FORECASTING IN FINANCIAL MARKET USING MARKOV REGIME SWITCHING AND PRINCIPAL COMPONENT ANALYSIS.

NOP SOPIPAN

September 13, 2012
Motivation

Research Objectives

Methodology

Scope and limitations

Research Methodology

- Heteroskedasticity Model
- Spurious high persistence problem
- Multicollinearity

Descriptive of model
การพยากรณ์ในตลาดทางการเงิน ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก และวิธีสับเปลี่ยนสถานะมาร์คอฟ
The Beginning

Prof. Dr. Pairote Sattayatham

Prof. Dr. Bhusana Premanode
Characteristics of returns

![Graphs showing financial price, log return, and volatility of log return over time.]

**Outline**
- Motivation
- Research Objectives
- Methodology
- Scope and limitations
- Research Methodology
- Descriptive of model

**NOP SOPIPAN**

**FORECASTING IN FINANCIAL MARKET USING MARKOV**
Property in Financial returns

- (Weakly) Stationary process.
- Look liked White noise process
- The model assume random walk model.
Forecast Financial Return

Let \( \{P_t\} \) denote the series of the financial price at time \( t \) and \( \{r_t\}_{t>0} \) be the logarithmic return (in percent) on the financial price at time \( t \) be a sequence of random variables on a probability space \((\Omega, \mathcal{F}, \mathcal{P})\), i.e.

\[
r_t = 100 \cdot \ln\left(\frac{P_t}{P_{t-1}}\right)
\]  

Let \( r_t \) be explain in mean equation, i.e.

\[
r_t = \mu_t + \varepsilon_t
\]  

with mean and variance:

\[
\mu_t = E(r_t | F_{t-1}), 
\]

\[
h_t = \text{Var}(r_t | F_{t-1}) = E[(r_t - \mu_t)^2 | F_{t-1}]
\]
Mehmet A. (2008), Marcucci J. (2005) assume mean equation in returns is constant.

Easton and Faff (1994), and Kyimaz and Berument (2001) assume mean equation in returns with a one week delay into the regression model.

Supot C. (2003) assume mean equation in returns with Autoregressive process.
Construct Model

- The constant mean equation cannot be forecast accurately due to the inaccuracy of the financial data, since the financial returns depend concurrently and dynamically on many economic and financial variables.

- The fact that the return has a statistically significant autocorrelation indicates that lagged returns might be useful in predicting future returns.

- We consider add some explanatory variables and stationary Autoregressive Moving-average order p and q (ARMA (p,q)). Then we add them into a mean equation in order to increase accuracy.
The objectives of the research are the following:

1. Construct model of equation (3) for forecasting of financial return.
The objectives of the research are the following:

1. Construct model of equation (3) for forecasting of financial return.
2. Solve multicollinearity problem in explanatory variables of equation (3) by using Principal Component Analysis (PCA).
The objectives of the research are the following:

1. Construct model of equation (3) for forecasting of financial return.

2. Solve multicollinearity problem in explanatory variables of equation (3) by using Principal Component Analysis (PCA).

3. Construct heteroskedasticity model for forecasting of volatility with GARCH-type model and Markov Regime Switching GARCH model.
The objectives of the research are the following:

1. Construct model of equation (3) for forecasting of financial return.

2. Solve multicollinearity problem in explanatory variables of equation (3) by using Principal Component Analysis (PCA).

3. Construct heteroskedasticity model for forecasting of volatility with GARCH-type model and Markov Regime Switching GARCH model.

4. Compare and evaluate the performance of models between constant mean equation and model of equation (3) with loss functions.
The process of research had been following:

The process of research had been following:


- **2\textsuperscript{nd} Paper.** P. Sattayatham, N. Sopipan and B. Premanode (2012), "Forecasting the Stock Exchange of Thailand using Day of the Week Effect and Markov Regime Switching GARCH.", Journal of Forecasting. (in progress)
The process of research had been following:


Next Step Use PCA to solve multicollinearity problem in explanatory variables in mean equation forecast return.
The forecasting in Markov Regime Switching for simplicity, this thesis assume presence of two regimes and order of the GARCH-type is (1,1).
The forecasting in Markov Regime Switching for simplicity, this thesis assume presence of two regimes and order of the GARCH-type is (1,1).

The standardized errors follow student-t, generalized error distributions (GED) and normal distribution.
Heteroskedasticity Model
Spurious high persistence problem
Multicollinearity
Heteroskedasticity

The generalized autoregressive conditional heteroskedasticity (GARCH)-type models mainly capture for problem of heteroskedasticity.

- GARCH model
- The Exponential GARCH (EGARCH) model
- GJR-GARCH model
Heteroskedasticity model

- **GARCH(1,1):** [Bollerslev (1986)]

\[
\varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}
\]  

(3)
Heteroskedasticity Model

Spurious high persistence problem
Multicollinearily

Heteroskedasticity model

- **GARCH(1,1):** [Bollerslev (1986)]

  \[ \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \]  

- **EGARCH(1,1):** [Nelson (1991)]

  \[ \ln(h_t) = \alpha_0 + \alpha_1 \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta_1 \ln(h_{t-1}) + \xi \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \]
Heteroskedasticity Model

- **GARCH(1,1):** [Bollerslev (1986)]
  \[ \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3} \]

- **EGARCH(1,1):** [Nelson (1991)]
  \[ \ln(h_t) = \alpha_0 + \alpha_1 \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta_1 \ln(h_{t-1}) + \xi \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \tag{4} \]

- **GJR-GARCH(1,1):** [Glosten, Jagannathan and Runkle (1993)]
  \[ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 (1 - I_{\{\varepsilon_{t-1} > 0\}}) + \beta_1 h_{t-1} + \xi \varepsilon_{t-1}^2 (I_{\{\varepsilon_{t-1} > 0\}}) \tag{5} \]
Spurious high persistence problem
Review of Markov Regime Switching

- Hamilton and Susmel (1994) and Cai (1994) proposed MRS ARCH.
- Gray (1996) proposed MRS GARCH.
- Klassen (2002) modification of Gray’s MRS GARCH.
Concept of Markov Regime Switching

- Generalization of the simple dummy variables approach.
- Allow regimes (called states) to occur several periods over time.
- In each period $t$ the state is denoted by $S_t$.
- There can be $N$ possible states: $S_t = 1, \ldots, N$. 
Markov Regime Switching

- Probability $S_t$ equals some particular value $j$ depends on the past only through the most recent value $S_{t-1}$:
  
  $$P\{S_t = j|S_{t-1} = i, S_{t-2} = k, \ldots\} = P\{S_t = j|S_{t-1} = i\} = p_{ij}.$$
Markov Regime Switching

- Probability $S_t$ equals some particular value $j$ depends on the past only through the most recent value $S_{t-1}$:
  \[ P\{S_t = j | S_{t-1} = i, S_{t-2} = k, \ldots\} = P\{S_t = j | S_{t-1} = i\} = p_{ij}. \]
- Process is described as an $N$-state Markov Chain with transition probabilities
  \[ \{p_{ij}\}_{i,j=1,2,\ldots,N} \text{ and } p_{i1} + p_{i2} + \ldots + p_{iN} = 1. \]
Markov Regime Switching

The conditional probability density function for the observations $r_t$ given the state variables $S_t, S_{t-1}$ and $F_{t-1}$ is $f(r_t | S_t, S_{t-1}, F_{t-1})$
The conditional probability density function for the observations $r_t$ given the state variables $S_t, S_{t-1}$ and $F_{t-1}$ is $f(r_t|S_t, S_{t-1}, F_{t-1})$.

The chain rule for conditional probabilities yields then for the joint probability density function for the variables $r_t, S_t, S_{t-1}$ given $F_{t-1}$ is $f(r_t, S_t, S_{t-1}|F_{t-1}) = f(r_t|S_t, S_{t-1}, F_{t-1})Pr(S_t, S_{t-1}|F_{t-1})$. 
Markov Regime Switching

- The conditional probability density function for the observations $r_t$ given the state variables $S_t, S_{t-1}$ and $F_{t-1}$ is
  $f(r_t|S_t, S_{t-1}, F_{t-1})$

- The chain rule for conditional probabilities yields then for the joint probability density function for the variables $r_t, S_t, S_{t-1}$ given $F_{t-1}$ is
  $f(r_t, S_t, S_{t-1}|F_{t-1}) = f(r_t|S_t, S_{t-1}, F_{t-1}) \Pr(S_t, S_{t-1}|F_{t-1})$

- The log-likelihood function to be maximized with respect to the unknown parameters becomes
  \[ l(\theta) = \sum_{t=1}^{T} \left[ \log \left( \sum_{S_t=1}^{N} \sum_{S_{t-1}=1}^{N} \right) f(r_t, S_t, S_{t-1}|F_{t-1}) \right] \]
The MRS-GARCH model with only two regimes can be represented as follows.

\[ r_t = \mu_{t,S_t} + \varepsilon_t = \mu_{t,S_t} + \eta_t \sqrt{h_{t,S_t}} \]  
\[ h_{t,S_t} = \alpha_{0,S_t} + \alpha_{1,S_t} \varepsilon_{t-1}^2 + \beta_{1,S_t} h_{t-1} \]  

where \( S_t = 1, 2 \), \( \mu_{t,S_t} \) is the mean and \( h_{t,S_t} \) is the volatility under regime \( S_t \) on \( F_{t-1} \).
Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated.
For the mean equation in (3) i.e.

\[ r_t = \mu_t + \varepsilon_t, \]

\[ \mu_t = \phi_0 + \sum_{i=1}^{k} \beta_i X_{it} + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \]

**Multicollinearity**, or high correlation between the explanatory variables in a regression equation. One method for removing such multicollinearity and redundant independent variables is **principal component analysis (PCA)**.
The idea of PCA is to find linear combinations $\alpha_i$ such that $Z_i$ and $Z_j$ are uncorrelated for $i \neq j$ and the variances of $Z_i$ are as large as possible. More specifically:

- The first principal component of $X$ is the linear combination $Z_1 = \alpha'_1 X$ that maximizes $\text{Var} (Z_1)$ subject to the constraint $\alpha'_1 \alpha_1 = 1$.

- The second principal component of $X$ is the linear combination $Z_2 = \alpha'_2 X$ that maximizes $\text{Var} (Z_2)$ subject to the constraint $\alpha'_2 \alpha_2 = 1$ and $\text{Cov}(Z_2, Z_1) = 0$. 
The $k$th principal component of $X$ is the linear combination $Z_k = \alpha_k'X$ that maximizes $\text{Var}(Z_k)$ subject to the constraint $\alpha_k'\alpha_k = 1$ and $\text{Cov}(Z_k, Z_{k'}) = 0$ for $k \neq k'$. 
Principal Component Analysis (Cont.)

The new variables from the PCA become ideal to use as predictors in a regression equation since they optimize spatial patterns and remove possible complications caused by multicollinearity.

\[ \hat{Y} = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i X_i \quad \text{PCA} \quad \hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{m} \hat{\beta}_i Z_i \]
The model of returns (2) as follow:

\[ r_t = \mu_t + \varepsilon_t \]

This thesis use three mean equations of return as follow:

(1) **Mean equation are constants.**

\[ \mu_t = \delta. \]

\[ \mu_t = \delta S(t). \]
(2) Mean equation as stationary ARMA\((p,q)\) includes day of the week effect.

\[
\mu_t = \sum_{i=1}^{5} \beta_i D_{it} + \sum_{j=1}^{p} \phi_j r_{t-j} - \sum_{k=1}^{q} \theta_k \epsilon_{t-k},
\]

where

- \(D_{it}, i = 1, \ldots, 5\) are dummy variables which take on the value of 1 if the corresponding return of the five days respectively and 0 otherwise,
- \(\beta_i, i = 1, \ldots, 5\) are coefficients which represent the average return for each day of the week,
- \(\phi_j, \theta_k, i = 1, \ldots, p, k = 1, \ldots, q\), are coefficients which represent the ARMA \((p, q)\).
THANK YOU FOR ATTENTION