_j \geq 0$ for $j = 1, 2$ in order to guarantee the positivity of h_t.

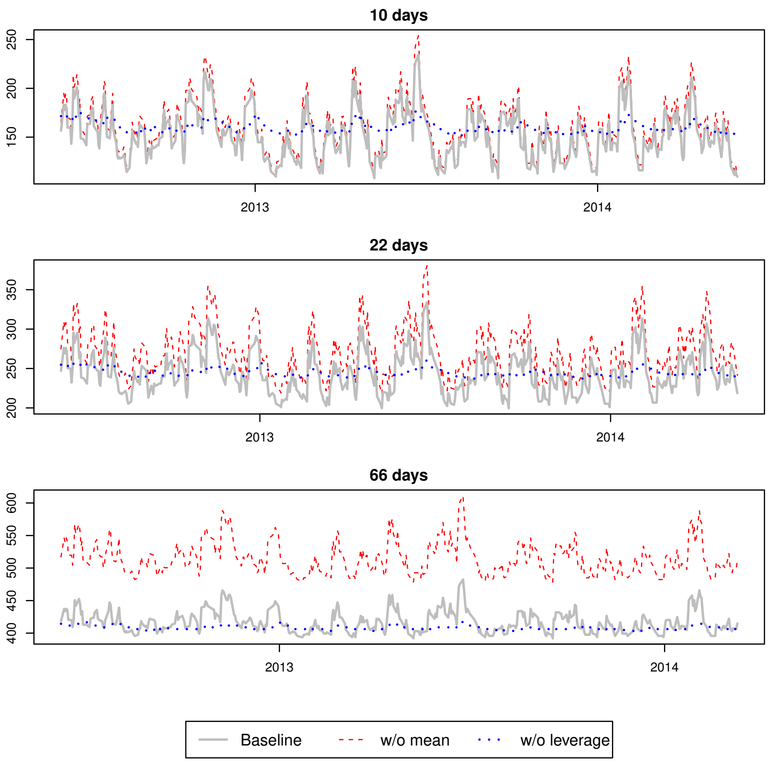
Table 35: estimation with first 1997 observations

	Baseline			w/o mean		
Parameter	Estimates	S.e.	t-stat	Estimates	S.e.	t-stat
κ	2.37e-02	2.71e-02	8.72e-01	3.30e-02	3.34e-02	9.90e-01
$\alpha1$	1.21e-04	1.40e-04	8.62e-01	8.54e-05	8.74e-05	9.77e-01
$\beta1$	2.39e-05	7.66e-02	3.11e-04	1.82e-15	7.70e-02	2.36e-14
$\gamma1$	3.56e+02	2.59e+01	1.38e+01	3.62e+02	2.66e+01	1.36e+01
$\omega1$	5.24e-100	2.22e-04	2.36e-96	6.39e-113	1.42e-04	4.48e-109
$\alpha2$	2.50e-06	2.84e-07	8.81e+00	2.53e-06	2.87e-07	8.82e+00
$\beta2$	0.00e+00	6.77e-02	0.00e+00	0.00e+00	6.99e-02	0.00e+00
$\gamma2$	4.94e+02	4.99e+01	9.90e+00	4.93e+02	4.95e+01	9.97e+00
$\omega2$	2.75e-11	5.24e-06	5.24e-06	9.62e-17	5.10e-06	1.89e-11
σ	6.35e-06	5.94e-07	1.07e+01	6.38e-06	5.98e-07	1.07e+01
ρ	2.05e-01	1.93e-02	1.06e+01	2.03e-01	1.90e-02	1.07e+01
λ	2.50e+00	1.97e+00	1.27e+00			
r	1.41e-224	2.02e-04	6.97e-221	1.34e-229	1.75e-04	7.67e-226
Log likelihood	24753.86			24752.63		
Obs.	1997			1997		

	w/o leverage		
Parameter	estimates	S.e.	t-stat
κ	5.12e-01	6.39e-02	8.02e+00
$\alpha 1$	6.80e-06	2.58e-06	2.63e+00
$\beta 1$	9.27e-01	2.75e-02	3.38e+01
$\gamma 1$			
$\omega 1$	6.72e-18	3.05e-06	2.20e-12
$\alpha 2$	4.18e-05	5.95e-07	7.02e+01
$\beta 2$	8.40e-01	1.31e-02	6.42e+01
$\gamma 2$			
$\omega 2$	1.88e-49	1.11e-05	1.69e-44
σ	1.34e-04	3.49e-07	3.83e+02
ρ			
λ	-1.06e-01	3.05e+00	-3.46e-02
r	3.68e-04	4.59e-04	8.02e-01
Log likelihood	22237.11		
Obs	1997		

The mean parameter λ is estimated to be 2.50 which is smaller than the estimates with the Heston-Nandi model and it is not significant according to its t-statistic. Also we fail to reject the null hypothesis that $\lambda = 0$ by the likelihood ratio test. The leverage effect parameter γ_1 and γ_2 are estimated to be 356 and 494 respectively and these are both significant according to its t-statistics. Also we reject the null hypothesis that $\gamma_1 = \gamma_2 = 0$ by the likelihood ratio test. Figure 32 shows the VaR with three different models, and Table 36 shows their average value. The baseline VaR is, on average, smaller than the VaR without the mean effect. This is the same pattern as in the Heston-Nandi model. The difference at 10 horizon is 8.74 million dollars. When we compare the baseline VaR and the VaR without the leverage effect exhibits nonmonotone pattern. VaR without the leverage effect is larger on average except for the 66 day horizon.

Figure 32: VaR with three different models (Forecast window = last 500 days)



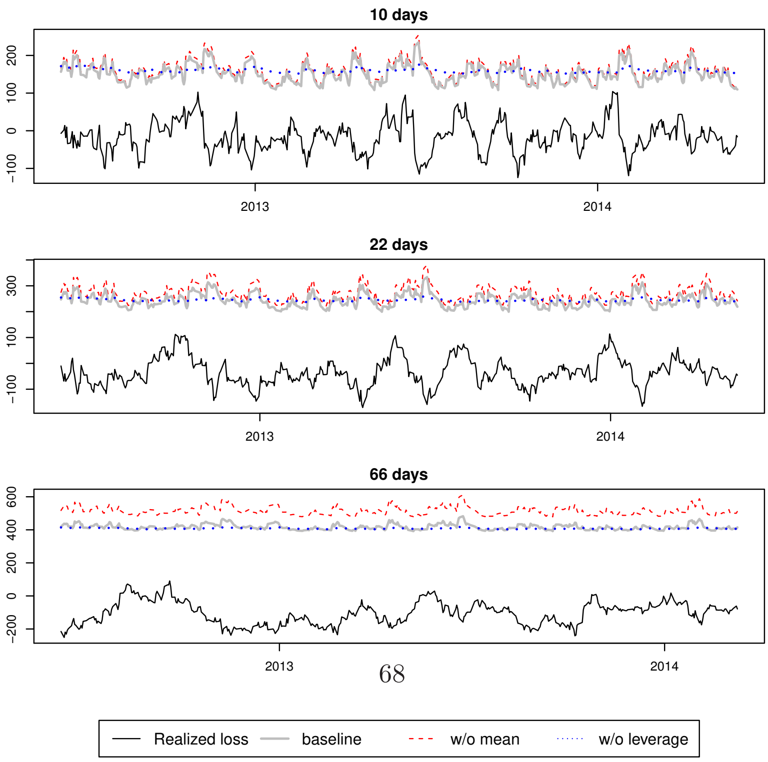


Table 36: Mean of VaR (forecast window = last 500 days)

		Horizon (days)					
		1	10	22	66	125	250
Baseline	(A)	33.87	151.82	242.77	416.95	525.14	637.38
w/o mean	(B)	34.08	160.57	269.52	514.83	694.54	899.19
Difference	(B)-(A)	0.21	8.74	26.76	97.88	169.40	261.81
w/o leverage	(C)	42.52	159.53	244.47	407.82	526.63	671.66
Difference	(C)-(A)	8.65	7.71	1.70	-9.13	1.49	34.28

(Million dollars)

Table 37 shows the violation rate. It is close to 1% at 1 day horizon but it becomes zero for horizons larger than 10 days.

Table 37: Violation rate (forecast window=last 500 days)

Horizon	Baseline	w/o mean	w/o leverage
1	1.80	1.80	1.00
10	0.00	0.00	0.00
22	0.00	0.00	0.00
66	0.00	0.00	0.00
125	0.00	0.00	0.00
250	0.00	0.00	0.00

(%)

Finally Table 38 shows the Diebold-Mariano test statistics comparing the predictive ability of the baseline model and the model without the mean effect. From 10 day horizons, we reject the null hypothesis of the equal predictive ability.

Table 38: Diebold-Mariano test: Baseline vs w/o mean

Horizon	Mean	S.e.	t-stat	p-value
1	-2.26e-07	4.79e-07	-4.71e-01	6.81e-01
10	4.66e-05	1.44e-06	3.24e+01	0.00e+00
22	1.51e-04	3.56e-06	4.23e+01	0.00e+00
66	6.19e-04	6.54e-06	9.46e+01	0.00e+00
125	1.18e-03	4.86e-06	2.43e+02	0.00e+00
250	2.05e-03	3.04e-06	6.75e+02	0.00e+00

Note: $\overline{d}$ is a sample average of loss difference between the two models, i.e., loss from the model without the mean minus the loss from the baseline model.

Table 39 shows the test statistic to compare the predictive ability of the baseline model and the model without the leverage effect. The result is not monotone: we reject the null hypothesis for horizons 10 days and above 66 days.

Table 39: Diebold-Mariano test: Baseline vs w/o leverage effect

Horizon	Mean	S.e.	t-stat	p-value
1	-1.24e-05	2.18e-05	-5.70e-01	7.16e-01
10	3.75e-05	1.15e-05	3.27e+00	5.38e-04
22	3.93e-06	1.42e-05	2.77e-01	3.91e-01
66	-5.66e-05	8.92e-06	-6.34e+00	1.00e+00
125	9.20e-06	3.73e-06	2.46e+00	6.87e-03
250	2.45e-04	1.53e-06	1.60e+02	0.00e+00

6 Conclusion

In this paper, we consider the effect of ignoring the time-variability of the conditional mean on the computation of the VaR. We show that, even though the constant mean may not be rejected at 5% level, we should estimate the model with the time-varying mean since they will have a large impact on the VaR.

7 Appendix

7.1 Other Calibration: constant volatility model

We first pick $\theta = 0.999$ (so that its counterpart of annually aggregated return is $\theta^{250} = 0.78$) and the annual $R^2 = 0.06$ (one of the posterior mode of PS(2012)). Making them as a benchmark, we vary the values of θ and the annual R^2, fixing the other parameters to be constant or keeping the first and second moments of y_t. The calibration designs are shown in Table 40.

Table 40: Calibration design

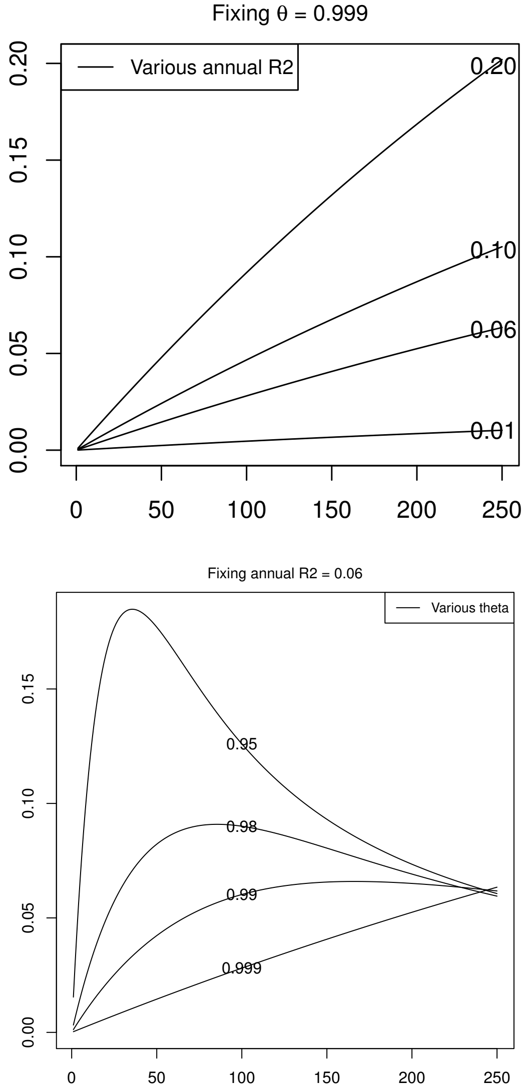
Fix $\theta = 0.999$	Annual $R^2 =$	0.01
		$\underline{0.06}$
		0.10
		0.20
Fix Annual $R^2 = 0.06$	$\theta =$	$\underline{0.999}$
		0.99
		0.98
		0.95

(underline is drawn for the benchmark calibration.)

In Table 40, the first four rows show the calibration fixing $\theta = 0.999$. We vary the annual R^2 to be 0.01, 0.06, 0.10 and 0.20. 0.01 and 0.20 are extreme values that are not likely according to PS(2012), and 0.10 is the other postrior mode of PS(2012). The second four rows shows the calibration fixing the annual $R^2 = 0.06$. We vary θ to be 0.999, 0.99, 0.98 and 0.95. We consider that μ_t is very persistent and believe that this range is reasonable.

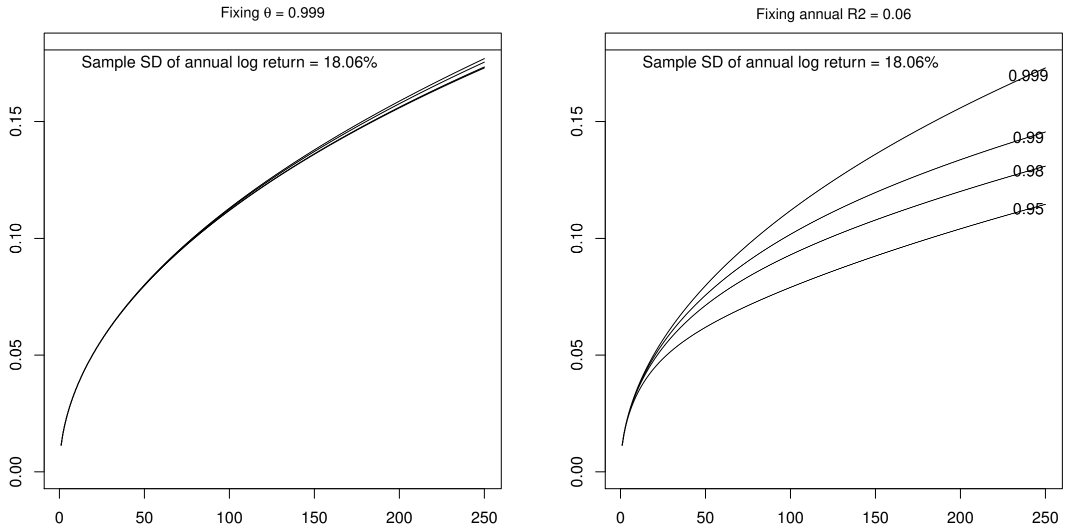
Figures 33 and 34 show some properties of different calibrations. Figure 33 As shown in Figure 33, the R^2 grows monotonically when $\theta = 0.999$, whereas it is hump-shaped. Figure 34 shows the unconditional variance of aggregated log return, y_t^h as a function of h. From the data, the variance of annual log return is about 18%. The figure shows that the variance of annual log return is the most close to the data when $\theta = 0.999$ and it is not sensitive to the annual R^2.

Figure 33: Term structure of R^2: R^2 of $y_t^{(h)}$ as a function of h



Note: R^2 is defined to be the variance of μ_t divided by the variance of y_t, which is the R^2 if we regress y_t on μ_t.

Figure 34: Variance of aggregated return $y_t^{(h)}$ as a function of h



7.1.1 Mean and Variance effect with various calibrations

In this section, let us compare the properties of $E(y_{t+1:t+\tau})$, $Var(y_{t+1:t+\tau})$ when (i) an agent have a misspecified iid model (M_{iid}), and (ii) an agent has the correct model and observes μ_t (full information). In all figures, the red and dashed lines correspond to case (i), and the black and solid lines correspond to the case (ii). For the full information case, the following five values for μ_t are chosen for the demonstration:

Table 41: Five values of μ_t to compute oracle VaR

μ_1	μ_2	μ_3	μ_4	μ_5
Mean-2sd	Mean - 1sd	Mean	Mean + 1sd	Mean + 2sd

where each values are drawn from the unconditional distribution, $\mu_t \sim N(\overline{\mu}, \sigma_\mu^2)$.

Figures 35 and 36 show $E(y_{t+1:t+\tau})$ as a function of τ with various calibrations. When $\theta = 0.999$, the mean grows almost linearly with the horizon. On the other hand, when θ takes smaller values, it converges quickly. The same pattern is observed by Figures 37 and 38.

Figure 35: Mean of $y_{t+1:t+\tau}$ as a function of τ: fixing $\theta = 0.999$

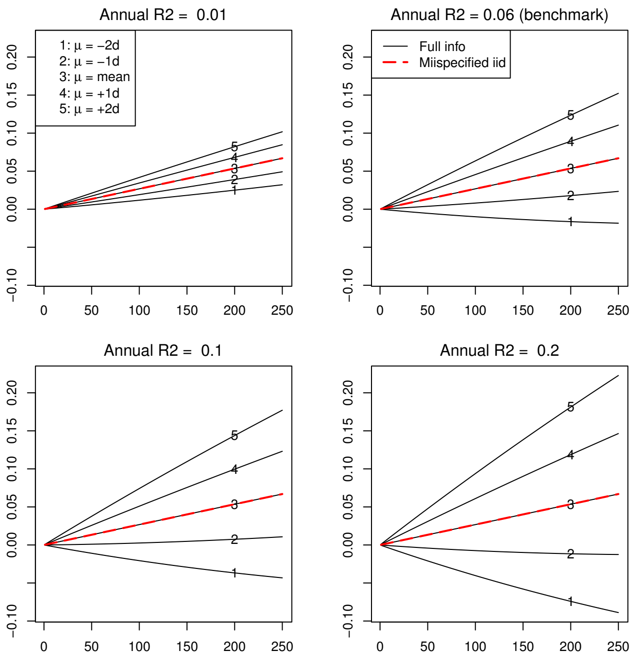


Figure 36: Mean of $y_{t+1:t+\tau}$ as a function of τ: fixing annual $R^2 = 0.06$

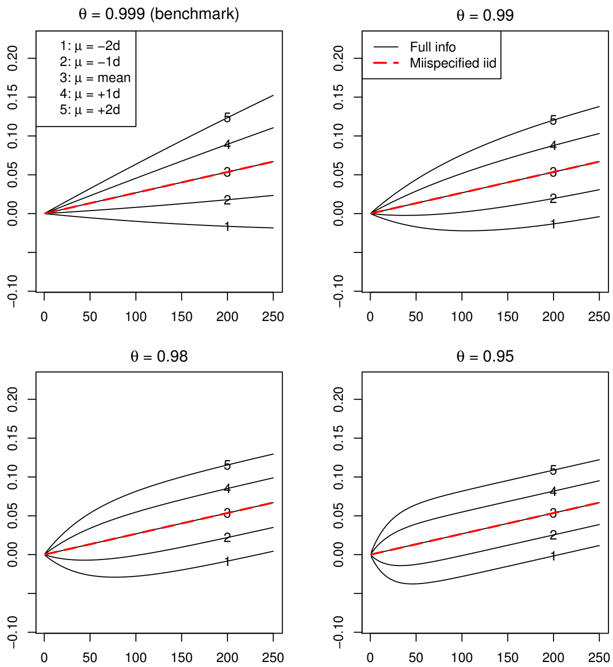


Figure 37: Per-period mean of $y_{t+1:t+\tau}$ as a function of τ: fixing $\theta = 0.999$

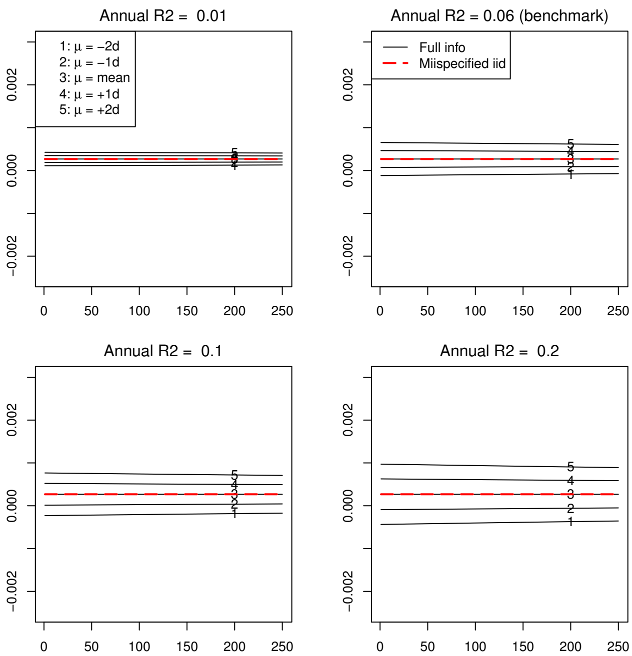


Figure 38: Per-period mean of $y_{t+1:t+\tau}$ as a function of τ: fixing annual $R^2 = 0.06$

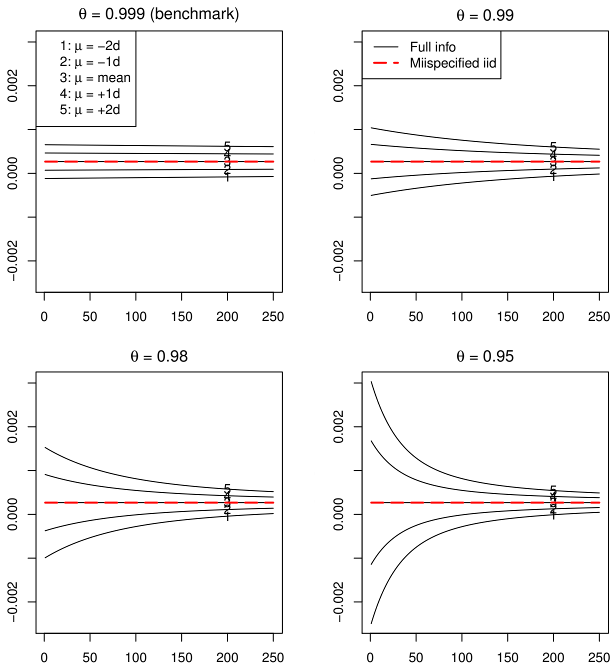
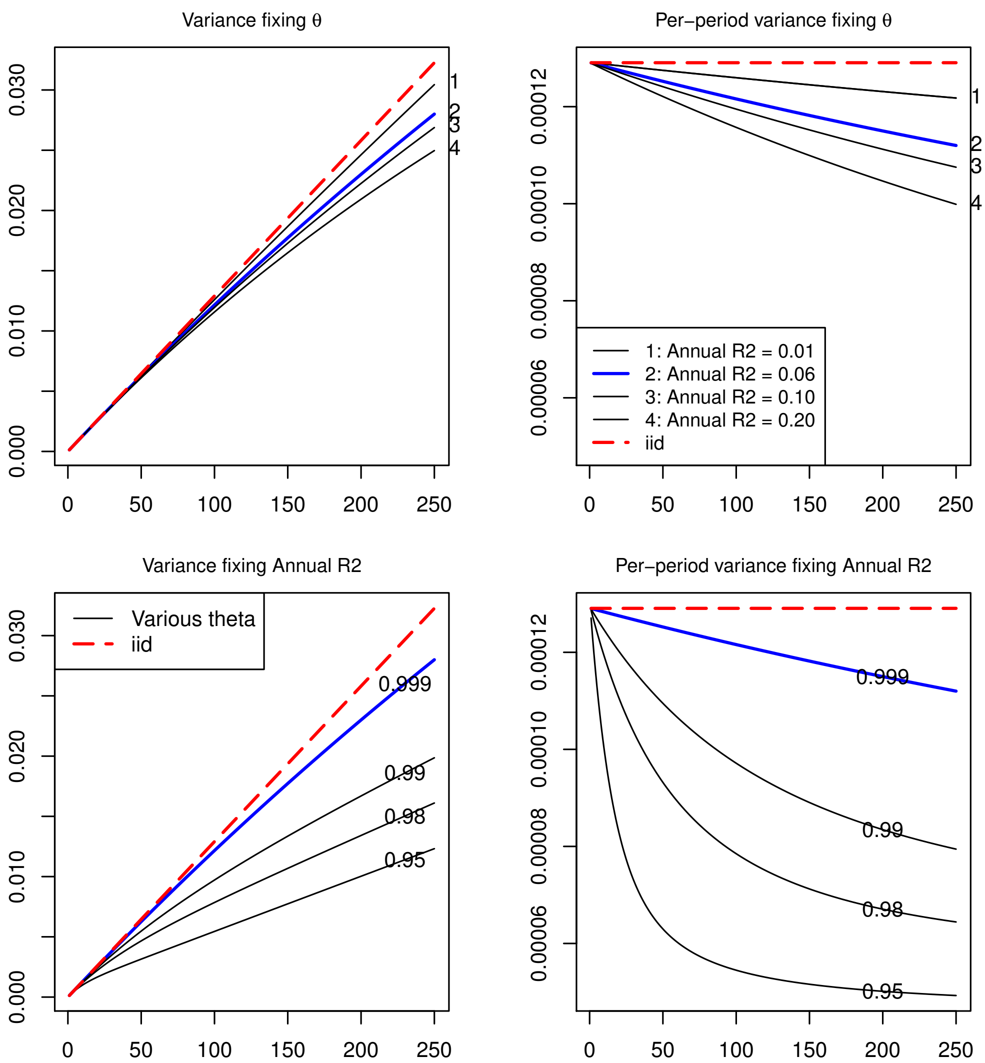


Figure 39 shows $Var(y_{t+1:t+\tau})$ in the left panels and the per-period variance ($\mathbf{v}_{\tau,iid}$ and $\mathbf{v}_{\tau,full}$) in the right panels. The per-period variance is constant for misspecified iid model, and those with full information is smaller than that. The difference is larger as the annual R^2 is larger (fixing θ), and θ is smaller (fixing annual R^2).

Figure 39: Variance and per-period variance



7.1.2 VaR

Figure 40 shows the VaR with calibrations fixing $\theta = 0.999$ while varying the annual R^2. As the annual R^2 increases, the variation of VaR based on different values of μ_t increases. However, the overall relationship between the VaR with misspecified iid model is the same, i.e., the iid-based VaR lies between the line 2 and line 3 of each panel. Table 3 shows the values of selected horizons. Table 42 summarizes it.

Figure 40: VaR: fixing $\theta = 0.999$

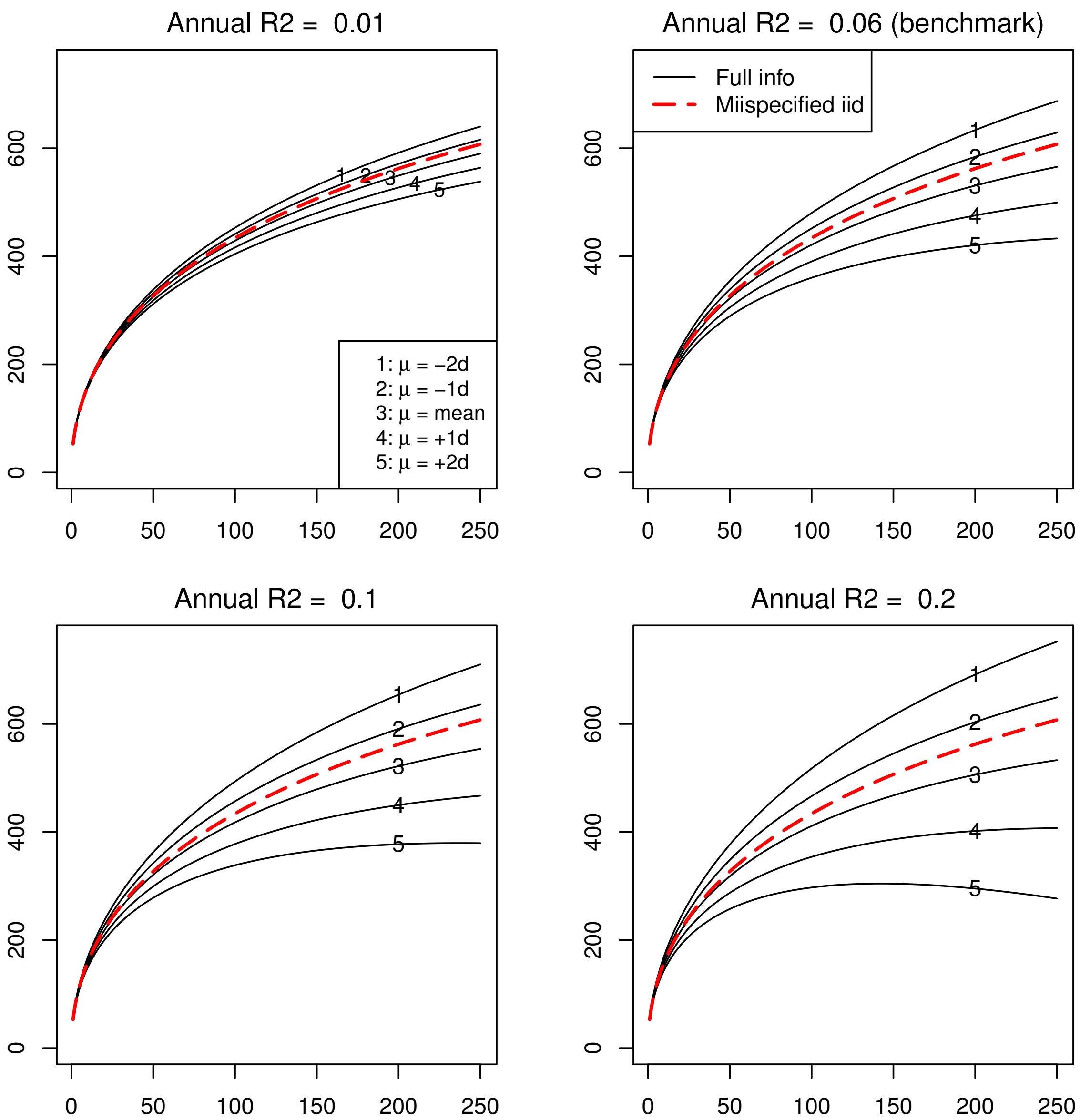


Table 42: VaR: fixing $\theta = 0.999$

Annual R^2 = 0.01

Days	oracle VaR μ_1	μ_2	μ_3	μ_4	μ_5	iid VaR
1	53.3	53.2	53.0	52.8	52.7	53.0
10	162.3	160.8	159.3	157.8	156.4	159.5
22	234.2	231.2	228.0	224.8	221.7	228.6
66	381.1	372.8	364.2	355.5	347.1	367.1
125	495.2	481.0	466.1	451.0	436.5	473.1
250	640.2	615.9	590.1	563.9	538.2	607.5

Annual R^2 = 0.06 (benchmark)

Days	oracle VaR μ_1	μ_2	μ_3	μ_4	μ_5	iid VaR
1	53.8	53.4	53.0	52.6	52.2	53.0
10	166.3	162.7	159.0	155.3	151.8	159.5
22	242.4	234.9	227.1	219.3	211.7	228.6
66	401.1	381.1	359.9	338.5	317.7	367.1
125	526.8	492.6	456.1	418.7	382.0	473.1
250	687.2	628.9	565.5	499.4	433.1	607.5

Annual R^2 = 0.10

Days	oracle VaR μ_1	μ_2	μ_3	μ_4	μ_5	iid VaR
1	54.0	53.5	53.0	52.5	52.0	53.0
10	168.3	163.7	158.9	154.1	149.5	159.5
22	246.3	236.7	226.6	216.5	206.7	228.6
66	410.9	385.2	357.9	330.1	303.1	367.1
125	542.2	498.3	451.2	402.7	354.7	473.1
250	710.2	635.7	553.8	467.2	379.2	607.5

Annual R^2 = 0.20

Days	oracle VaR μ_1	μ_2	μ_3	μ_4	μ_5	iid VaR
1	54.4	53.7	53.0	52.3	51.6	53.0
10	171.9	165.4	158.6	151.8	145.3	159.5
22	253.6	240.0	225.8	211.5	197.6	228.6
66	428.8	392.6	354.1	314.6	275.9	367.1
125	570.3	508.9	442.4	372.9	303.4	473.1
250	752.2	649.1	533.0	407.2	276.8	607.5

(million dollars)

Figure 41 shows the VaR with calibrations fixing annual $R^2 = 0.06$ while varying θ. As θ decreases, VaR based on misspecified iid model deviates more from the oracle VaR. Table 43 summarizes it.

Figure 41: VaR: fixing annual $R^2 = 0.06$

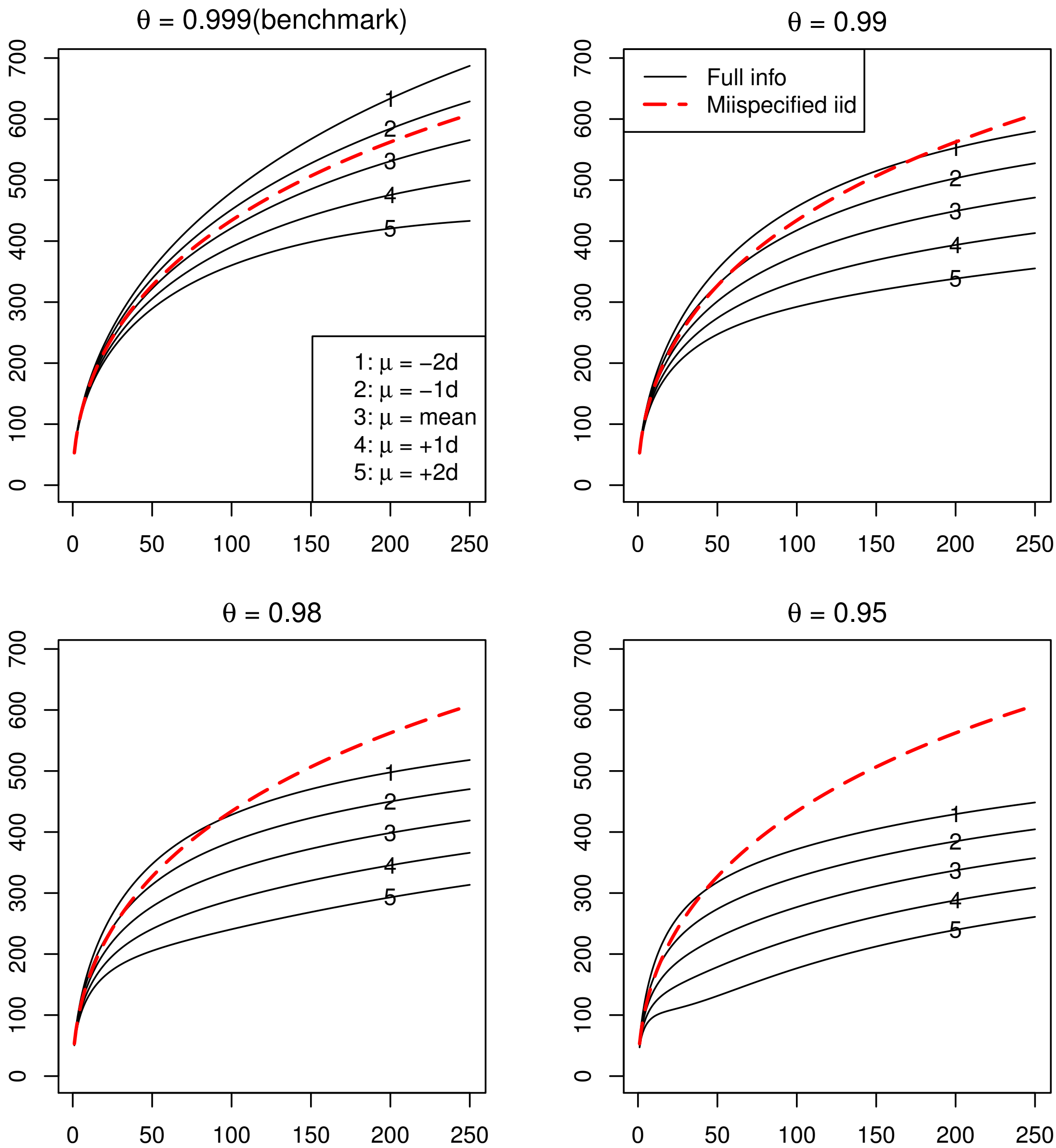


Table 43: VaR: fixing annual $R^2 = 0.06$

$\theta = 0.999$ (benchmark)

	oracle VaR					iid VaR
Days	μ_1	μ_2	μ_3	μ_4	μ_5	
1	53.8	53.4	53.0	52.6	52.2	53.0
10	166.3	162.7	159.0	155.3	151.8	159.5
22	242.4	234.9	227.1	219.3	211.7	228.6
66	401.1	381.1	359.9	338.5	317.7	367.1
125	526.8	492.6	456.1	418.7	382.0	473.1
250	687.2	628.9	565.5	499.4	433.1	607.5

$\theta = 0.99$

	oracle VaR					iid VaR
Days	μ_1	μ_2	μ_3	μ_4	μ_5	
1	54.5	53.8	53.0	52.2	51.4	53.0
10	170.7	163.8	156.7	149.6	142.7	159.5
22	247.8	234.2	219.9	205.6	191.7	228.6
66	394.3	363.6	331.1	297.9	265.5	367.1
125	489.0	446.2	400.3	353.1	306.6	473.1
250	579.5	527.5	471.3	413.1	355.2	607.5

$\theta = 0.98$

	oracle VaR					iid VaR
Days	μ_1	μ_2	μ_3	μ_4	μ_5	
1	55.4	54.2	52.9	51.6	50.4	53.0
10	175.1	164.4	153.3	142.1	131.3	159.5
22	251.4	231.4	210.3	188.9	168.1	228.6
66	380.7	342.3	301.3	259.4	218.2	367.1
125	451.9	405.8	356.4	305.5	255.2	473.1
250	518.0	470.3	419.0	366.0	313.5	607.5

$\theta = 0.95$

	oracle VaR					iid VaR
Days	μ_1	μ_2	μ_3	μ_4	μ_5	
1	58.1	55.4	52.6	49.8	47.1	53.0
10	182.3	161.8	140.3	118.6	97.5	159.5
22	249.1	215.8	180.4	144.3	109.1	228.6
66	339.5	294.1	245.5	195.6	146.4	367.1
125	389.8	344.2	295.5	245.3	195.8	473.1
250	448.4	404.4	357.3	308.8	261.0	607.5

(million dollars)

7.2 ARMA(1,1) representation

We have M_0. Letting L be a back-shift operator, and by $y_t = \mu_{t-1} + u_t$,

$$\begin{aligned}(1-\theta L)y_t &= (1-\theta L)(\mu_{t-1}+u_t) = (1-\theta L)\mu_{t-1} + (1-\theta L)u_t. \\ &= c + w_{t-1} + u_t - \theta u_{t-1}\end{aligned}$$

where the last equality comes from the fact that $\mu_t = c + \theta\mu_{t-1} + w_t$. Let us define z_t as follows:

$$z_t \equiv (1-\theta L)y_t - c.$$

Since $z_t = w_{t-1} + u_t - \theta u_{t-1}$, $Cov(z_t, z_{t-1}) = \sigma_{uw} - \theta\sigma_u^2$ and $Cov(z_t, z_{t-h}) = 0$ for all h such that $|h| > 1$. Therefore, z_t has an MA(1) representation with $E(z_t) = 0$. Thus, letting $\{\eta_t\}$ be a second-order white noise, z_t can be expressed as the following:

$$z_t = \eta_t - \gamma\eta_{t-1}$$

Let ACF(1) of z_t be λ which is $\frac{\sigma_{uw}-\theta\sigma_u^2}{\sigma_w^2+(1+\theta^2)\sigma_u^2-2\theta\sigma_{uw}}$. Then, $\lambda = \frac{Cov(z_t,z_{t-1})}{Var(z_t)} = \frac{-\gamma}{1+\gamma^2}$ and thus $\lambda\gamma^2 + \gamma + \lambda = 0$. First, since the discriminant of this equation is positive,[14] the two solutions are real. Second, since the product of the solutions are one and we are interested in the case where $|\gamma| < 1$, we select the smaller solution, i.e., $\gamma = \frac{-(1-\sqrt{1-4\lambda^2})}{2\lambda}$. With the parameter values above, y_t has the following ARMA(1,1) representation:

$$y_t = c + \theta y_{t-1} + \eta_t - \gamma\ \eta_{t-1}$$

Since $(u_t, w_t)'$ is iid and normally distributed, η_t is a m.d.s. $Var(\eta_t) \equiv \sigma_\eta^2 = \frac{\theta\sigma_u^2-\sigma_{uw}}{\gamma} = \frac{\sigma_w^2+(1+\theta^2)\sigma_u^2-2\theta\sigma_{uw}}{1+\gamma^2}$.

7.3 Characteristic function of Heston Nandi Model

The characteristic function of $y_{t+1:t+\tau}|\underline{y_t}, S_t$ is

$$\begin{aligned}C(\tau,u) &= \exp\left(a(\tau,u) + b(\tau,u)h_t\right) \\ a(\tau,u) &= a(\tau-1,u) + iur + b(\tau-1,u)\omega - \frac{1}{2}\log(1-2\alpha b(\tau-1,u)) \\ b(\tau,u) &= iu(\lambda+\gamma) - \frac{1}{2}\gamma^2 + \beta b(\tau-1,u) + \frac{1/2(iu-\gamma)^2}{1-2\alpha b(\tau-1,u)} \\ a(0,u) &= 0,\ \ b(0,u) = 0\end{aligned}$$

For example, when $\tau = 1$,

$$a(1,u) = iur, \quad b(1,u) = -\frac{1}{2}u^2 + iu\lambda, \quad C_1(u) = -\frac{1}{2}u^2h_t + iu(r+\lambda h_t)$$

which is the characteristic function of $N(r+\lambda h_t,\ h_t)$.

[14] Since $\lambda = \frac{-\gamma}{1+\gamma^2}$, $|\lambda| \leq \frac{1}{2}$. Thus, $\Delta = 1 - 4\lambda^2 \geq 0$.

7.4 Inverting characteristic function and obtaining CDF

we invert the characteristic function to a CDF of $y_{t+1:t+\tau}|\underline{y_t}$ by Gil-Palaez formula:

$$Pr(y_{t+1:t+\tau} \leq y|\underline{y_t}) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{Im(e^{-iuy}C_h(\tau,u))}{u}du$$

where $Im(z)$ denotes the imaginary part of a complex number z. We evaluate the integral by numerical integration. Let us denote by $I(u) = Im(e^{-iuy}C_h(\tau,u))$,[15]

$$Pr(y_{t+1:t+\tau} \leq y|\underline{y_t}) \approx \frac{1}{2} - \frac{\Delta}{\pi}\left(\sum_{k=0}^{n-1}\frac{I((k+1/2)\Delta)}{(k+1/2)\Delta}\right)$$

I pick $\Delta = 0.01$, $n = 50,00$ so that I evaluate the function from 0 to 500.[16]

7.5 Various values of h_t: analytical approach

The characteristic function of $h_{t+\tau}$ denoted by $C_h(\tau,u)$is

$$\begin{aligned}
C_h(\tau,u) &= E_t(\exp(iuh_{t+\tau})) = \exp(\widetilde{a}(\tau,u) + \widetilde{b}(\tau,u)h_t) \\
\widetilde{a}(\tau,u) &= \widetilde{a}(\tau-1,u) + \widetilde{b}(\tau-1,u)\omega - \frac{1}{2}\log(1-2\alpha\widetilde{b}(\tau-1,u)) \\
\widetilde{b}(\tau,u) &= \widetilde{b}(\tau-1,u)\beta + \frac{\alpha\widetilde{b}(\tau-1,u)\gamma^2}{1-2\alpha\widetilde{b}(\tau-1,u)}
\end{aligned}$$

with $\widetilde{a}(0,u) = 0$, $\widetilde{b}(0,u) = iu$.

Proof. Let us guess that $C_h(\tau,u) = \exp(\widetilde{a}(\tau,u)+\widetilde{b}(\tau,u)h_t)$. For simplicity, let us define $\widetilde{a}(\tau,u) \equiv \widetilde{a}_\tau$, $\widetilde{b}(\tau,u) \equiv \widetilde{b}_\tau$. Then,

$$\begin{aligned}
C_h(\tau,u) &= \exp(\widetilde{a}_\tau + \widetilde{b}_\tau h_t) \\
&= E_t[\exp(\widetilde{a}_{\tau-1} + \widetilde{b}_{\tau-1}h_{t+1})] \\
&= E_t[\exp(\widetilde{a}_{\tau-1} + \widetilde{b}_{\tau-1}(\omega + \beta h_t + \alpha(z_{t+1} - \gamma\sqrt{h_t})^2))] \\
&= E_t[\exp(\widetilde{a}_{\tau-1} + \widetilde{b}_{\tau-1}\omega + \widetilde{b}_{\tau-1}\beta h_t + \alpha\widetilde{b}_{\tau-1}(z_{t+1} - \gamma\sqrt{h_t})^2))]
\end{aligned}$$

Using the fact that, for standard normal variable z,

$$E[\exp(a(z+b)^2] = \exp\left(-\frac{1}{2}\log(1-2a) + \frac{ab^2}{1-2a}\right)$$

[15]Composite midpoint rule states that $\int_a^b f(x)dx = h(f(x_{1/2}) + f(x_{3/2}) + \cdots + f(x_{(2n-1)/2})) + O(n^{-2})$.

[16]With $\Delta = 0.01$ and $n = 10,000$ so that evaluating the function up to 1000 makes difference of order $e-15$ at $h = 2$ and less than this for $h \geq 3$. With $\Delta = 0.005$, $n = 100,000$ have even smaller differences.

$$
\begin{aligned}
C_h(\tau,u) &= \exp\left(\widetilde{a}_{\tau-1}+\widetilde{b}_{\tau-1}\omega+\widetilde{b}_{\tau-1}\beta h_t-\frac{1}{2}\log(1-2\alpha\widetilde{b}_{\tau-1})+\frac{\alpha\widetilde{b}_{\tau-1}\gamma^2 h_t}{1-2\alpha\widetilde{b}_{\tau-1}}\right)\\
&= \exp\left(\widetilde{a}_{\tau-1}+\widetilde{b}_{\tau-1}\omega-\frac{1}{2}\log(1-2\alpha\widetilde{b}_{\tau-1})+\left(\widetilde{b}_{\tau-1}\beta+\frac{\alpha\widetilde{b}_{\tau-1}\gamma^2}{1-2\alpha\widetilde{b}_{\tau-1}}\right)h_t\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\widetilde{a}(\tau,u) &= \widetilde{a}_{\tau-1}+\widetilde{b}_{\tau-1}\omega-\frac{1}{2}\log(1-2\alpha\widetilde{b}_{\tau-1})\\
\widetilde{b}(\tau,u) &= \widetilde{b}_{\tau-1}\beta+\frac{\alpha\widetilde{b}_{\tau-1}\gamma^2}{1-2\alpha\widetilde{b}_{\tau-1}}
\end{aligned}
$$

□

The unconditional characteristic function is computed as follows:

$$
\begin{aligned}
\exp(C_h(u)) &\equiv E(\exp(iuh_{t+1}))=E[E(\exp(iuh_{t+1})|h_t)]=E[\exp(\widetilde{a}(1,u)+\widetilde{b}(1,u)h_t)]\\
&= \exp(\widetilde{a}(1,u))E[\exp(\widetilde{b}(1,u)h_t)]\\
&= \exp(\widetilde{a}(1,u))\exp(C_h(\widetilde{b}(1,u))\\
&= \exp(\widetilde{a}(1,u)+C_h(\widetilde{b}(1,u))
\end{aligned}
$$

Plugging in $\widetilde{a}(1,u)$ and $\widetilde{b}(1,u)$,

$$
C_h(u) = iu\omega-\frac{1}{2}\log(1-2iu\alpha)+C_h\left(iu\beta+\frac{i\alpha\gamma^2 u}{1-2iu\alpha}\right)
$$

References

Adrian, T., and M. K. Brunnermeier (2011): "CoVaR," Working paper.

Barberis, N. (2000): "Investing for the Long Run when Returns Are Predictable," *Journal of Finance*, 55(1), 225–264.

Brownlees, C., and R. F. Engle (2017): "SRISK: A Conditional Capital Shortfall Measure of Systemic Risk," *Review of Financial Studies*, 30(1), 48–79.

Campbell, J. Y. (2001): "Why long horizons? A study of power against persistent alternatives," *Journal of Empirical Finance*, 8(5), 459–491.

Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi (2014): "The Economic Value of Realized Volatility: Using High-Frequency Returns for Option Valuation," *Journal of Financial and Quantitative Analysis*, 49(03), 663–697.

Christoffersen, P., K. Jacobs, and C. ornthanalai (2009): "Exploring Time-Varying Jump Intensities: Evidence from S&P500 Returns and Options," Working paper.

——— (2011): "Dynamic Jump Intensities and Risk Premia: Evidence from S&P500 Returns and Options," Working paper.

Christoffersen, P. F. (2012): *Elements of Financial Risk Management*. Academic Press, second edn.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–263.

Duffie, D., and J. Pan (2001): "Analytical value-at-risk with jumps and credit risk," *Finance and Stochastics*, 5(2), 155–180.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68(6), 1343–1376.

Engle, R. F. (2011): "Long-Term Skewness and Systemic Risk," *Journal of Financial Econometrics*, 9(3), 437–468.

Fama, E. F., and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–27.

Giacomini, R., and I. Komunjer (2005): "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business & Economic Statistics*, 23, 416–431.

Giacomini, R., and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

Heston, S. L., and S. Nandi (2000): "A closed-form GARCH option valuation model," *Review of Financial Studies*, 13(3), 585–625.

Ludvigson, S. C., and S. Ng (2007): "The empirical risk-return relation: A factor analysis approach," *Journal of Financial Economics*, 83(1), 171–222.

Meddahi, N. (2002): "ARMA Representation of Two-Factor Models," CIRANO Working Papers 2002s-92, CIRANO.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Pástor, L., and R. F. Stambaugh (2009): "Predictive Systems: Living with Imperfect Predictors," *Journal of Finance*, 64(4), 1583–1628.

——— (2012): "Are Stocks Really Less Volatile in the Long Run?," *Journal of Finance*, 67(2), 431–478.

Patton, A. J. (2016): "Comparing Possibly Misspecified Forecasts)," Working paper.

Sargent, T. J., and L. Hansen (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91(2), 60–66.

Stock, J., and M. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.