SEMIFARMA-HYGARCH Modeling of Dow Jones Return Persistence

Mohamed Chikhi, Anne Peguin-Feissolle, Michel Terraza

Working Papers / Documents de travail

SEMIFARMA-HYGARCH Modeling of Dow Jones Return Persistence

Mohamed Chikhi
Anne Péguin-Feissolle
Michel Terraza

WP 2012 - Nr 14

SEMIFARMA-HYGARCH MODELING OF DOW JONES RETURN PERSISTENCE

Mohamed Chikhi[1]

Anne Péguin-Feissolle[2] **Michel Terraza**[3]

June, 2012

Abstract

This paper analyzes the cyclical behavior of Dow Jones by testing the existence of long memory through a new class of semiparametric ARFIMA models with HYGARCH errors (SEMIFARMA-HYGARCH); this class includes nonparametric deterministic trend, stochastic trend, short-range and long-range dependence and long memory heteroscedastic errors. We study the daily returns of the Dow Jones from 1896 to 2006. We estimate several models and we find that the coefficients of the SEMIFARMA-HYGARCH model, including long memory coefficients for the equations of the mean and the conditional variance, are highly significant. The forecasting results show that the informational shocks have permanent effects on volatility and the SEMIFARMA-HYGARCH model has better performance over some other models for long and/or short horizons. The predictions from this model are also better than the predictions of the random walk model; accordingly, the weak efficiency assumption of financial markets seems violated for Dow Jones returns studied over a long period.

Keywords: *SEMIFARMA model, HYGARCH model, nonparametric deterministic trend, kernel methodology, long memory*

JEL Classification: *C14, C22, C58, G17*

[1] Université de Ouargla & LAMETA/CNRS, Université Montpellier I, Faculté des Sciences Economiques, Espace Richter, Avenue de la Mer, C.S. 79606, 34960 Montpellier Cedex 2, France. E-mail: chikhi@lameta.univ-montp1.fr

[2] **Corresponding author:** GREQAM/CNRS, Université d'Aix-Marseille 2, Centre de la Charité, 2 rue de la Charité, 13236 Marseille cedex 02, France, Tel: +33 (0)4.91.14.07.70, Fax: +33.(0)4.91.90.02.27, E-mail: Anne.Peguin@univmed.fr

[3] LAMETA/CNRS, Université Montpellier I, Faculté des Sciences Economiques, Espace Richter, Avenue de la Mer, C.S. 79606, 34960 Montpellier Cedex 2, France. E-mail: mterraza@lameta.univ-montp1.fr

1. Introduction

Stock prices have specific statistical properties whose consideration is fundamental to address the problem of modeling. Thus, the presence of long memory in financial series has important implications especially concerning the property of weak efficiency of markets. Indeed, if a series exhibits long memory, this implies significant autocorrelations between observations that, although they are distant in time, can help predict future returns, which violates the assumption of market efficiency (see, among others, Fama 1965, 1970 and 1998; Grossman 1976; Jensen 1978; Lillo and Farmer 2004; Christodoulou-Volos and Siokis 2006; Barkoulas and Baum 1997; Gursakal 2010). Therefore, modeling persistence in financial time series has been a high priority in the field of economic research. Moreover, time series often exhibit deterministic or stochastic trends. Thus, Beran (1999) proposed a SEMIFAR (*Semiparametric fractional autoregressive*) model; this model is interesting because it takes into account both short-term behavior of the series through autoregressive parameters, long-term behavior through the parameter of fractional integration and also the nonparametric deterministic trend. Beran and Feng (2002a, 2002b) then introduce the moving average part into the SEMIFAR model, calling it SEMIFARMA (SEMIFAR *Moving Average*) model. The SEMIFARMA model thus includes the ARIMA model and the fractionally autoregressive process (Granger and Joyeux 1980; Hosking 1981). The assumption of white noise on the SEMIFARMA model residuals ignores the presence of conditional heteroskedasticity; however, the financial series are generally characterized by a time-varying volatility that can be modeled by ARCH-type models (Engle 1982; Bollerslev 1986). The SEMIFARMA-GARCH model proposed by Feng et al. (2007) corresponds to a specific representation of nonlinearity allowing for simplified modeling of uncertainty. We will extend this model by using the HYGARCH (*Hyperbolic* GARCH) model (Davidson 2004) that provides a direct measure of persistence through the fractional integration parameter.

The present article belongs to the field of the above mentioned research work. It is applied to analysis of the persistence of informational shocks and to the search for a possible long memory in Dow Jones returns. By studying the daily returns of this stock price over a long period, from 1896 to 2006 (30292 observations), we show that the SEMIFARMA-HYGARCH model has a significant predictive superiority over all other proposed models for long horizons. The predictions from this model are also better than the predictions of the random walk model, either short term or long term; accordingly, the weak efficiency assumption of financial markets seems violated for Dow Jones returns studied over a long period.

The rest of the paper is organized as follows. Section 2 focuses on presentation of the SEMIFARMA-HYGARCH model. Section 3 is devoted to empirical study of the daily series of Dow Jones; we compare the predictive quality of SEMIFARMA-GARCH, SEMIFARMA-FIGARCH and SEMIFARMA-HYGARCH models with that of a random walk. We give some conclusions in Section 4.

2. Presentation of the SEMIFARMA- HYGARCH model

Some authors have used the HYGARCH model or the SEMIFAR model to study financial time series, but always separately. Thus, concerning the HYGARCH model, Davidson (2004) examines the volatility dynamics of three Asian currencies. Tang and Shieh (2006) analyze the value-at-risk in the case of three stock indexes (S&P 500, Dow Jones and Nasdaq 100). Cardamone and Folkinshteyn (2007) study the sensitivity of U.S. interest rates and exchange rates. Härdle and Mungo (2008) focus on the value-at-risk of some stock indexes. McMillan and Kambouroudis (2009) compare the predictions of the HYGARCH model and other models for 31 stock indexes. Kasman et al. (2009) investigate the presence of long memory in eight stock indexes of countries from Central Europe and Eastern Europe. Aloui and Mabrouk (2010) consider the value-at-risk for oil and gas. Conrad (2010) derives the conditions for non-negativity of the conditional variance in a HYGARCH model and presents two empirical applications, one for the daily FTSE, the other for the DAX30. Wei et al. (2010) are concerned with the volatility of the oil market. Finally, Kwan et al. (2011) study some theoretical properties of the HYGARCH model and give an empirical illustration devoted to the exchange rates of the Korean currency. Regarding the SEMIFAR modeling, Beran and Ocker (1999a) apply SEMIFAR to model some European and Asian exchange rates; Beran and Ocker (1999b) and Feng et al. (2007) study different stock series.

The *SEMIFARMA-HYGARCH* model we use is as follows (see Beran and Feng (2002a) for the *SEMIFAR* part and Davidson (2004) for the HYGARCH part). $\{Y_t\}$ is a fractional semiparametric process with hyperbolic GARCH error, called *SEMIFARMA-HYGARCH*, if it verifies the following relationship:

$$\phi(B)(1-B)^{d_2}\{(1-B)^{d_1}Y_t - g(x_t)\} = \theta(B)\varepsilon_t \tag{1}$$

with $$\varepsilon_t = u_t\sigma_t\ ,\quad \sigma_t > 0\ ,\ u_t \sim iid(0,1) \tag{2}$$

and $$\sigma_t^2 = \omega^* + \sum_{i=1}^{\infty}\lambda_i B^i \varepsilon_t^2 = \omega^* + \lambda(B)\varepsilon_t^2 \tag{3}$$

where $$\lambda(B) = 1 - [1-\beta(B)]^{-1}\psi(B)\{1+\alpha[(1-B)^d - 1]\} \tag{4}$$

and $$\omega^* = \omega(1-\beta(B))^{-1} \tag{5}$$

with $(1-B)^d = \sum_{k=0}^{\infty}(-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}B^k = 1 - \sum_{k=1}^{\infty}c_k(d)B^k$;moreover, $c_1(d) = d$, $c_2(d) = \frac{1}{2}d(1-d)$, and $\Gamma(.)$ is the gamma function. Furthermore, we have $\phi(B) = 1-\phi_1 B - \phi_2 B^2 - .... - \phi_p B^p$, $\theta(B) = 1-\theta_1 B - \theta_2 B^2 - .... - \theta_q B^q$ and $\beta(B) = 1-\beta_1 B - ... - \beta_r B^r$; $\psi(B) = 1+\psi_1 B + \psi_2 B^2 + ... + \psi_s B^s$ and $\lambda(B)$ denote polynomials in the lag operator with degrees given by *p*, *q*, *r*, *s* and ∞ respectively, with roots outside the unit circle. *B* is the lag operator, d_1 is an integer: $d_1 \in \{0,1\}$, $x_t = t$ is the trend and g : [0, 1] → R is the smoothing function. Finally, the process is stationary and invertible, $-\frac{1}{2} < d_2 < \frac{1}{2}$, and u_t is an i.i.n. process. Davidson (2004) shows that the HYGARCH process generalizes the FIGARCH model (Baillie et al. 1996). If $\alpha = 1$, the HYGARCH model corresponds to a FIGARCH model. The process is stationary if $0 < \alpha < 1$ and nonstationary if $\alpha > 1$. The long memory intervenes in the mean equation (1) (through the parameter d_2) and in the variance equation (3) (through the parameter *d*). It is important to note that the definition of long memory is an asymptotic definition; in other words, what matters is how the autocorrelations converge to zero when the lag increases. To summarize, the distinction, in terms of autocorrelations, between a short memory process and a long memory process results from the speed of geometric or hyperbolic convergence of the autocorrelations toward zero.

Finally, the $g(x_t)$ function is a nonlinear deterministic trend. The kernel estimation in the case of nonparametric regression has been studied especially by Hall and Hart (1990), Ray and Tsay (1997) and Beran (1999) in the case of long memory errors. We estimate g by the kernel method using the following model:

$$(1-B)^{d_1} Y_t = g(x_t) + X_t \qquad (6)$$

It is equivalent to $Y_t = g(x_t) + X_t$ if $d_1 = 0$ and to $Y_t - Y_{t-1} = g(x_t) + X_t$ if $d_1 = 1$, where X_t is a process of long memory stationary errors if d_2 >0 and short memory errors if d_2 =0 ; the case d_2 <0 is sometimes called "anti-persistence". We consider the polynomial kernel defined by:

$$K(x) = \sum_{l=0}^{\tau} \alpha_l x^{2l} \text{ with} |x| \le 1 \qquad (7)$$

and $K(x) = 0$ if $|x|$>1. Here we have $\tau \in \{0, 1, 2, \ldots\}$ and the α_t coefficients are such that $\int_{-1}^{1} K(x)dx = 1$ (see Beran and Feng (2002a and b) for details on the estimation method).

3. Empirical analysis of the Dow Jones

The empirical study focuses on the logarithmic series of daily Dow Jones covering a historical period from May 26, 1896 to August 17, 2006 (n = 30292). Unit root tests results (Dickey and Fuller 1981; Phillips and Perron 1988; Schmidt and Phillips 1992; Kwiatkowski et al. 1992; Elliott et al. 1996) show that this series is characterized by a unit root (see Table 1). The series is finally differentiated to obtain the returns (see Figure 1).

[Insert Table 1 here]

[Insert Fig. 1 here]

The assumption of normality of returns is clearly rejected (see Table 2 and Figure 2). The observed asymmetry may indicate the presence of nonlinearities in the evolution process of returns. The scatter plot of the series (Figure 3) does not appear in the form of a regular ellipsoid, and confirms nonlinearity. In addition, the series is heteroscedastic according to the results of White and Breusch-Pagan tests reported in Table 3, since the null hypothesis of homoscedasticity is rejected at 5%. The conditional heteroskedasticity test result shows that Dow Jones returns are characterized by the presence of an ARCH effect frequently encountered in financial time series [$nR^2 = 1114.3201 > \chi^2_{0.05}(2)$].

[Insert Table 2 here]

[Insert Fig. 2 here]

[Insert Table 3 here]

[Insert Fig. 3 here]

In view of Table 4, the random walk hypothesis is clearly rejected. Indeed, Mizrach (Mizrach 1995) and BDS (Brock et al. 1987) statistics, which test the presence of linear or nonlinear dependences, are strictly greater than the critical value at 5%.

[Insert Table 4 here]

These first tests generally highlight the presence of significant non-zero autocorrelations in the short term; even if these autocorrelations are significant from a statistical point of view, they are not significant from an economic point-of-view in that it is impossible to exploit these autocorrelations to establish speculative rules leading to abnormal profits. Tests lead us to reject the null hypothesis of no autocorrelation, but do not detect the presence of a structure of long-term dependence. Given this situation, we analyze the cyclical behavior of returns by considering longer horizons. By plotting the periodogram of this series (see Figure 4) (with Tuckey window), we note that the spectral density is concentrated around low frequencies and tends to infinity when the frequency tends to zero. This is a sign of long memory that we have to verify with statistical tests.

[Insert Fig. 4 here]

For this reason, we choose different values for the periodogram ordinates around the square root of the number of observations. This choice aims to examine the stability of the estimators when the number of the periodogram ordinates varies. From Table 5, it is obvious that the series of Dow Jones returns is generated by a long memory process. Some values of the Student statistic (with a power of 0.8) are strictly greater than the critical value at 5%. In addition, the memory parameter estimated by the Andrews and Guggenberger (2003) method is positive and significant. The estimation result is very close to those found with the GPH (Geweke and Porter-Hudak 1983) method. The returns are long-term predictable: the presence of a long memory indicates that agents can anticipate their returns to a sufficiently long time horizon. Indeed, the observed movements appear as the result of lasting exogenous shocks which affect the New York market, i.e. the return will not come back to its fundamental value.

[Insert Table 5 here]

The results of the SEMIFARMA model estimation by the exact maximum likelihood method are shown in Table 6. After estimating the deterministic trend, the optimal window and the cross-validation criteria by the kernel method based on the methodology of Nadaraya-Watson (Nadaraya 1964; Watson 1964), the results indicate that the Dow Jones series is characterized by a long memory: the estimated fractional integration parameter is significantly different from zero. This result is consistent with the different spectral method estimators.

[Insert Table 6 here]

We note that the residuals (Figure 5) are not characterized by a Gaussian distribution (Table 7) and are leptokurtic (Figure 6). The asymmetry may indicate the presence of nonlinearities in the residuals. However, these residuals can be modeled by GARCH models because the presence of an ARCH effect is confirmed by the result of the ARCH-LM test on residuals ($nR^2 = 20.1646 > \chi^2(1)$).Thus, the spectrum of squared residuals (see Figure 7) is concentrated around low frequencies and tends to infinity as the frequency tends to zero. It is likely that the conditional variance has a persistent long memory structure. The financial asset prices often exhibit heteroscedastic behavior with persistence. For this reason, we will study the conditional variance of Dow Jones returns, to consider the possibility of lasting shocks on volatility.

[Insert Fig. 5 here]

[Insert Table 7 here]

[Insert Fig. 6 here]

[Insert Fig. 7 here]

The process followed by Dow Jones returns is then estimated by the simultaneous maximization of the mean and variance processes. The tests allow the adequacy of the model to be appraised, firstly, concerning the absence of autocorrelation and of heteroscedasticity, and secondly, concerning the existence of long memory and the good modeling of the ARCH effect. In practical terms, we estimate several models for different lags: a SEMIFARMA (*p, d, q*) jointly with a GARCH, FIGARCH and HYGARCH model. For each model, we calculate both Akaike (1970) and Schwarz (1978) information criteria. The estimation of different models is only based on 30032 observations, in order to make further comparisons with the predictions of the 260 remaining observations.

In view of Table 8, we find that the information criteria are minimum for the SEMIFARMA-HYGARCH model and the coefficients of this model are highly significant. In addition, long memory coefficients for the equations of the mean and the conditional variance are also significant. The residuals of the models are characterized by the absence of conditional heteroskedasticity: the ARCH-LM statistics are strictly less than the critical value of χ_2^2 at 5%. It should be noted that the normality assumption of residuals of the models is clearly rejected because the Jarque-Bera statistics are strictly greater than the critical value of χ_2^2 at 5%. In view of Table 9, the series of the SEMIFARMA-HYGARCH residuals show no dependence structure where the BDS statistics are strictly less than the critical value 1.96.

[Insert Table 8 here]

[Insert Table 9 here]

To compare the forecasting performance of the proposed models and the random walk model, two criteria are used: the mean squared error (MSE) and the mean absolute error (MAE) given by

$$MSE = H^{-1}\sum_{h=1}^{H}(\hat{Y}_{n-H+h} - Y_{n-H+h})^2 \qquad (8)$$

$$MAE = H^{-1}\sum_{h=1}^{H}\left|\hat{Y}_{n-H+h} - Y_{n-H+h}\right| \qquad (9)$$

where *h* is the forecasting horizon and *H* is the total number of predictions for the horizon h over the forecast period.

Table 10 contains the results of in-sample predictions provided by the different models. MSE and MAE criteria generally give the same results. We note that, whatever the forecast horizon, the random walk model is beaten by all the other models. We generally find good predictive results from the SEMIFARMA-FIGARCH model and especially the SEMIFARMA-HYGARCH model with a horizon of 90, 180 and 260 days. As shown in Table 11, which gives the statistical comparisons of out-of-sample forecasts, all the models beat also the random walk model and we find good predictive results from the SEMIFARMA-HYGARCH with a horizon of 30, 90 and 180 days. Indeed, the random walk takes into account only the short-term memory of the series and therefore completely neglects the long-term memory. MSE and MAE criteria give nearly the same results. Consequently, as shown in Tables 10 and 11, the long memory models of conditional variance, such as the SEMIFARMA-HYGARCH model, provide superior quality forecasts over a long horizon. Taking into account the long memory implies that, for a long-term forecast (180 or 260 days), not only the last observed value but also the entire weighted history of this series should be used.

[Insert Table 10 here]

[Insert Table 11 here]

Given that the Dow Jones returns are characterized by the presence of long-term dynamics in the equations of the mean and conditional variance and by heteroscedasticity, the SEMIFARMA-HYGARCH model allows computation of better short-term and long term forecasts than the random walk model. This model is also clearly superior to all other models for long horizons.

In order to test the statistical significance of the forecasting improvements obtained with the SEMIFARMA-HYGARCH model over the random-walk model, we can use also a battery of tests based on loss functions: the asymptotic test, the sign tests, then Wilcoxon's test, the Naive benchmark test, the Morgan-Granger-Newbold test and the Meese Rogoff test (all these tests are summarized in Diebold and Mariano (1995)). The p-values in table 12 clearly indicate that the null hypothesis of equal accuracy of the two models is strongly rejected (the p-values are less than 0.05). So, we accept different predictive accuracy; it means that, in this case, the SEMIFARMA-HYGARCH model beats the random walk process. The evidence is encouraging for the predictive ability of non-linear fractional models for the Dow Jones returns (the same conclusion is given in Barkoulas and Baum (2006) for some US monetary indices).

[Insert Table 12 here]

Although the gap between the MSE and MAE criteria does not appear very important, what is significant in this study is the systematic nature of the models. Indeed, the price movements appear as the result of lasting exogenous shocks which affect the U.S. stock market; in other words, the consequences of a shock will be sustainable, the Dow Jones returns will not come back to their previous fundamental value and the shock will be persistent in the long term. This suggests that, due to the long-term predictability of returns, it will be possible *a priori* to establish remunerative strategies on the New York stock market.

4. Conclusion

In this article, we investigated the presence of long memory in the Dow Jones returns. In this context, we proposed a semiparametric long memory model called SEMIFARMA with hyperbolic GARCH errors. We implemented the exact maximum likelihood method to estimate exactly this class of models by taking into account the phenomenon of long-term persistence for the conditional variance. From the results, informational shocks have lasting effects on volatility and the SEMIFARMA-HYGARCH model shows a clear superiority over all the other proposed models for long horizons. Specifically, the forecasts of the long memory model show a clear improvement compared to the random walk model at all horizons. The returns have long term memory structure because distant observations are positive autocorrelated: future predictions are improved by past returns and this put into question the validity of the efficient market hypothesis. It appears that the Dow Jones returns are long-term dependent, suggesting stock market inefficiency. Consequently, low efficiency of financial markets seems violated for the Dow Jones returns studied over a long period. Thus, recent works on volatility modeling through FIGARCH or HYGARCH processes seem particularly promising and may provide new evidence to better understand the dynamics of financial series.

References

Akaike, H. (1970). Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics*, 22, 203-217.

Aloui, C, & Mabrouk, S. (2010). Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. *Energy Policy*, 38, 2326–2339

Andrews, D. W. K., & Guggenberger, P. (2003). A Bias-Reduced Log-Periodogram Regression Estimator for the Long-memory Parameter. *Econometrica*, 71, 675-712.

Baillie, R.T., Bollerslev, T, & Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74, 3-30.

Barkoulas, J.T., & Baum C.F. (1997). Long memory and forecasting in Euroyen deposit rates. *Financial Engineering and the Japanese Market*, 4, 189-201.

Barkoulas, J.T., & Baum C.F. (2006). Long-memory forecasting of US monetary indices, *Journal of Forecasting*, Vol. 25, Issue 4, 291-302.

Beran, J. (1999). SEMIFAR models—a semiparametric framework for modelling trends, long-range dependence and nonstationarity. Discussion Paper No. 99/16, Center of Finance and Econometrics, University of Konstanz.

Beran, J., & Feng, Y. (2002a). SEMIFAR models - A semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statististics & Data Analysis*, 40, 393–419.

Beran, J., & Feng, Y. (2002b). Iterative plug-in algorithms for SEMIFAR models -definition, convergence and asymptotic properties. *Journal of Computational and Graphical Statistics*, 11, 690-713.

Beran, J., & Ocker, D. (1999a). SEMIFAR forecasts, with applications to foreign exchange rates. *Journal of Statistical Planning and Inference*, 80, 137-153.

Beran, J., & Ocker, D. (1999b). Volatility of stock market indices - an analysis based on SEMIFAR models. Working Paper, Department of Economics and Statistics, University of Konstanz, Germany.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

Brock, W. A., Dechert, W. D., & Scheinkman, J. (1987). A test for independence based on the correlation dimension. Discussion Paper 8702, University of Wisconsin-Madison.

Cardamone E., & Folkinshteyn D. (2007). HYGARCH Approach to Estimating Interest Rate and Exchange Rate Sensitivity of a Large Sample of U.S. Banking Institutions. Working Paper, Temple University, Department of Finance, USA.

Christodoulou-Volos, C.C, & Siokis, F. M. (2006). Long Range Dependence in Stock Market Returns. *Applied Financial Economics*, 16, 1331-1338.

Conrad, C. (2010). Non-negativity conditions for the hyperbolic GARCH model. *Journal of Econometrics*, 157, 441-457.

Davidson, J., (2004). Moments and memory properties of linear conditional heteroskedasticity models and a new model. *Journal of Business and Economic Statistics*, 22, 16-29.

Dickey, D., & Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427-431.

Diebold, F.X and Mariano, R.S. (1995). Comparing Predictive accuracy. *Journal of Business and Economic Statistics*, 13, 3, 253-263.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64, 4, 813–836.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimation of U.K. inflation. *Econometrica*, 50, 987–1008.

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21(5), 55–59.

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383-417.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49, 283–306.

Feng, Y., Beran, Y., & Yu, K. (2007). Modelling financial time series with SEMIFAR–GARCH model. *IMA Journal of Management Mathematic*, 18, 395-412.

Geweke, J., & Porter-Hudak, S., (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis*, 4, 221-238.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15-30.

Grossman, S.J. (1976). On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information. *The Journal of Finance*, 31(2), 573-585.

Gursakal, S. (2010). Detecting Long Memory in Bulls and Bears Markets: Evidence from Turkey. *Journal of Money, Investment and Banking*, 18, 95-104.

Hall, P., & Hart, J., 1990. Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications*, 36, 339–351.

Härdle, W., & Mungo, J. (2008). Value at risk and expected short fall when there is long range dependence. Discussion Paper 2008-006, Humboldt-universität zu Berlin, Germany.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68, 165–176.

Jensen, M.C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2-3), 95-101.

Kasman, A., Kasman, S., & Torun, E. (2009). Dual long memory property in returns and volatility: Evidence from the CEE countries' stock markets, *Emerging Markets Review*, 10, 122–139.

Kwan, W., Li, W.K., & Li, G. (2011). On the estimation and diagnostic checking of the ARFIMA-HYGARCH Model. *Computational Statistics and Data Analysis*, forthcoming.

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root: How Sure are we that Economic Time Series have a Unit Root?. *Journal of Econometrics*, 54, 159-178.

Lillo, F., & Farmer, J.D. (2004). The Long Memory of the Efficient Market, *Studies in Nonlinear Dynamics & Econometrics*, 8, 3.

McMillan, D.G., & Kambouroudis, D. (2009). Are Risk Metrics forecasts good enough? Evidence from 31 stock markets, *International Review of Financial Analysis*, 18, 117–124.

Mizrach, B. (1995). A Simple Nonparametric Test for Independence. Working Paper 1995-23, Rutgers University, USA.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Their Applications*, 9, 134-137.

Phillips, P.C.B., & Perron, P. (1988). Testing for Unit Roots in Time Series Regression, *Biometrika*, 75, 335-346.

Ray, B.K., & Tsay, R.S., 1997. Bandwidth selection for kernel regression with long-range dependence. *Biometrika*, 84, 791–802.

Schmidt, P., & Phillips, P.C.B. (1992). LM Test for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics*, 54, 257–287.

Schwarz, G. (1978). Estimating the dimension of a Model. *Annals of Statistics*, 6, 461-464.

Tang, T.L., & Shieh, S.J. (2006). Long memory in stock index future markets: a value-at- risk approach. *Physica A*, 366, 437–448.

Watson, G.S. (1964). Smooth regression analysis. *Sankhyä*, A26, 359-372.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32, 1477–1484

Figure 1 – Dow Jones (logarithmic series and returns)

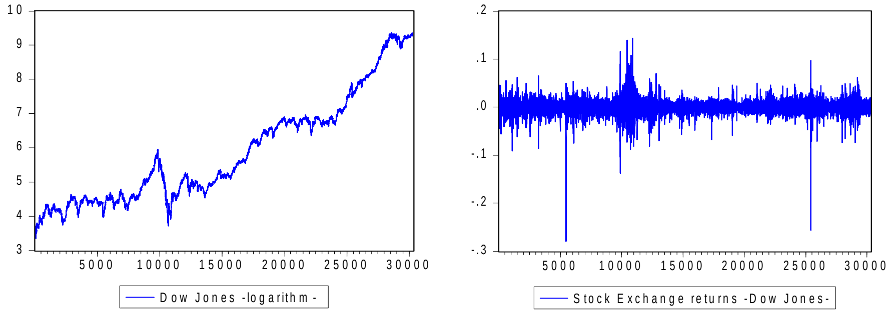

Figure 2 – Kernel estimation of density

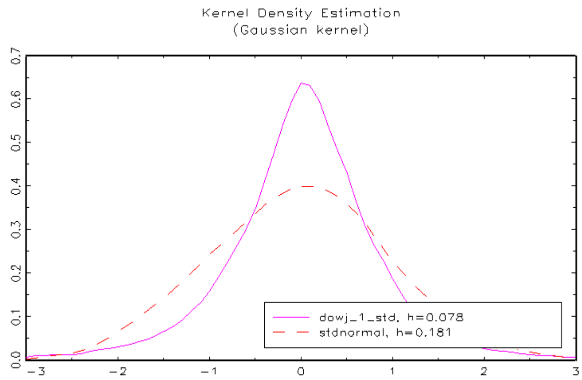

Figure 3 – Scatter plot of Dow Jones variations

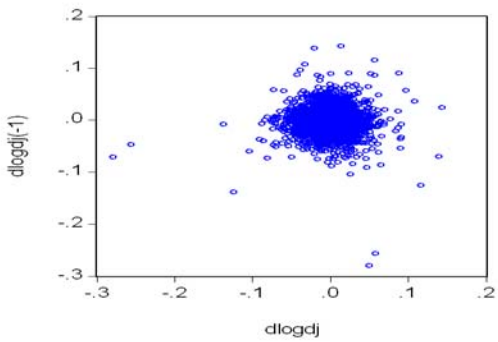

Figure 4 –Periodogram of Dow Jones returns

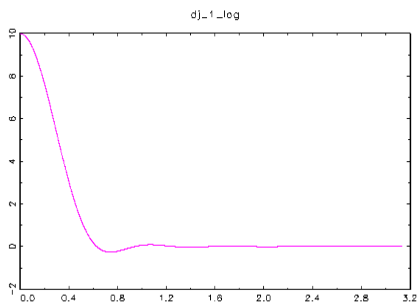

Figure 5 – Evolution of estimation residuals

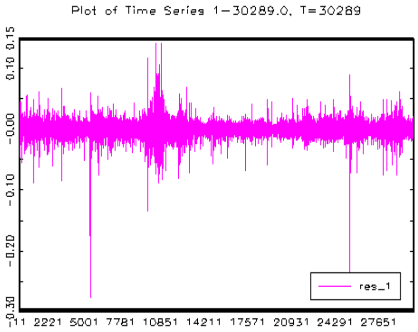

Figure 6 – Kernel estimator of residuals' density

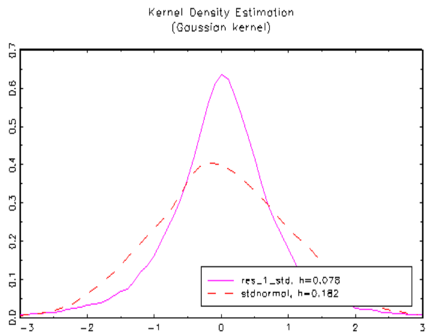

Figure 7 – Simple correlogram and periodogram of squared residuals

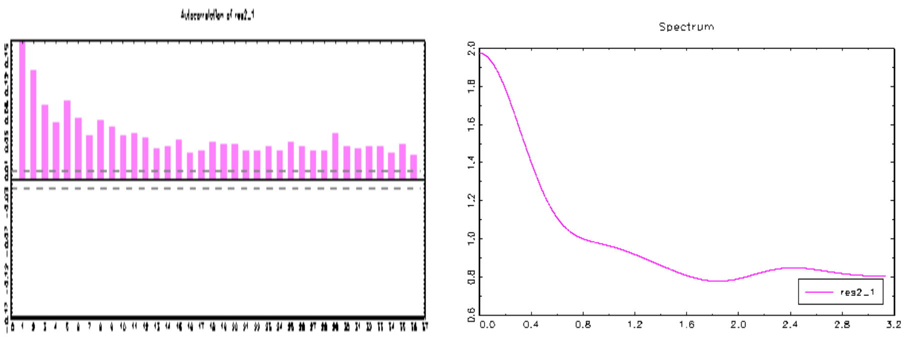

Table 1 – Unit root tests

Test	Logarithmic series		Returns	
Dickey-Fuller	(I)	-1.6316 (-3.4127)	(I)	-73.810 (-3.4127)
Phillips-Perron	(III)	-1.6323 (-3.4127)	(II)	-167.53 (-3.4127)
Schmidt-Phillips	*Z(Rho)*	-6.4786 (-18.1)	*Z(Rho)*	-27677.6873 (-18.1)
	Z(Tau)	-1.7970 (-3.02)	*Z(Tau)*	-159.0685 (-3.02)

Test	Window	Spectral estimation method							
		Bartlett kernel				Quadratic spectral kernel			
		Logarithmic series		Returns		Logarithmic series		Returns	
		(II)	(III)	(II)	(III)	(II)	(III)	(II)	(III)
KPSS H_0 : Stationary series	Newey-West	20.413 (0.463)	3.564 (0.146)	0.165 (0.463)	0.033 (0.146)	72.771 (0.463)	12.631 (0.146)	0.173 (0.463)	0.034 (0.146)
	Andrews	0.7086 (0.463)	0.150 (0.146)	0.177 (0.463)	0.035 (0.146)	2.29 (0.463)	0.159 (0.146)	0.181 (0.463)	0.036 (0.146)
Elliott-Rothenberg-Stock H_0 :unit root	Newey-West	0.00272 (3.26)	0.0066 (5.62)	138.388 (3.26)	16.367 (5.62)	0.0028 (3.26)	0.0069 (5.62)	144.822 (3.26)	17.151 (5.62)
	Andrews	0.0029 (3.26)	0.007 (5.62)	147.891 (3.26)	17.527 (5.62)	0.0029 (3.26)	0.0070 (5.62)	151.514 (3.26)	17.959 (5.62)

(I): model without constant and deterministic trend (5%)
(II): model with constant and without deterministic trend (5%)
(III): model with constant and deterministic trend (5%)

Table 2 – Normality tests on distribution of returns

Skewness	Kurtosis	J.B	A.D	K.S
-1.1208	38.2164	1571629	1015.4123	0.581 (0.092

J.B is the Jarque-Bera statistic,
K.S is the Kolmogorov-Smirnov statistic,
A.D is the Anderson-Darling statistic.
The Kolmogorov-Smirnov critical values are given between parentheses at 5%.

Table 3 – Homoscedasticity tests

Breusch-Pagan statistic	White statistic	LM - ARCH(2) statistic
1481.7629	1370.9883	1114.3201

Table 4 – BDS and Mizrach test results on the series of returns

m	BDS		Mizrach
	Fraction of pairs	Standard Deviation	$T \leq 1000$
2	41.85412	37.41965	4.3802
3	53.06906	48.96358	3.1270
4	60.44449	57.82082	2.1684
5	67.01797	67.16815	2.1346
6	73.79956	78.60053	1.1964
7	80.97002	92.77881	1.7861
8	88.54294	110.1803	1.8503
9	97.17348	132.5872	1.9537
10	106.8643	161.0307	1.5520
11	117.9473	198.9349	-
12	131.0207	250.3263	-
13	146.1977	318.0141	-
14	163.8683	408.8996	-
15	184.3967	530.3570	-
16	208.5077	694.4048	-
17	237.1340	922.2850	-
18	271.0007	1242.474	-
19	311.1570	1689.988	-
20	359.0700	2325.839	-

The BDS statistic is computed by two methods with $\varepsilon = 0.7$

Table 5 – Results from the ARFIMA estimation using spectral methods on Dow Jones returns

Windows	Ordinates				
	$n^{0.4}$	$n^{0.5}$	$n^{0.6}$	$n^{0.7}$	$n^{0.8}$
GPH	-0.0489 (-0.5431)	0.0329 (0.6527)	0.0411 (1.4095)	0.0305 (1.7834)	**0.0311** **(3.0528)**
Rectangular	-0.0501 (-0.5045)	0.0114 (0.2061)	0.0322 (1.0023)	0.0282 (1.4961)	**0.0293** **(2.6160)**
Bartlett	-1.1681 (-0.0670)	0.0132 (0.4104)	0.0304 (1.6382)	**0.0253** **(2.3253)**	**0.0271** **(4.1915)**
Daniell	-0.0694 (-0.9878)	0.0122 (0.3098)	0.0306 (1.3452)	0.0255 (1.9079)	**0.0271** **(3.4235)**
Tukey	-0.0646 (-1.0326)	0.0115 (0.3288)	0.0309 (1.5249)	**0.0260** **(2.1866)**	**0.0277** **(3.9173)**
Parzen	-0.0614 (-1.1900)	0.5466 (0.0158)	0.0307 (1.8373)	**0.0255** **(2.6063)**	**0.0274** **(4.7018)**
B-priest	-0.0606 (-0.7878)	0.0103 (0.2406)	0.0317 (1.2713)	0.0270 (1.8500)	**0.0285** **(3.2790)**
Andrews-Guggenberger	-	-	-	-	**0.0294** **(3.047)**

The values in parentheses are the Student statistics. The statistically significant values are in bold.

Table 6 – Semiparametric estimation by the exact maximum likelihood method

Exact maximum likelihood	Nonparametric estimation of the deterministic trend	
	Optimal CV Criteria	$\hat{h}_{opt}$
(2, 0.0423, 2) t_d*=12,5369*	0.9382	*0.1384*

CV : cross validation criteria and $\hat{h}_{opt}$ **:** optimal window

Tableau 7 – Main characteristics of residuals

Skewness	Kurtosis	J.B statistic	ARCH-LM statistic
-1.02	37.47	1505311.66	20.1646

Table 8 – Maximum likelihood estimation – BHHH algorithm

Parameters	SEMIFARMA-GARCH	SEMIFARMA-FIGARCH	SEMIFARMA-HYGARCH
$\hat{\phi}_1$	-0.1767 (-2.3134)	-0.2036 (-2.5513)	-
$\hat{\theta}_1$	0.2668 (3.7227)	0.5011 (0.7244)	0.1067 (12.3140)
$\hat{d}_2$	0.0326 (3.4389)	0.0231 (2.6932)	0.0226 (2.7058)
$\hat{\alpha}_0$	1.3564 (0.8644)	1.1025 (2.0876)	1.8451 (2.0039)
$\hat{\alpha}_1$	0.0895 (2.1494)	0.0612 (2.1511)	0.0567 (2.1765)
$\hat{\beta}_1$	0.0949 (4.1676)	0.1032 (3.9842)	0.1256 (2.4783)
$\hat{d}$	-	0.04311 (4.2745)	0.04420 (4.8655)
$\hat{\alpha}$	-	-	0.1592 (2.9252)
$\hat{h}_{opt}$	0.1384	0.1384	0.1384
IMSE	0.9382	0.9382	0.9382
Akaike	5.6842	5.6438	**5.6395**
JB statistic	3436.18	3407.22	**3391.46**
Schwarz	5.6903	5.6422	**5.6416**
ARCH(1)	2.8847	2.6512	**2.4956**

The Student statistics are in parentheses. IMSE: Minimum Integrated Mean Squared Error

Table 9 – BDS test on SEMIFARMA-HYGARCH residuals

m	BDS statistic
2	0.2531
3	0.2844
4	0.3142
5	0.3578
6	0.4812
7	0.6740
8	0.8124

Table 10 – Comparison of predictive qualities (in-sample predictions)

	Horizon	Criteria	SEMIFARMA-GARCH	SEMIFARMA-FIGARCH	SEMIFARMA-HYGARCH	Random Walk
Conditional mean (Returns)	1 day	MSE	**6.0644**	6.1632	6.1665	7.0328
		MAE	2.1432	**2.0323**	2.0342	3.2744
	2 days	MSE	**6.0648**	6.0728	6.0813	7.1214
		MAE	**2.1263**	2.1344	2.1397	3.3328
	15 days	MSE	**6.6231**	6.6420	6.6422	7.1488
		MAE	**2.6302**	2.6481	2.6486	3.3611
	30 days	MSE	**6.2003**	6.2078	6.2076	7.2201
		MAE	**2.8804**	2.8902	2.8869	3.4027
	90 days	MSE	7.0201	7.0194	**7.0072**	8.1746
		MAE	2.6327	2.6319	**2.6289**	3.4726
	180 days	MSE	6.9303	6.9211	**6.9072**	8.3011
		MAE	2.4411	2.4207	**2.3995**	3.5814
	260 days	MSE	6.7012	6.6981	**6.6837**	8.1233
		MAE	2.1354	2.1349	**2.1219**	3.4916
Conditional variance (Volatility)	1 day	MSE	**5.0333**	5.3682	5.3684	-
		MAE	2.1112	**1.0721**	1.0725	-
	2 days	MSE	5.0232	**5.0188**	5.1029	-
		MAE	2.1152	2.1084	**2.0891**	-
	15 days	MSE	5.5178	**5.4411**	5.4677	-
		MAE	**2.1317**	2.1466	2.1578	-
	30 days	MSE	**5.1002**	5.1287	5.1366	-
		MAE	2.7901	2.8023	**2.7811**	-
	90 days	MSE	6.0104	6.0044	**5.4607**	-
		MAE	1.9346	1.9222	**1.8344**	-
	180 days	MSE	5.8322	5.8188	**5.7702**	-
		MAE	1.3208	1.2024	**1.1109**	-
	260 days	MSE	5.2019	5.1298	**5.0913**	-
		MAE	2.1453	2.0933	**2.0532**	-

Table 11 – Comparison of predictive qualities (out-of-sample predictions)

	Horizon	Criteria	SEMIFARMA-GARCH	SEMIFARMA-FIGARCH	SEMIFARMA-HYGARCH	Random Walk
Conditional mean (Returns)	1 day	MSE	**6.1232**	6.1353	6.1422	7.0632
		MAE	2.1432	**2.0323**	2.0344	3.4811
	2 days	MSE	**6.1271**	6.1452	6.1488	7.1568
		MAE	**2.1824**	2.1956	2.1397	3.6232
	15 days	MSE	**6.6401**	6.6511	6.6823	7.2402
		MAE	**2.6902**	2.7105	2.7113	3.8600
	30 days	MSE	6.2602	6.2588	**6.2552**	7.6314
		MAE	2.9460	2.9410	**2.9400**	3.5602
	90 days	MSE	7.0911	7.0812	**7.0253**	8.2679
		MAE	2.7033	2.6963	**2.6289**	3.9601
	180 days	MSE	6.9219	6.9008	**6.8814**	8.7246
		MAE	2.4903	2.4855	**2.3836**	3.9901
Conditional variance (Volatility)	1 day	MSE	**5.0965**	5.0836	5.0844	-
		MAE	2.1708	**1.8814**	1.9203	-
	2 days	MSE	5.0811	**5.0993**	5.1036	-
		MAE	2.1758	2.1801	**2.1107**	-
	15 days	MSE	5.5701	**5.5233**	5.5420	-
		MAE	**2.1908**	2.2012	2.1897	-
	30 days	MSE	**5.1685**	5.1786	5.1383	-
		MAE	2.9067	2.8511	**2.8423**	-
	90 days	MSE	6.0210	5.6049	**5.5417**	-
		MAE	1.9458	1.8922	**1.9015**	-
	180 days	MSE	5.5589	5.4274	**5.4110**	-
		MAE	1.2840	1.1871	**1.1701**	-

Table 12 – Comparing predictive accuracy of SEMIFARMA-HYGARCH model over the random walk: Diebold-Mariano test

Forecast observations numbers	S_1	S_2	S_3	*Naïve Benchmark test*		*MGN*	*MR*
30	2.2856 (0.0223)	2.9212 (0.0035)	3.6303 (0.0003)	$F = 0.1400$ (1.0000)	$F = 7.1452$ (0.0000)	-7.2450 (0.0000)	-2.3343 (0.0211)

The P-Values are in parentheses, S_1 : Asymptotic test statistic, S_2 : Sign test statistic, S_3 : Wilkoxon test statistic, MGN: Morgan-granger-Newbold test statistic, MR: Meese-Rogoff test statistic. Truncation lag used for asymptotic test = 3, Truncation lag used for Meese-Rogoff test = 8.