Time Series Analysis on Stock Returns

Yiqiao Yin

Columbia University

March 27, 2017

Abstract

This papers applies ARMA model to daily, weekly, and monthly price returns of S&P 500 Index Fund. We present model selection, forecasts, and residual tests. The results for weekly and monthly residuals are consistent with short- and long-run reversal strategy.

Contents

1 Introduction

Sharpe (1964) [3] had long influenced academics and practitioners to think about risk and return. The contribution brought us the idea a portfolio invested with mean-variance efficient upon the foundation of Markowitz (1959) [2]. However, model using a market factor does a poor job at forecasting future returns. One needs to forecast market returns to identify returns of particular stocks. Market returns act like a martingale. Therefore, this phenomenal deserves attention to conduct time series analysis. This paper conducts a series of time series analysis to find the best fitted model of market returns. We use S&P 500 Index Fund as original data set.

Section Two introduces mathematical models. We start from stationary process and take readers through definition of covariance function. We define ARMA(p, q) model and stationarity. Section Three takes real dataset for tests and forecasts. We follow the same order in previous section and present a series of analysis from the results of data.

2 Mathematical Model

2.1 Stationary Process

Definition 2.1. A **time series model** for the observed data $\{x_t\}$ is specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

Remark 2.2. We will frequently use the term *time series* to mean both the data and the process of which it is a realization.

□

Remark 2.3. A complete probabilistic time series model for the sequence of random variables $\{X_1, X_2, ...\}$ would specify all of the **joint distributions** of the random vectors $(X_1, ..., X_n)'$, for $n = 1, 2, ...$, or equivalently all of the probabilities

$$P[X_1 \leq x_1, ..., X_n \leq x_n],\ -\infty < x_1, ..., x_n < \infty,\ n = 1, 2, ...$$

□

The random walk $\{S_t, t = 0, 1, 2, ...\}$ (starting at zero) is obtained by cumulatively summing (or "integrating") iid random variables. Thus a random walk with zero mean is obtained by defining $S_0 = 0$ and

$$S_t = X_1 + X_2 + \cdots + X_t, \ for \ t = 1, 2, ...,$$

where $\{X_t\}$ is iid noise.

In the least squares procedure we may attempt to fit a parametric family of functions, e.g.,

$$m_t = a_0 + a_1 t + a_2 t^2,$$

to the data $\{x_1, ..., x_n\}$ by choosing the parameters, in this illustration a_0, a_1, and a_2, to minimize $\sum_{t=1}^{n}(x_t - m_t)^2$. This method of curve fitting is called **least squares regression**.

Definition 2.4. Let $\{X_t\}$ be a time series with $\mathbb{E}(X_t^2) < \infty$. The **mean function** of $\{X_t\}$ is

$$\mu_X(t) = \mathbb{E}(X_t).$$

The **covariance function** of $\{X_t\}$ is

$$\gamma_X(r, s) = Cov(X_r, X_s) = \mathbb{E}[(X_r - \mu_X(r))(X_s - \mu_s(s))]$$

for $r \in \mathbb{Z}$ and $s \in \mathbb{Z}$.

Proposition 2.5. *The MA(q) Process: $\{X_t\}$ is a **moving-average process of order** q if*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\theta_1, ..., \theta_q$ are constants.

Definition 2.6. The time series $\{X_t\}$ is an **ARMA(1,1) process** if it is stationary and satisfies (for every t)

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1},$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\phi + \theta \neq 0$.

Definition 2.7. $\{X_t\}$ is **(weakly) stationary** if

(i) $\mu_X(t)$ is independent of t, and

(ii) $\gamma_X(t+h,t)$ is independent of t for each h.

Definition 2.8. Let $\{X_t\}$ be a stationary time series. The **autocovariance function** (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = Cov(X_{t+h}, X_t).$$

The **autocorrelation function** (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t).$$

Definition 2.9. Let $x_1, ..., x_n$ be observations of a time series. The **sample mean** of $x_1, ..., x_n$ is

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

$$\hat{\gamma}(h) := n^{-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}),\ -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},\ -n < h < n.$$

Consider a stationary time series defined as

$$X_t = \phi X_{t-1} + Z_t,\ t = 0, \pm 1, ...,$$

where $|\phi| < 1$ and $\{Z_t\} \sim WN(0, \sigma^2)$. The best linear predictor of X_{n+1} in terms of $\{1, X_n, ..., X_1\}$ is, for $n \geq 1$,

$$P_n X_{n+1} = \mathbf{a}_n' \mathbf{X}_n,$$

where $\mathbf{X}_n = (X_n, ..., X_1)'$ and

$$\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}$$

A solution would be

$$\mathbf{a}_n(\phi, 0, ..., 0)',$$

and hence the best linear predictor of X_{n+1} in terms of $\{X_1, ..., X_n\}$ is

$$P_n X_{n+1} = \mathbf{a}_n' \mathbf{X}_n = \phi X_n,$$

with mean squared error

$$\mathbb{E}(X_{n+1} - P_n X_{n+1})^2 = \gamma(0) - \mathbf{a}_n \gamma_n(1) = \frac{\sigma^2}{1 - \phi^2} - \phi\gamma(1) = \sigma^2.$$

2.2 Durbin-Levinson

The coefficients $\phi_{n1} < ..., \phi_{nn}$ can be computed recursively from the equations

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1},$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1.n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

and

$$v_n = v_{n-1}[1 - \phi_{nn}^2],$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $v_0 = \gamma(0)$. See Appendix for Proof.

2.3 Innovations Algorithm

The coefficients $\theta_{n1}, ..., \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1}\big(\kappa(n=1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\big),\ 0 \leq k < n,$$

and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

3 Analysis

3.1 Dataset

This paper uses S&P 500 Index Fund, Gold Futures (GLD), and 20-Year Treasury Bill Futures (TLT) with daily, weekly, and monthly closing price as original data set. We look at two groups of data set: price minus mean, and price returns. For each group, we compute that of daily, weekly, and monthly values for all three indices (SPY, GLD, and TLT). Data set for all three indices (closing prices, adjusted) range from Jan. 1st 1980 to Mar. 27th 2017. We used the first 60% of data set as training set to compute all statistics required. We plot the rest 40% along with forecasting to see how well the forecasts predict the actual values.

For all three indices (SPY, GLD, and TLT), there were 6083, 3108, and 3690 values during the time range selected above. Their correlations among adjusted prices and price returns were presented in Table 1. Observe that SPY adjusted closing price and TLT adjusted closing price have the highest correlation among them all, 0.81. Their returns, although highly correlated, turned out to be negatively correlated at -0.42.

3.2 Analysis

Starting with daily, weekly, monthly adjusted closing prices and price returns of all three indices, we compute all possible ARMA(p,q) models and we select the best fitted models by sigma square, log likelihood, AIC, AICC, and BIC values. Table 2 presents all best fitted results for all three indices

(SPY, GLD, TLT) with all three different time units. In Panel (A), the table presents three indices with corresponding ARMA models including coefficients, standard errors, and t-stats. Panel (B) shows testing parameters, sigma square, log likelihood, AIC, AICC, and BIC for all ARMA models for all three indices with all three time units. Table 2 allowed ARMA models to have drift terms.

Observe that results show significant t-stat for almost all coefficients except for TLT, which failed to reject null hypothesis.

3.3 Forecasts

After the selection of the best fitted model, we forecast in daily, weekly, monthly time frame and compare the forecasts with real data. Figure 1 Panel (B), Figure 2 Panel (B), and Figure 3 Panel (B) present the forecasting results of all three indices among all three time units. In Panel (B) of the first three figures, we plot nine graphs. The first row of graphs present daily, weekly, and monthly closing prices. The second row of graphs present daily, weekly, and monthly adjusted closing prices minus mean price of that data set.

Panel (A) of Figure 1, 2, and 3 plot ACF and PACF of adjusted closing prices and price returns of all three time frames. Observe that Figure 2 Panel (A) present ACF for Weekly Returns that have spikes (higher than $1/\sqrt{n}$ occurring at lag-4 (this is a month since there are 4 weeks in a month). In the same panel, PACF shows a relative high spike at lag-12 for monthly returns (this is a year since there are 12 months in a year). We can find pattern in roughly same lags for GLD (spike at lag-6 in weekly ACF/PACF, spike at lag-13 in monthly ACF/PACF). Such patterns can be observed in for SPY as well (lag-32 for daily returns). Such results are consistent with short- and long-run reversals from De Bondt and Thaler (1985) [1].

Panel (B) of Figure 1, 2, and 3 plot forecasts based on the best fitted ARMA model of each of the three indices for all three time units. The first row of Panel (B) of each figure plots all of the closing prices (for all time units). The second row of Panel (B) of each figure plots 60% of the data from the first row as if there were used for training data. ARMA models were tested and the best fitted ARMA models were selected out of the training data. Based on best fitted model, we forecast the rest 40% of the original data (which are the third row). Hence, the total length of each graph in the third row is the same as the length of corresponding plot in the first row.

4 Conclusion

This paper takes readers through a series of time series analysis. We present model selection, forecasts, and residual tests. We find similar patterns as short- and long-run reversals described in De Bondt and Thaler (1985) [1], which can be a consistent trading strategy.

5 Appendix

5.1 Proof of Durbin-Levinson

Proof: The definition of ϕ_{11} ensures that the equation

$$R_n \phi_n = \rho_n$$

(where $\rho_n = (\rho(1), ..., \rho(n))'$) is satisfied for $n = 1$. The first step in the proof is to show that ϕ_n, defined recursively by ϕ_{nn} and $\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix}$, satisfies $R_n \phi_n = \rho_n$ for $n = k$. Then, partitioning R_{k+1} and defining

$$\rho_k^{(r)} := (\rho(k), \rho(k-1), ..., \rho(1))'$$

and

$$\phi_k^{(r)} := (\phi_{kk}, \phi_{k,k-1}, ..., \phi_{k1})',$$

we see that the recursions imply

$$\begin{aligned} R_{k+1}\phi_{k+1} &= \begin{bmatrix} R_k & \rho_k^{(r)} \\ \rho_k^{(r)'} & 1 \end{bmatrix} \begin{bmatrix} \phi_k - \phi_{k+1,k+1}\phi_k^{(r)} \\ \phi_{k+1,k+1} \end{bmatrix} \\ &= \begin{bmatrix} \rho_k - \phi_{k+1,k+1}\rho_k^{(r)} + \phi_{k+1,k+1}\rho_k^{(r)} \\ \rho_k^{(r)'}\phi_k - \phi_{k+1,k+1}\rho_k^{(r)'}\phi_k^{(r)} + \phi_{k+1,k+1} \end{bmatrix} \\ &= \rho_{k+1}, \end{aligned}$$

as required. Here we have used the fact that if $R_k\phi_k = \rho_k$, then $R_k\phi_k^{(r)} = \rho_k^{(r)}$. This is easily checked by writing out the component equations in reverse order. Since $R_n\phi_n$ is satisfied for $n = 1$, it follows by induction that the coefficient vectors ϕ_n defined recursively by

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1},$$

and

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

to satisfy $R_n \phi_n = \rho_n$ for all n.

It remains only to establish that the mean squared errors

$$v := \mathbb{E}(Xn + 1 - \phi_n' \mathbf{X}_n)^2$$

satisfy $v_0 = \gamma(0)$ and $v_n = v_{n-1}[1 - \phi_{nn}^2]$. The fact that $v_0 = \gamma(0)$ is an immediate consequence of the definition $P_0 X_1 := \mathbb{E}(X_1) = 0$. Since we have shown that $\phi_n' \mathbf{X}_n$ is the best linear predictor of X_{n+1}, we can writ,e

$$v_n = \gamma(0) - \phi_n' \gamma_n = \gamma(0) - \phi_{n-1}' \gamma_{n-1} + \phi_{nn} \phi_{n-1}^{(r)'} \gamma_{n-1} - \phi_{nn} \gamma(n).$$

Applying

$$\begin{aligned} \mathbb{E}(X_{n+h} - P_n X_{n+h})^2 &= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma(h + i - 1) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma(i - j) a_j \\ &= \gamma(0) - \mathbf{a}_n' \gamma_n(h), \\ &= \gamma(0) - \mathbf{a}_n' \Gamma_n \mathbf{a}_n \end{aligned}$$

again gives us

$$v_n = v_{n-1} + \phi_{nn} \left(\phi_{n-1}^{(r)'} \gamma_{n-1} - \gamma(n) \right),$$

and hence, by

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n - j) \right] v_{n-1}^{-1},$$

there is

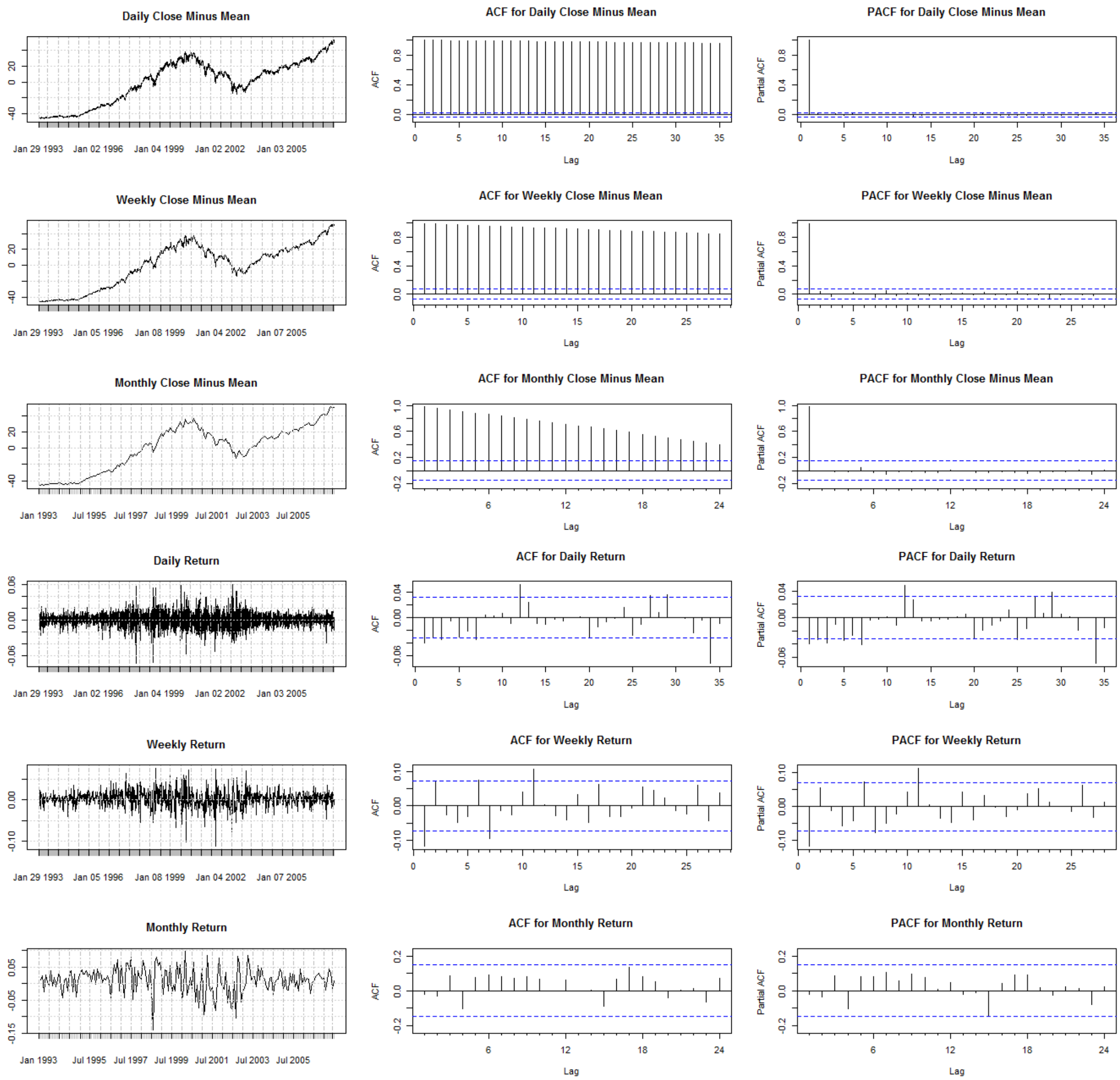
$$v_n = v_{n-1} - \phi_{nn}^2 (\gamma(0) - \phi_{n-1}' \gamma_{n-1}) = v_{n-1}(1 - \phi_{nn}^2).$$

Q.E.D.

6 Figures

Figure 1: This figure plots, in Panel A, daily closing price, weekly closing price, and monthly closing price of S&P 500 Index Fund. Moreover, the figure presents ACF and PACF for closing price and returns for daily, weekly, and monthly data. In Panel B, we present forecasting resulting.

(Panel A)



(Panel B)

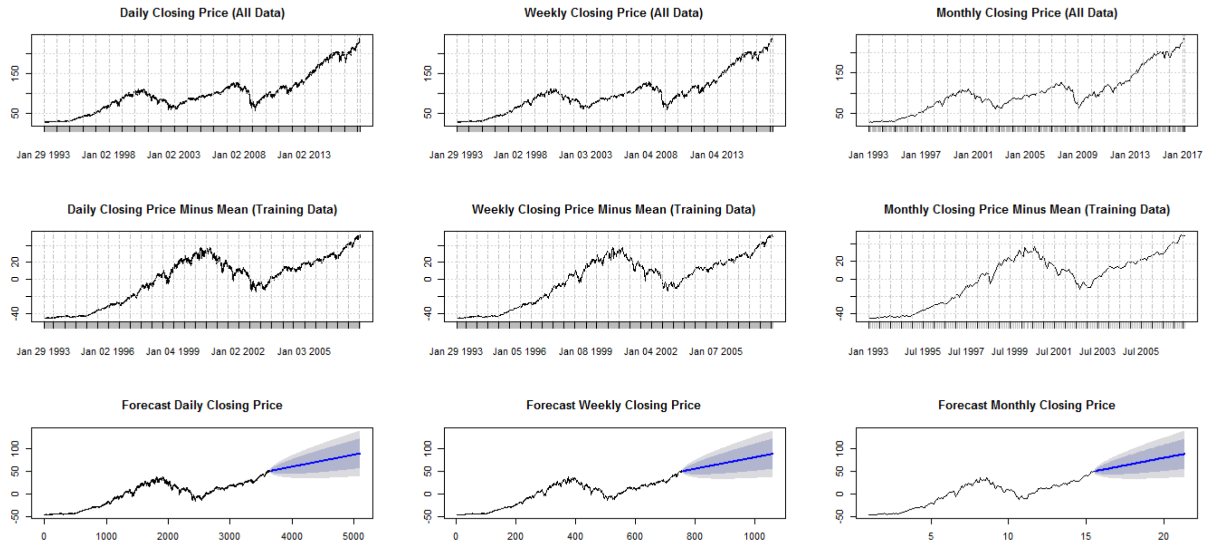
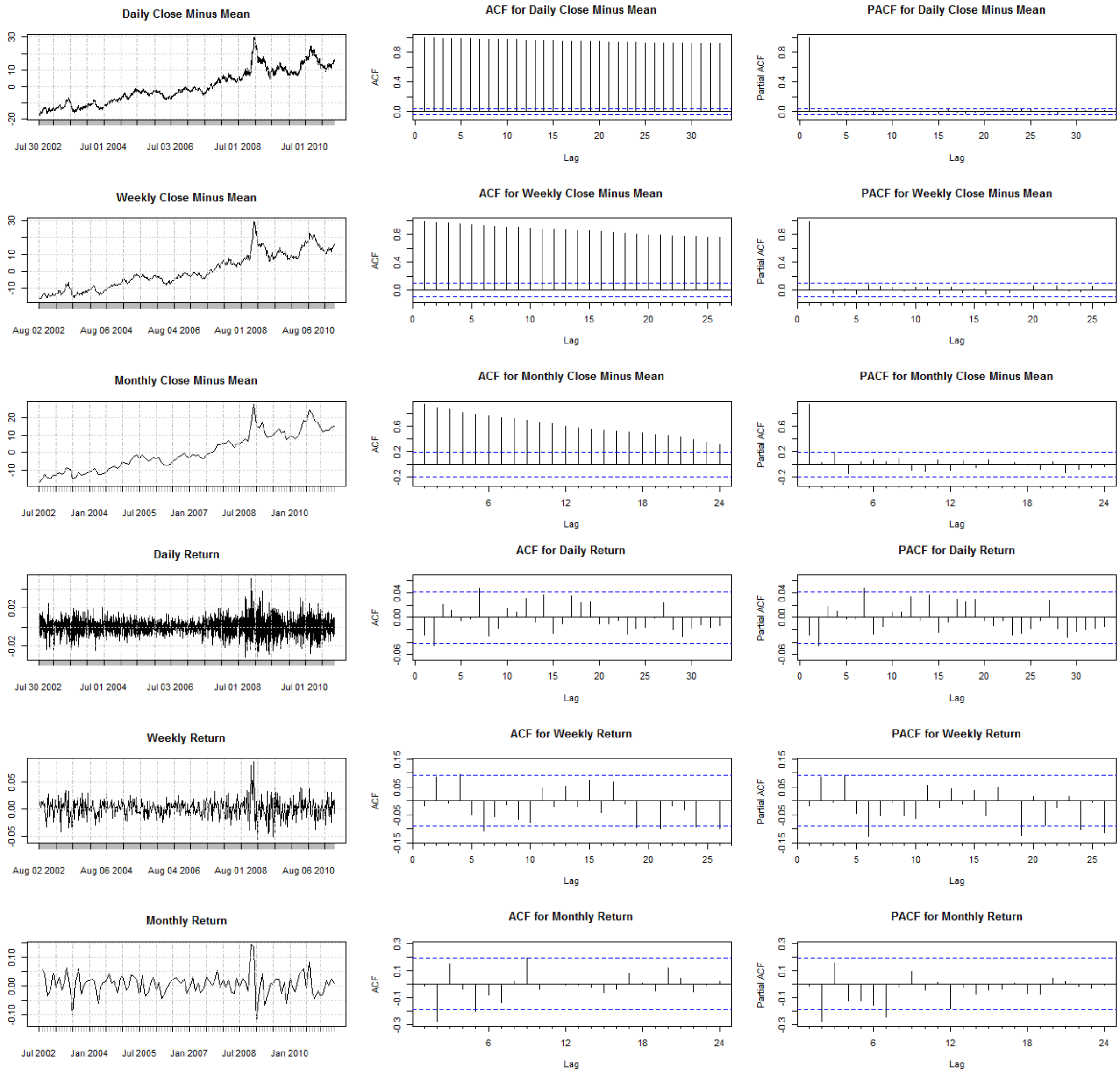

Daily Closing Price (All Data)
Weekly Closing Price (All Data)
Monthly Closing Price (All Data)
Jan 29 1993
Jan 02 1998
Jan 02 2003
Jan 02 2008
Jan 02 2013
Jan 03 2003
Jan 04 2008
Jan 04 2013
Jan 1993
Jan 1997
Jan 2001
Jan 2005
Jan 2009
Jan 2013
Jan 2017
Daily Closing Price Minus Mean (Training Data)
Weekly Closing Price Minus Mean (Training Data)
Monthly Closing Price Minus Mean (Training Data)
Jan 02 1996
Jan 04 1999
Jan 02 2002
Jan 03 2005
Jan 05 1996
Jan 08 1999
Jan 04 2002
Jan 07 2005
Jul 1995
Jul 1997
Jul 1999
Jul 2001
Jul 2003
Jul 2005
Forecast Daily Closing Price
Forecast Weekly Closing Price
Forecast Monthly Closing Price

Figure 2: This figure plots, in Panel A, daily closing price, weekly closing price, and monthly closing price of 20-Year Treasury-Bill Futures (TLT). Moreover, the figure presents ACF and PACF for closing price and returns for daily, weekly, and monthly data. In Panel B, we present forecasting resulting.

(Panel A)



(Panel B)

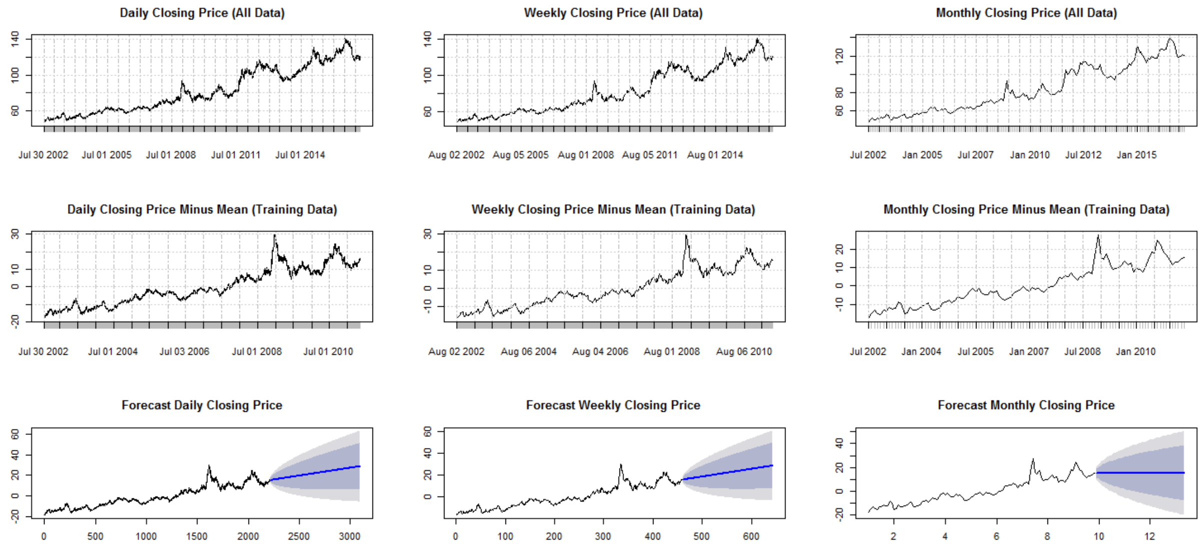
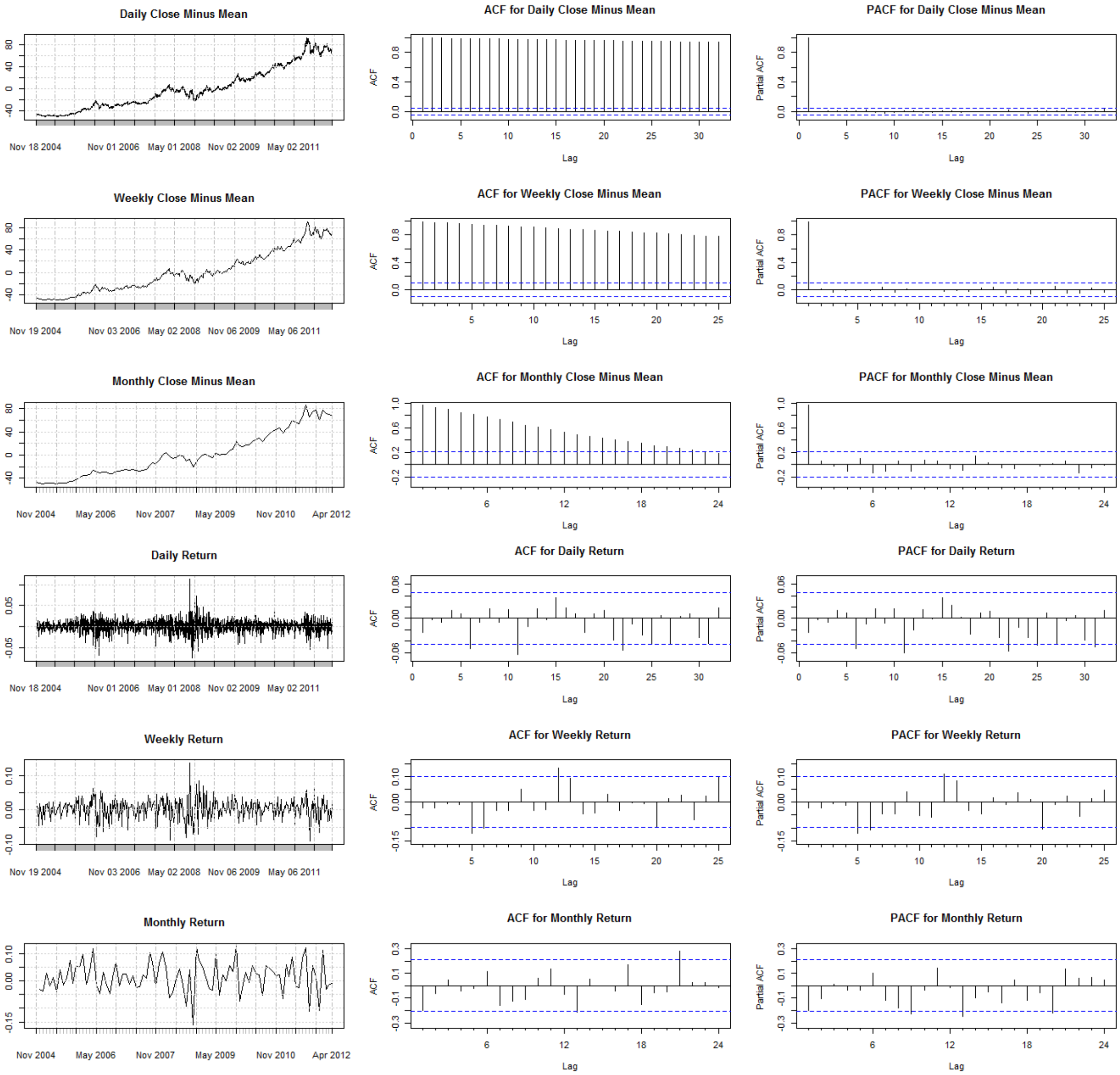
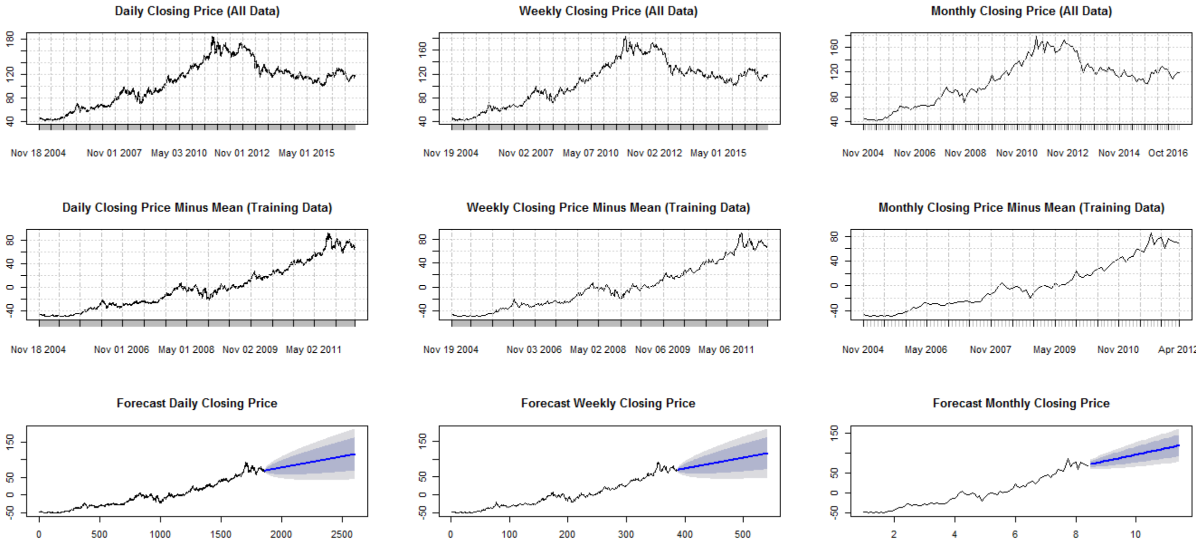

Daily Closing Price (All Data)
Weekly Closing Price (All Data)
Monthly Closing Price (All Data)
Jul 30 2002 Jul 01 2005 Jul 01 2008 Jul 01 2011 Jul 01 2014
Aug 02 2002 Aug 05 2005 Aug 01 2008 Aug 05 2011 Aug 01 2014
Jul 2002 Jan 2005 Jul 2007 Jan 2010 Jul 2012 Jan 2015
Daily Closing Price Minus Mean (Training Data)
Weekly Closing Price Minus Mean (Training Data)
Monthly Closing Price Minus Mean (Training Data)
Jul 30 2002 Jul 01 2004 Jul 03 2006 Jul 01 2008 Jul 01 2010
Aug 02 2002 Aug 06 2004 Aug 04 2006 Aug 01 2008 Aug 06 2010
Jul 2002 Jan 2004 Jul 2005 Jan 2007 Jul 2008 Jan 2010
Forecast Daily Closing Price
Forecast Weekly Closing Price
Forecast Monthly Closing Price
0 500 1000 1500 2000 2500 3000
0 100 200 300 400 500 600
2 4 6 8 10 12

Figure 3: This figure plots, in Panel A, daily closing price, weekly closing price, and monthly closing price of Gold Futures (GLD). Moreover, the figure presents ACF and PACF for closing price and returns for daily, weekly, and monthly data. In Panel B, we present forecasting resulting.

(Panel A)



(Panel B)


Daily Closing Price (All Data)
Nov 18 2004
Nov 01 2007
May 03 2010
Nov 01 2012
May 01 2015
Weekly Closing Price (All Data)
Nov 19 2004
Nov 02 2007
May 07 2010
Nov 02 2012
May 01 2015
Monthly Closing Price (All Data)
Nov 2004
Nov 2006
Nov 2008
Nov 2010
Nov 2012
Nov 2014
Oct 2016
Daily Closing Price Minus Mean (Training Data)
Nov 18 2004
Nov 01 2006
May 01 2008
Nov 02 2009
May 02 2011
Weekly Closing Price Minus Mean (Training Data)
Nov 19 2004
Nov 03 2006
May 02 2008
Nov 06 2009
May 06 2011
Monthly Closing Price Minus Mean (Training Data)
Nov 2004
May 2006
Nov 2007
May 2009
Nov 2010
Apr 2012
Forecast Daily Closing Price
Forecast Weekly Closing Price
Forecast Monthly Closing Price

7 Tables

Table 1. This table presents correlation with all three indices including adjusted closing price (w/ dividends) and price returns.

Panel (A)			
	GLD Adjusted	SPY Adjusted	TLT Adjusted
GLD Adjusted	1	0.2964067	0.6538499
SPY Adjusted	0.2964067	1	0.8103385
TLT Adjusted	0.6538499	0.8103385	1
	GLD Return	SPY Return	TLT Return
GLD Return	1	0.03912509	0.1073758
SPY Return	0.03912509	1	-0.4277832
TLT Return	0.10737578	-0.42778323	1

Panel (B)			
Min.:	1/29/1993	Min.:	27.54
1st Qu.:	2/6/1999	1st Qu.:	70.47
Median:	2/24/2005	Median:	94.58
Mean:	2/22/2005	Mean:	100.69
3rd Qu.:	3/9/2011	3rd Qu.:	117.75
Max.:	3/24/2017	Max.:	238.74
Index	GLD.Adjusted		
Min.:	11/18/2004	Min.:	41.26
1st Qu.:	12/19/2007	1st Qu.:	77.34
Median:	1/20/2011	Median:	112.73
Mean:	1/21/2011	Mean:	106.14
3rd Qu.:	2/24/2014	3rd Qu.:	127.75
Max.:	3/24/2017	Max.:	184.59
Index	TLT.Adjusted		
Min.:	7/30/2002	Min.:	46.69
1st Qu.:	3/28/2006	1st Qu.:	61.15
Median:	11/23/2009	Median:	76.5
Mean:	11/25/2009	Mean:	83.22
3rd Qu.:	7/25/2013	3rd Qu.:	105.4
Max.:	3/24/2017	Max.Qu.:	141.31

Table 2. This table presents correlation with all three indices including adjusted closing price (w/ dividends) and price returns.

Panel (A)				Coefficients	SE	T-stat
SPY	Daily	(1,1)	AR1	0.7535	0.0705	10.68794326
			MA1	-0.8055	0.0634	-12.70504732
			Drift	0.0264	0.0111	2.378378378
	Weekly	(1,0)	AR1	-0.1388	0.036	-3.855555556
			Drift	0.1268	0.0545	2.326605505
	Monthly	(0,0)	Drift	0.5502	0.2347	2.34426928
GLD	Daily	(0,0)	Drift	0.0625	0.0309	2.022653722
	Weekly	(0,0)	Drift	0.3007	0.1424	2.111657303
	Monthly	(0,1)	MA1	-0.4002	0.0987	-4.054711246
			Drift	1.3386	0.3658	3.659376709
TLT	Daily	(0,0)	Drift	0.0151	0.0125	1.208
	Weekly	(0,0)	Drift	0.0703	0.0562	1.25088968
	Monthly	(0,0)	Drift			

Panel (B)			Sigma Square	Log Likelihood	AIC	AICC	BIC
SPY	Daily	(1,1)	0.7169	-4567.83	9143.65	9143.66	9168.46
	Weekly	(1,0)	2.92	-1476.83	2959.66	2959.69	2973.54
	Monthly	(0,0)	9.643	-443.56	891.11	891.18	897.43
GLD	Daily	(0,0)	1.784	-3182.25	6368.49	6368.5	6379.55
	Weekly	(0,0)	7.852	-944.93	1893.87	1893.9	1901.78
	Monthly	(0,1)	33.31	-281.37	568.74	569.03	576.21
TLT	Daily	(0,0)	0.3433	-1956.72	3917.44	3917.44	3928.84
	Weekly	(0,0)	1.45	-734.39	1472.79	1472.81	1481.04
	Monthly	(0,0)	7.699	-258.58	1472.79	519.21	521.83

References

[1] De Bondt, Werner F. M. and Thaler, Richard (1985), "Does the Stock Market Overreact?" *The Journal of Finance*, 40(3), 793-805.

[2] Markowitz, Harry (1959), "Portfolio Selection: Efficient Diversification of Investments", *Wiley, New York*.

[3] Sharpe, William, F. (1964), "Capital Asset Prices: a Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance*, 19(3), 425-442.