May 18, 2015

Master Thesis in Finance

Stockholm School of Economics

RISING HOUSEHOLD CONSUMER DEBT: GOOD OR BAD?

EMPIRICAL RESEARCH ON U.S. STOCK MARKET VOLATILITY USING NORMAL MIXTURE GARCH-MIDAS MODEL

Hang Le[1]

ABSTRACT

Household debt has been on the continuous rise to raise concern for its sustainability and its consequences to the financial system and the macro-economy as a whole. In this paper, I review empirical work on the growth of total household consumer debt ratio on long-term component of stock market volatility. The paper applies GARCH-MIDAS model with a mean-reverting unit *weekly* GARCH process, and a MIDAS (Mixed Data Sampling) applying to *monthly* household consumer debt ratio and errors following normal distribution, student-t distribution and two-component normal mixture distribution. The results show that household consumer debt ratio explains more than 12% of total stock volatility for the full sample from 01-01-1964 to 01-01-2015. In addition, household consumer debt ratio has mixed effects on stock market volatility. Income inequality seems to be plausible to explain the mixed effect. In low income inequality period, household consumer debt decreases stock market volatility. In high income inequality period, household consumer debt increases stock market volatility. It shapes part of a renewed interest in whether or not rising income inequality is the source of financial instability.

Keywords: Household debt, income inequality, stock market volatility, mixed data sampling, long term variance component, normal mixture distribution.

SUPERVISORS: ASSOCIATE PROFESSOR IRINA ZVIADADZE

[1] 40517@student.hhs.se

ACKNOWLEDGEMENT

I would like to express a very special thank to my thesis supervisor, Assistant Professor Irina Zviadadze at Department of Finance of Stockholm School of Economics (SSE) for her continuous support, comments and suggestions. I would also like to thank Swedish Institute (SI) for great supports of my master study. Last but not least, I would like to thank my family and friends for encouragement and constructive feedback.

CONTENT

LIST OF FIGURES

Figure 8. (Rows) From left to right, Normal Mixture GARCH-MIDAS, Normal Distribution GARCH-MIDAS, Student-t distribution GARCH-MIDAS for period from 01-01-1964 to 01-12-1984 (dd-mm-yy).

(Columns) From up to down: QQ plot, Sample Autocorrelation Function of z, Sample Autocorrelation Function of squared z, conditional volatility and its long-run component, optimal weight function and errors distribution for period from 01-01-1964 to 01-12-1984 (dd-mm-yy).

Figure 9. (Rows) From left to right, Normal Mixture GARCH-MIDAS, Normal Distribution GARCH-MIDAS, Student-t distribution GARCH-MIDAS for period from 01-01-1985 to 01-01-2015 (dd-mm-yy).

(Columns) From up to down: QQ plot, Sample Autocorrelation Function of z, Sample Autocorrelation Function of squared z, conditional volatility and its long-run component, optimal weight function and errors distribution for period from 01-01-1985 to 01-01-2015 (dd-mm-yy).

LIST OF TABLES

1. INTRODUCTION

The rising household debt, both in absolute and relative term, has raised concern for financial stability. One important component of household debt is household consumer debt. The value of household consumer debt is larger than the size of the private equity market (Vissing-Jørgensen, 2007). Aggregate household consumer debt has increased from $0.36 trillion in 1980 to $3.4 trillion in 2014. During the same period, total household consumer debt per disposable personal income (Consumer Debt Ratio – CDR) has grown from 18.45% to 25.19% (Figure 1). Despite the fact that consumer debt is about 1/3 the size of mortgage debt and has fluctuated in a relatively narrow range than mortgage debt, the required payments on consumer debt are actually higher than those on mortgage debt because of the shorter maturities on consumer debt (Figure 2). The returns to understanding the implication of rising household consumer debt to financial markets, therefore, are high.

Overall, households are today much more leveraged than they used to be in the past. This fact may reflect the deepening of the financial market, as household access to credit was substantially increased. Household indebtedness can increase households' lifetime welfare by allowing households to maintain a stable level of consumption when there is a temporary loss of income. However, if households maintain the level of indebtedness regarding a permanent loss of income (during extending period of rising income inequality); household's indebtedness might pose a risk to the financial stability. It emphasizes the fact that excessive indebtedness may exhaust households' balance sheet and ultimately lead them to curtail their spending. Therefore, the level of household borrowing in period of rising income inequality implies significant exposure to the financial sector.

In spite of the increasing attention to household consumer debt, the effect of household consumer debt has not been a major focus of research economics. In response to the view of growth in household consumer debt as a negative force in the economy, in this paper, I attempt to examine the effect of household consumer debt to stock market volatility. In efficient markets, when stock price reflects all available information, stock market volatility will reflect the volatility of economic fundamentals and an inherent part of a well-functioning financial system.

Even relatively large short-term volatility can be the result of a rational reaction by market participants to rapidly changing events and increased uncertainty about future returns (Global Financial Stability Report, Sep 2003). The extreme stock volatility (often referred to as "tail events") reflects potential sources that are associated with financial instability. The effect of household debt on stock market volatility, therefore, may shed light on the implication of household consumer debt.

The objective of this paper is to empirically investigate the effect of household debt to stock market volatility by using the model that is able to take into consideration both high and low frequency components of stock volatility during different periods of income inequality. At this purpose, the GARCH-MIDAS model, proposed by Engle et al (2013) is used by combining the information provided by *weekly* stock prices and *monthly* household consumer debt ratio while applying two-component normal mixture errors distribution. The model is then compared to GARCH-MIDAS model with normal and student-t errors distribution. Among these models, GARCH-MIDAS model with two-component normal mixture allows us to capture the effect of household debt variables in different states of the economy, thus provides the best fit.

The paper is organized as follows. The literature review is presented in Section 2. Section 3 outlines the hypothesis. Section 4 explains the data and methodology. Section 5 discusses results and analysis. Lastly, Section 6 presents conclusions and suggests further research.

2. LITERATURE REVIEW

2.1. Why is household consumer debt rising?

To understand the effect of rising household consumer debt, it is important to understand what causes household consumer debt to rise. The understanding of its causes reveals information of whether it is sustainable, and consequently, reveals information about its effect to stock market volatility.

As explained by Rinaldi and Sanchis-Arellano (2006), household debt increases due to macroeconomic stability, financial product innovations and legal or institutional regulations, while market imperfections, together with the effect of moral hazard on the behavior of some lenders, may have boosted household debt to excessive levels, resulting in the growth of non-performing loans. If it is the case, the fact that household consumer debt is rising do not imply whether the economy is in macroeconomic stability or instability states. Excessive rising of household consumer debt implies the market instability, however, the level of excessive borrowing is difficult to determine.

Dynan (2009) and Edelberg (2006) explained this growth as due to technical change in loan production, including but not limited to reductions in distributional costs, risk-based pricing, monitoring and repossession, and securitization and other secondary market innovation. A related explanation is that the technology of persuasion has improved as well, in other words, lenders have gotten more effective at convincing consumers to borrow (Gabaix and Laibson 2006, Bertrand et al. 2010, Gine, Martinez, and Keenan 2014; Gurun, Matvos and Seru 2014). The growing in direct marketing in the post-war period, and especially in the IT era seems to be consistent with this view, in addition to innovation in pricing, with bank checking account overdrafts (Stango and Zinman 2014a), credit card introductory rates and penalty fees (Agarwal, Chomsisengphet, et al. 2014; S.DellaVigna and Malmendier 2004; Heidhues and Koszegi 2010). Financial deregulation, which has decreased the effect of credit rationing, as well as lower interest rates, in nominal and real terms, can also explain the rapid growth in household consumer indebtedness. The fact is proved by U.S. Consumer Survey Report from the beginning of the 1990s. From 1992 to 2007, a percentage of households' loan application that was partially, or fully, rejected in the previous five years decreases from 32% to 24%. During the same period, a percentage of failed applicants that succeed to obtain a subsequent loan (e.g. by applying to another lender) rises from 37% to 45%. Other possible factors include demographic shift (Dynan and Kohn 2007; Christelis, Ehrmann, and Georgarakos 2013) and reduced generosity in social insurance (Hacker 2008). Rising household consumer debt in response to those factors, however, reveals little information about its sustainability to the financial markets.

Another factor that is the main focus of this paper is income inequality. Many authors investigate the link between household borrowing and income inequality. Among those, Barry and Steven (2014) connect rising income inequality to increasing debt-to-income ratio by below equation:

$$d/_{dt}\left(D/_{Y}\right) = \mathrm{A}/_{Y} + C/_{Y} - 1 + (i - \pi)\left(D/_{Y}\right) - g_{Y}\left(D/_{Y}\right)$$

Where: A equals Asset Purchases minus Asset Sales (at book values),

C is consumption

Y is disposable income

D is the stock of debt

g_Y is the real growth rate of income

π is the inflation rate

i is interest rate

Rising income inequality contribute to the equation as g_Y for bottom group decreases (the income growth rate of the bottom group falls while the growth rate of the upper group rises). In this case, the income share of the top-income group will increase. If g_Y decreases, the debt-to-income ratio rises more quickly for the lower group, other things equal. As $\left(D/_{Y}\right)$ increases, the interest term in the equation becomes larger which implies the rise in financial instability. If real interest rate rises, the effect is even stronger. The adverse change in real interest rate $(i - \pi)$, even minor, could significantly increase debt service payments of bottom income households. Data from Mason and Jayadev (2014a, Table 2) infer the real interest rate is the most important factor explaining the rising household debt-income ratio in the early 1980s, as households borrow to fund their debt. Households, however, may draw down assets ($^{A}/_{Y}$ decreases) to stabilize the debt-income ratio. But if the drop in g_Y is permanent, the rate of asset accumulation would have to drop permanently to keep D/Y from rising. For a lower-income group, this response will well drive asset accumulation negative, which would be unsustainable. A more sustainable response, especially low-income households, would be adjust to lower income growth or higher real interest rates by reducing the ratio of outlays to income (recognizing that outlays consists of personal consumption expenditure (PCE) plus interest expense and personal transfer) ($^{C}/_{Y}$ decreases). Therefore, in period of rising inequality, rising household consumer debt, especially for low-income households would implies unsustainable borrowing, if households do not stabilize their debt-to-income ratio. Consequently, in the next section, I examine the responses of U.S households' indebtedness in period of rising income inequality. Whether households response to rising income inequality by reducing their outlays ratio or selling assets implies the sustainability of rising household consumer debt, thus its effect to stock market volatility.

2.2. Households indebtedness in rising income inequality period

As demonstrated in Section 2.1, in response to rising income inequality, household can stabilize the debt-income ratio by either drawing down assets (which is unsustainable) or reducing the expenditure-income ratio (which is more sustainable).

The period of rising income inequality marks around 1980s, when top 5% income share increase quickly after long period of stabilization from 1964 to 1980 (Figure 4). From 1964 to 1980, top 5% income share increased slightly from 20.62% to 21.17% while it accelerated to 33.84% in 2007. Barry and Steven (2014) estimate that between 1960 to 1980, the annualized income growth rate of households in the bottom 95% (1.9%) is slightly lower than the growth rate of households in the top 5% (2.1%). Income inequality, however, starts rising from 1980s. Annualized growth rate of real income per household for top 5% increased to 3.9% in 2007 while for the bottom 95%, it decreased to 1.1%.

To examine whether household reduce their expenditure-income ratio during rising income inequality period, I take into consideration the outlays ratio after 1980s. From 1989 to 2007, the bottom 95% consumes 10% more per disposable income than top 5% (Barry and

Steven 2014). Given the strong increasing trend of the consumption rate (PCE/DPI – Figure 3) during the same period, it implies that bottom 95% households did not reduce their consumption rate. The conclusion is consistent with results from Dirk and Fabrizio (2005) that income inequality in the U.S. has not been associated with consumption inequality. Meyer, Bruce and James (2013) also points out that while the 90/10 income share ratio was 19% higher from 2000 to 2011, the 90/10 ratio for consumption was slightly lower. Barry and Steven (2014) also show that top income households cut their consumption rate while bottom income household maintain or slightly increase their consumption rate during period of high income inequality. Furthermore, given that the outlay rate from the bottom 95% increases somewhat more than the consumption rate from 1989 (Barry and Steven 2014), we confirm that there is no decline in the outlay rate for bottom 95% income households either (Figure 3).

Overall, low-income households did not reduce their outlay rate in period of rising income inequality. It implies that the debt-income ratio for this group should have risen. The data from Survey of Consumer Finance confirm this point. The debt-income ratio for bottom 95% income rises from 77% to 177% from 1983 to 2007 (Survey of Consumer Finance 2007).

The other option of offsetting debt by selling assets is also not possible. Household rising net worth to disposable income for bottom 95% (as seen in Figure 5) suggests that rising household net worth may offset rising household consumer debt. However, Barry and Steven (2014) suggest analyzing the composition of net worth for the bottom 95%. The value of primary residence and retirement account should be excluded in the value of net worth. Rising equity in an owner-occupied home is offset by a rising opportunity cost of living in that house, unless the homeowner literally sells the house and moves into a less costly one. Similarly, the purpose of retirement account is to fund a future consumption plan, not to offset a rise in debt. Excluding those two accounts from net worth reveals the downward trend of household's net worth to disposable income. It implies that it is also not possible for households to offset debt by selling assets.

In summary, a fall in income growth implies that the bottom income groups should cut its consumption rate to maintain a sustainable debt-income ratio. However, facts show that bottom 95% did not appear to smooth their consumption. Instead they borrow at unsustainable rates, therefore, deteriorating their balance sheet.

The rising inequality implies that the bottom groups should reduce their consumption rate; however, rising inequality is likely to explain why the bottom groups did not cut their consumption. Some studies suggest that rising inequality leads to increased supply by increasing loanable funds (e.g., Kumhof, Ranciere and Winant 2014, Barry and Steven 2014). Others have suggested that rising inequality leads to increase increased loan demand through social preferences (reference points and/or peer effects) (Georgarakos et al 2014, Barry and Steven 2014). In relative income hypothesis of Duesenberry (1952) and developed recent work on "expenditure cascade" (Levine et. al., 2010; Belabed et al. 2013), households whose incomes are falling behind try to keep up with norms of spending set by those who benefit from rising inequality. Van Treeck and Sturn (2012) argue that the rising income inequality in the United States has led to a change in the consumption and borrowing behavior of American households. After having increased working hours, and having easy access to credit, for the purposes of both consumption and housing, middle-income Americans have reacted to the growing gap between their revenues and those of their better-to-do neighbors by increasing the extent of their borrowing and thus reducing their saving rates. Marglin (1984) proposed the "habit persistence theory" that prevents household from deviating from recent consumption patterns.

In conclusion, in period of rising income inequality, households' behavior response in the U.S. implies unsustainable rising household consumer debt. Rising income inequality also explains the households' behavior responses for not stabilizing their debt-to-income ratio. Rising household debt, therefore, is expected to indicate important effect to stock market volatility in different level of income inequality. In the next section, the hypothesis is formulated based on the implication of literature review.

3. HYPOTHESIS

As shown in Section 2, the link of rising income inequality to rising household consumer debt implies important implication to stock volatility. Rising income inequality implied that the bottom income household would have to cut its consumption rate to keep a sustainable debt-income ratio. The bottom income households in the U.S., however, maintain their consumption rate partly due to income inequality.

Therefore, rising income inequality amplifies household consumer debt in the U.S to excessive level, which is unsustainable. This is likely to increase uncertainty in the stock market, which in turn increase stock market volatility. Therefore, I propose the following as the main hypothesis of my study:

H1: Rising household consumer debt affects stock market volatility.

H2: Rising household consumer debt has mixed effects to stock volatility.

H2a: In low income inequality period, rising household consumer debt decreases stock market volatility in the U.S.

H2b: In high income inequality period, rising household consumer debt increases stock market volatility in the U.S.

The hypothesis will be test using the data and methodology explained in the next section.

4. DATA AND METHODOLOGY

4.1. Descriptive statistics of U.S. stock market return

From Kenneth R. French's website, I obtained weekly[2] U.S. stock market returns from 1964 to 2015 (2666 data points). The stock market return includes all NYSE, AMEX and NASDAQ firms. This sample is used to analyze the effect of household consumer debt ratio available monthly from 01-01-1959 to 01-01-2015. There is difference in the time period of U.S stock market return and household consumer debt ratio since I also need five year of lags of household consumer debt to compute the long-term component of stock market volatility (as detailed in Section 4.3).

The first left plot of Figure 6 shows the weekly stock return sample. The weekly return series seem to be a stationery process with a mean close to zero but with volatility exhibiting relatively calm periods followed by more volatile periods. The first middle plot in Figure 6 shows the Sample Autocorrelation Function for the weekly returns of lags 0 to 20. Based on the Sample Autocorrelation Function plot, it is not confirmed whether the data is serially correlated or not, even though it seems to have significant serial correlation at lag 6 and 15.

However, Ljung-Box Q-test (see Appendix A for the explanation of the Ljung-Box Q-test) rejects the null hypothesis that all autocorrelations up to the tested lags are zero for lags from lags 6 to 20. This suggests that to model this return series, the conditional mean is needed. Although the paper focuses on the conditional variance (the effect of household variable on long term component in particular), for the conditional variance model to perform properly, the conditional mean needs to be taken into account as well.

The next descriptive statistics is that of the squared returns. The second left plot of Figure 6 shows the weekly squared returns for the full sample. The second middle plot of Figure 6 shows the Sample Partial Autocorrelation Function which clearly demonstrated significant autocorrelation. Engle's ARCH test confirms that the square returns are serially correlated by

[2] I used several alternative time frequencies for stock market returns, for example daily data and weekly data. My results show that the effect of household consumer debt is the most significant for weekly data. It can be explained that daily stock market return may incorporate even non-economic fundamentals. I therefore elect to present only the results with weekly stock market return. The results for other time frequencies are available on request.

rejecting the null hypothesis that there is no autocorrelation for lags from 1 to 20 and even at a 5% significant level (see Appendix B for the description of Engle's ARCH test)

Next the empirical distribution of the weekly stock returns is examined. In Figure 6 (the higher right plot), q-q plot of the empirical distribution (y-axis) against the best fitted normal distribution (x-axis) is presented. The q-q plot shows clearly that even the best fitted normal distribution is not considered as good reference distribution. The empirical distribution of the weekly returns exhibits significantly heavier tails than the normal distribution which implies that another choice of parametric family should be considered. From the right plot of Figure 1, it is evident that t-location scale distribution is a much better reference. In case of the q-q plot against t-location distribution, there are only limited numbers of points that are off the linear red dot-line. Therefore, in the paper, student-t distribution is considered in response to the need of heavier tail distribution.

In Table below, summary statistics of U.S. stock returns is presented. The kurtosis is larger than 3 confirms the need of distribution with heavier tails than the normal distribution.

Mean	Standard Error	Standard Deviation	Sample Variance	Kurtosis	Skewness	Minimum	Maximum
0.0018436	0.000427	0.022068	0.000487	6.316541	-0.69183	-0.19821	0.12739

4.2. Data source of household consumer debt

When looking at aggregate numbers, there are two measures for household consumer debt. The first is how much it costs to service the debt as a fraction of disposable (after tax) income (as shown in Figure 1). The second is how much debt there is with respect to the same disposable income measure. Where these numbers are high is difficult to say, but in the aggregate, both measures have clearly decreased during the past crisis, but different scale. While debt service payments decreased by almost one-third, the debt ratio decreased by only one-fifth. Since the service debt payment also takes into account the effect of interest rate, whenever interest rates go back up, service payment will increase. The aggregate debt per disposable personal income, on the other hand, only takes into account the level of household consumer debt. In the scope of this paper, I focus on the aggregate household consumer debt per disposable personal income due to three main reasons. The first reason is that I want to focus on the level of household consumer debt, without taking into account the effect of interest rate. Long period of lower interest rate implies lower household consumer debt service payment and may draw a wrong picture about households' financial wealth. The second reason is the availability of data. Aggregate household consumer debt is available monthly for long period from 1959 to 2015; therefore, covers the period with change in income inequality. The third reason is analyzed in Section 2. As the level of $D/_{Y}$ increases, the effect of real interest is more profound; therefore, higher aggregate household consumer debt ratio captures households' vulnerability to adverse shock.

The aggregate nominal household consumer debt and nominal disposable personal income is obtained from U.S. Bureau of Economic Analysis, retrieved from FRED, Federal Reserve Bank of St. Louis. I construct the aggregate consumer debt per disposable income by dividing aggregate monthly consumer debt by total monthly disposable personal income.

4.3. Methodology

For studies investigating the effect of household consumer debt to stock market volatility, a comprehensive analysis is limited because of the data availability as well as data frequency mismatch between high-frequency stock price (in this paper - weekly) and low-frequency household consumer debt (monthly). It apparently generates a trade-off between the possibility of efficiently exploiting all the information provided by the high-frequency data and the necessity to investigate the linkages to the low-frequency determinants. The practice to aggregate the high-frequency to low-frequency data by summing, averaging or using high frequency data at one point in time (for example at the end of month) as a proxy for low frequency data can lead to information loss. One way to overcome this problem was to apply Spline-GARCH model proposed by Engle and Rangel (2008). Unlike conventional GARCH or stochastic volatility models, the Spline-GARCH permits unconditional volatility to change over time. In the first step of Spline-GARCH model, they extract the slowly varying component of volatility together with high frequency data and construct volatility measure by taking the sample average of the low-frequency component. In the next step, they estimate a reduced-form mode to link the estimated price volatility to the low-frequency determinants. The method was able to achieve a better fit than the simple GARCH model; however, according to Ghysels and Wang (2011), the unconditional variance is modelled in a deterministic and non-parametric manner, preventing the possibility of directly incorporate low-frequency data. Furthermore, averaging daily/monthly data at monthly/annual level still leads to information loss. Finally, the impact of lags of the low-frequency drivers on price volatility is neglected, which is not likely the case in household debt. Consistently increasing household debt level has a high possibility of impacting stock market volatility.

Therefore, in this paper, I applied a recent class of component GARCH model which applied MIDAS (Mixed Data Sampling). MIDAS regression models are proposed by Ghysels et al (2006) and are used to incorporate low-frequency macroeconomic variables into the high-frequency financial series. This recent component GARCH model is referred to as GARCH-MIDAS, where macroeconomic variables are incorporated directly into long-term component of stock market volatility.

The new class of MIDAS structure has attracted much attention in recent years. Chen and Ghysels (2009) use MIDAS setting to analyze the effect of news on forecasting volatility. Ghysels et al. (2009) analyzes the Granger causality with mixed frequency data. Kotze (2007) uses MIDAS regression to investigate high-frequency asset prices and low-frequency inflation. In addition, a number of papers use MIDAS regression to get low frequency forecasts with high frequency data. Bai et al. (2009) and Tay (2007) use monthly data for quarterly forecasts. Alper et al. (2008) uses MIDAS regression to cross emerging market comparisons of stock market volatility. Forsberg and Ghysels (2006) demonstrated the relationship outperformances of MIDAS over the HAR-RV (Heterogeneous Autoregressive Realized Volatility) model proposed by Anderson (2007). Following these studies, in this paper, I apply GARCH-MIDAS approach to study the effect of low frequency household consumer debt (sampled monthly) to high frequency stock market volatility (sampled weekly).

The GARCH-MIDAS model can be described as below. The log returns are written as:

$$r_{i,t} = \mu + \sqrt{\tau_t . g_{i,t}}\varepsilon_{i,t}, \quad \forall i = 1, \dots, N_t \quad (1)$$

Where $r_{i,t}$ is the return on weekday i during month t.

N_t is the number of weekdays (trading days) in month t.

τ_t is the long-term/secular component.

$g_{i,t}$ is the short-term component of stock volatility.

The conditional short term variance follows GARCH (1,1) process:

$$g_{i,t} = (1-\alpha-\beta) + \alpha \frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta g_{i-1,t} \quad (2)$$

The household variables which are observed monthly have effect on long-term component. Household debt ratio is directly incorporated in the long-term component as:

$$\tau_t = \ln(m + \theta_l \sum_{k=1}^{K} \varphi_k(w_1, w_2) X_{t-k}^{l}\ ) \ (3)$$

Where K is the number of periods over which I smooth volatility[3]. τ_t shall be the same during month t and X_{t-k}^{l} represents the level of household debt ratio. By levels, I mean the growth rate of aggregate household consumer debt ratio. Variance during month t will be $\sigma^2{}_{i,t} = g_{i,t} * \tau_t$

The weighting function used in equation (3) is described by beta lag polynomial (discussed further in Ghysels, Sinko and Valkanov (2006)). The beta lag is popular for use to accommodate various lag structures. It can accommodate monotonically increasing, decreasing weighting function or hump-shaped weighting function although limited to unimodal shapes.

$$\varphi_k(w) = \frac{(1-\frac{k}{K})^{w-1}}{\sum_{j=1}^{K-1}(1-\frac{j}{K})^{w-1}} \qquad (4)$$

Equation (1)-(4) form the standard GARCH MIDAS model for time-varying conditional variance and parameter space. The parameters are estimated by using maximum likelihood estimation[4]. The log-likelihood function takes a specific form depending on the distribution assumption of $\varepsilon_{i,t}$. The likelihood function involves a large number of parameters, which conventional optimization algorithms do not always find global maximization. Therefore, I use

[3] A potential problem is to ensure non-negative τ_t. A possible solution is to use the log form (log τ_t) specification.

[4] Maximum likelihood function can be easily transformed into minimum likelihood function by taking the negative of maximum likelihood function. Taking the minimum likelihood function is for the use of optimization toolbox in Matlab.

the simulated annealing approach (Goffe et al., 1994) for estimation, which very robust even for very complicated maximization. In case of two-component normal mixture distribution, expectation maximization (EM) algorithms are used, which will be explained in more details in section 4.3.1.3.

4.3.1. Errors distribution

4.3.1.1. Normal distribution

If $\varepsilon_{i,t}$ is assumed to follow standard normal distribution $\varepsilon_{i,t} \sim N(0,1)$, the log-likelihood value is written as:

$$\frac{T}{2} * \ln(2\pi) + \frac{1}{2} * \sum_{t=1}^{T} \ln\left({\sigma^2}_{i,t}\right) + \frac{1}{2} * \sum_{t=1}^{T} \frac{(r_{i,t} - \mu)^2}{{\sigma^2}_t}$$

Where T is the number of stock return observation

In case of normal distribution errors, parameter space will be { μ, α, β, m, θ_l, w}

4.3.1.2. Student t-distribution

If $\varepsilon_{i,t}$ is assumed to follow student t-distribution, the log-likelihood value is written as:

$$T * \ln\left(\Gamma\left(\frac{v+1}{2}\right)\right) + T * ln\left(\Gamma\left(\frac{v}{2}\right)\right) + \frac{1}{2} * T * ln\big(\pi * (v-2)\big) + \frac{1}{2} * \sum_{t=1}^{T} \ln\left({\sigma^2}_{i,t}\right) + \frac{(\mathrm{v}+1)}{2} * \sum_{t=1}^{T} ln\left(1 + \frac{(r_{i,t} - \mu)^2}{{\sigma^2}_t * (v-2)}\right)$$

Where v: the number of degrees of freedom

Γ: the gamma function

T: the number of return observation.

In case of student t-distribution errors, parameter space will be { μ, α, β, m, θ_l, w,v}

4.3.1.3. Two-component normal mixture distribution.

Mixtures of normal distributions have long history in empirical finance for modelling assets returns (see Press 1967, Praetz 1972, Clark 1973, Blattberg and Gonedes 1974, Kon 1984). The use of normal mixtures to handle fat tails was first considered by Newcomb (1963). A good introduction of mixture distributions (theory and applications) can be found in Everitt and Hand (1981), Titterington et al. (1985) , McLachlan and Basford (1988), Lindsay (1995), McLachlan and Peel (2000), and Frühwirth-Schnatter (2006). Gridgeman (1970) proves that a mixture of normal distribution is leptokurtic, when all regimes have the same mean. Mixtures of normal distribution, therefore, are able to accommodate various shapes of continuous distribution and able to capture leptokurtic and skewed characteristics of financial times series data.

The paper, therefore, examines the GARCH-MIDAS model with $\varepsilon_{i,t}$ is assumed to follow a two-component normal mixture distribution. A mixture of normal distribution with the same means has heavier tails than the normal distribution of the same variance. Also note that the return distributions with thicker tales have a thinner and higher peak in the center compared to normal distribution. Fama (1965) claims that "the most popular approach to explain long-tailed distributions of price changes is a mixture of several normal distributions with possibly the same mean, but substantially different variances". Therefore, the results can be applied to the stock market returns with heavy-tailed as demonstrated in the descriptive statistics of excess kurtosis (Section 4.1). Other reason for using two-component normal mixture distribution is that the household variables may have different effect depending on the state of the economy. The first component usually carries a higher weight (more than 90% or more in the mixture). The second component can be considered the 'crash' component because it has much higher long-term volatility and occurs less than 10% of the time. By using GARCH-Midas model with errors follow two-component normal mixture distribution, I can measure the impact of household debt to market volatility during different states of the economy. In summary, the GARCH-MIDAS model with two-component normal mixture distribution can be explained as below.

I assume the error term follows $\varepsilon_{i,t}|I_{t-1} \sim NM(p, 1-p; 0_1, 0_2, \sigma_{1t}^2, \sigma_{2t}^2)$

Variance during month t of normal distribution 1 is ${\sigma^2}_{i,1t} = g_{i,1t} * \tau_{1t}$

Variance during month t of normal distribution 2 is ${\sigma^2}_{i,2t} = g_{i,2t} * \tau_{2t}$

τ_{1t}, τ_{2t} is the long-term/secular component of volatility and $g_{i,1t}$, $g_{i,2t}$ is the short-term component of volatility of normal distribution 1, 2 respectively. The conditional short-term variance follows GARCH (1,1) process:

$$g_{i,1t} = (1 - \alpha_1 - \beta_1) + \alpha_1 \frac{(r_{i-1,1t} - \mu)^2}{\tau_{1t}} + \beta g_{i-1,1t} \quad (2)$$

$$g_{i,2t} = (1 - \alpha_2 - \beta_2) + \alpha_2 \frac{(r_{i-1,2t} - \mu)^2}{\tau_{2t}} + \beta g_{i-1,2t} \quad (3)$$

And the long-term return variance is defined as:

$$\tau_{1t} = m_1 + \theta_{1l} \sum_{k=1}^{K} \varphi_k(w_1) X_{t-k}^l \quad (4)$$

$$\tau_{2t} = m_2 + \theta_{2l} \sum_{k=1}^{K} \varphi_k(w_2) X_{t-k}^l \quad (5)$$

Where τ_{1t}, τ_{2t} will be the same during weekday $\forall i = 1, \dots, N_t$ of month t.

X_{t-k}^l represents the level of household consumer debt ratio.

The parameters are estimated by maximum likelihood method. Since the conditional density of normal mixture equal $\sum_{m=1}^{2} p_m \varphi_i(e_i)$ (with $\varphi_i(e_i)$ is the normal density functions of component normal distributions), the log likelihood function is:

$$\sum_{t=1}^{T} \ln(p * \left(\frac{1}{\sigma_{i,1t} * \sqrt{2\pi}} * e^{\frac{-(r_{i,t}-\mu)^2}{2*\sigma^2{}_{i,1t}}}\right) + (1-p) * \left(\frac{1}{\sigma_{i,2t} * \sqrt{2\pi}} * e^{\frac{-(r_{i,t}-\mu)^2}{2*\sigma^2{}_{i,2t}}}\right)) \ (6)$$

Direct maximization of equation (6) is quite difficult, numerically, since it involves finding the maximization of the sum inside the logarithm. Therefore, I use a procedure called the Expectation Maximization (EM) algorithm. The key characteristics of the EM algorithm have been established by Dempster et al. (1977). The EM algorithm is a popular tool for simplifying maximum likelihood problems in the context of a mixture model. The EM algorithm has become the method of choice for estimating the parameters of a mixture model, since its formulation leads to straightforward estimators (Piscard 2007). The EM algorithm for two-component Gaussian mixture is summarized as below.

I consider unobserved latent variables Δ_t taking the values 0 or 1: if $\Delta_t = 1$ then the observation comes from distribution 1, otherwise it comes from distribution 2. Suppose I knew the values of the $\Delta_t's$. Then the log likelihood function would be:

$$\sum_{t=1}^{T}[\Delta_t * \left(\frac{1}{2}\ln(2\pi) + \frac{1}{2} * \ln\left({\sigma^2}_{i,1t}\right) + \frac{1}{2} * \frac{(r_{i,t}-\mu)^2}{{\sigma^2}_{i,1t}}\right) + (1-\Delta_t) * \left(\frac{1}{2}\ln(2\pi) + \frac{1}{2} * \ln\left({\sigma^2}_{i,2t}\right) + \frac{1}{2} * \frac{(r_{i,t}-\mu)^2}{{\sigma^2}_{i,2t}}\right)\right] + \sum_{t=1}^{N}[(1-\Delta_t) * ln\pi + \Delta_t * \ln(1-\pi)] \quad (7)$$

Since the values of the $\Delta_t's$ are actually known, I take iterative steps, substituting for each Δ_t its expected value $\gamma_i(\theta) = E(\Delta_t|\theta, z) = \Pr(\Delta_t = 1|\theta, z)$. $\gamma_i(\theta)$ also called the responsibility of model 1 for observation i. The summarized steps for EM algorithm are as below:

- Take initial guess for the parameters
- Expectation Step: Computes the responsibilities:

$$\hat{\gamma_i} = \frac{\hat{\pi}\frac{1}{\sigma_{i,1t} * \sqrt{2\pi}} * e^{\frac{-(r_{i,t}-\mu)^2}{2*{\sigma^2}_{i,1t}}}}{\hat{\pi}\frac{1}{\sigma_{i,1t} * \sqrt{2\pi}} * e^{\frac{-(r_{i,t}-\mu)^2}{2*{\sigma^2}_{i,1t}}} + (1-\hat{\pi})\frac{1}{\sigma_{i,2t} * \sqrt{2\pi}} * e^{\frac{-(r_{i,t}-\mu)^2}{2*{\sigma^2}_{i,2t}}}}$$

- Maximization Step: computed the new estimated mean and variances using the maximum likelihood function (7)
- Iterated these steps until convergence with tolerate difference of 10^{-6}.

The weighting scheme used in equation (4) and (5) is described by beta lag polynomial.

$$\varphi_k(w_1) = \frac{(1-\frac{k}{K})^{w_1-1}}{\sum_{j=1}^{K-1}(1-\frac{j}{K})^{w_1-1}} \qquad (7)$$

$$\varphi_k(w_2) = \frac{(1-\frac{k}{K})^{w_2-1}}{\sum_{j=1}^{K-1}(1-\frac{j}{K})^{w_2-1}} \qquad (8)$$

Equation (1)-(8) form the GARCH MIDAS model for conditional variance and parameter space { μ,α_1,β_1, α_2,β_2, m_1,m_2, θ_{1l}, θ_{2l}, w_1, w_2, p}

4.3.2. Evaluation of in-sample fit

4.3.2.1. Likelihood-Based Criteria and Tests

The Bayesian information criterion (BIC) is defined as

$$BIC = -2logL(\theta) + Klog(N),$$

Where $logL(\theta)$ is the maximized log likelihood function.

K: number of parameters.

N: the number of data points in the in-sample period.

The model that yields the lowest value of BIC is considered to be the best fit. However, the reliability of likelihood-based tests and criteria depends on the accuracy of error distribution assumption. Hence, BIC is used accompanying with a test for the assumed distribution of errors.

4.3.2.2. Unconditional Distribution Tests

These tests are based on a comparison of the empirical errors distribution with a simulated errors distribution generated by the estimated model. I used Kolmogorov-Smirnoff (KS) statistics for the comparison of two distributions. The statistics is:

$$KS = \max |F_1(x) - F_2(x)|$$

where $F_1(x)$: Model-based errors distribution (using kernel fitting)

$F_2(x)$: Empirical errors distribution (using kernel fitting)

KS is the maximum of the vertical differences between two cumulative distribution functions. The model that minimizes the value of this statistic is the preferred choice, since it indicates smaller differences between model-based errors distribution and empirical errors distribution. In this paper, I present the asymptotic p-value of the test. P-value is larger than 0.01 indicates that the KS test does not reject the null-hypothesis that the data are from the same continuous distribution at 1% significant level. Therefore, the model with higher p-value is the preferred choice.

4.3.2.3. Measure of proximity

This paper uses two loss functions, the Mean Square Error (MSE) and the Mean Absolute Error (MAE), defined as:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(\sigma_t^2 - E(\sigma_t^2)\right)^2$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\sigma_t^2 - E(\sigma_t^2)|$$

MSE is a quadratic loss function and gives a larger weight to large prediction errors compared to the MAE measure, and is therefore encouraged to use when large errors are more serious than small errors (Brooks and Persand, 2003). The model with lower MSE and/or MAE is preferred choice.

4.3.2.4. Moment Specification Tests

If a model is able to capture all the time variation in volatility over a sample then the time series of errors should be independent and has constant volatility. The errors will have the same functional form of distribution as specified in the model. Denote this distribution by F. We take the value of cumulative distribution $u_t = F(\varepsilon_{i,t})$. Under the null hypothesis, u_t will be independently and uniformly distributed, then z_t which is the invert cumulative distribution of u_t is independent and standard normal i.i.d. Ljung-Box Q-test is used to assess the presence of autocorrelation at individual lags.

In this paper, I present the p-value of the test. P-value is larger than 0.01 indicates that the Ljung-Box Q-test does not reject the null-hypothesis that the data are from the same continuous distribution at 1% significant level. Therefore, the model with higher p-value for Ljung-Box Q-test is the preferred choice.

4.3.3. Structural break testing

One possible limitation of the estimation provided with the full sample is that the low-frequency component of stock volatility can present structural breaks. As noted by Engel et al. (2008), GARCH-MIDAS models are not immune to breaks. Therefore, we would prefer to test whether our full sample presents or not and re-estimate the models considering possible changes in the stock volatility dynamics. One way to test for structural breaks is to compare the log-likelihood function for the full sample with those of the sub-samples suggested by the empirical evidence. In this case, testing the presence of a break in coincidence with the change in income inequality can shed light on the effect of household consumer debt ratio associated with income inequality. Following Engel et al (2013), I test the hypothesis calculating:

$$-2 * \left[LLF_{full} - \sum_{i=sub.samples} LLF_i \right] \sim x^2(df)$$

Where df indicates the number of parameters times the number of restrictions, which corresponds to the number of subsamples minus one. Two subsamples representing two different periods of income inequality (1964-1984 and 1985-2015) are used.

5. RESULT AND ANALYSIS

5.1. Model estimation

In Figure 7, I present the QQ-plot, Sample Autocorrelation Function for the z_t and squared z_t (as explained in ***4.3.2.4***), conditional volatility and its long-run component of stock market returns, optimal weight function and distribution of errors of the full sample (starting from 01-01-1964 to 01-01-2015) for GARCH-MIDAS with normal mixture, normal and student-t errors distribution respectively. Based on the Sample Autocorrelation Function plot, it is not completely clear whether the data is serially correlated or not. The QQ plot of sample versus normal distribution, however, clearly shows that normal mixture GARCH-MIDAS model provides the best fit. The errors distribution of normal mixture GARCH-MIDAS model shows that mixture distributions capture distributions with higher peak and heavier tails than normal distribution. The first normal distribution (first component - e1) is a low peak high variance regime and the second normal distribution (second component - e2) is a high peak low variance regime.

I confirm the fitness of different models based on tests summarized in Section 4. In Table 01, I present the test results of GARCH-MIDAS model with normal distribution (with compare to GARCH(1,1)), student-t distribution and two-component mixture distribution for the full sample. When the GARCH-MIDAS with normal mixture distribution errors (GARCH-MIDAS-NM) is chosen as the benchmark model, it clearly outperforms alternative models except for BIC. However, the reliability of BIC depends on the accuracy of error distribution assumption which clearly favors normal mixture GARCH-MIDAS. In details, compare to other models presented, GARCH-MIDAS with normal mixture distribution errors has the highest value of log likelihood function (LLF=6,752.98), lowest value of MSE and MAE (MSE=1, MAE=0.77), highest value of Kolmogorov-Smirnoff statistics (KS stat = 0.09), highest value of Ljung-Box Q-test (LBQ=0.08). Therefore, I analyze the result in the next section based on normal mixture GARCH-MIDAS model.

5.2. Result

First, it is shown that controlling for household variables in GARCH-MIDAS model increases the goodness of fit, compare to the standard GARCH (1,1). Therefore, household consumer debt is important explanatory variables to the long term component of stock market volatility. To understand how much of expected volatility can be explained by household consumer debt ratio, I compute the ratio: $Var(\ln(\tau_t))/Var(\ln(\tau_t * g_t))$. The variance ratio result is 12.41%, which means that household consumer debt ratio contributes to/explains more than 12% of the total stock volatility. Therefore, the result confirms the first hypothesis (H1), indicating that rising household consumer debt affects stock market volatility in the U.S.

The parameter estimates appear in Table 01 for different GARCH-MIDAS models. In the full sample, I take 5 year of lags, or 60 lags[5]. The parameter estimate is negative for the first component and positive for the second component. The parameters are statistically significant at 10% level and it implies the mix effect of the household consumer debt ratio in the full sample. For 89.42% of the time (p=0.8942), increasing household consumer debt ratio leads to decreasing stock market volatility. For 10.57% of the time, increasing household consumer debt ratio leads to increasing stock market volatility. Therefore, the result confirms the second hypothesis (H2), indicating the mixed effect of household consumer debt to stock market volatility in the U.S. Figure 7 illustrates the plot of the optimal weighting function for both components in the full sample. The function results in seemingly counterintuitive weighting patterns, a lower weight for more recent observations. The optimal weighting function implies that the information associated with rising household consumer debt needs time lags to be incorporated into stock market volatility. This can be interpreted as a delay between economic information and its effect to stock market volatility. It can also show the inefficiency of stock market, when all available information is not immediately incorporated in stock market volatility.

For testing of structural break, I have 12 parameters and two sub-samples (1964-1984, 1985-2015), that is 12 degree of freedom. The sub-sample from 1964 to 1984 represents

[5] I used several alternative lags (K in methodology equations). The total lags are determined by number of years, or so-called MIDAS years, and by time-span t (monthly) that will be used to calculate τ_t. The results show that the optimal value of the likelihood function increases with the number of lags with the optimal level at around 60 lags. I therefore limit the number of lags in the MIDAS equation to 60 lags, which results in 5 MIDAS years.

relatively low income inequality while the sub-sample from 1985 to 2015 marks the increase in income inequality.

Parameter estimates of different models for different sub-samples appear in Table 01. The focus is on the parameter θ which reveals the impact of household consumer debt ratio on stock market volatility. For the first period (1964-1984), the parameter estimate is negative and statistically significant for both components, especially in the first component, the parameter estimate is strongly significant at less than 0.0001% significant level. This means that higher household consumer debt ratio leads to lower stock market volatility. For the first component (which carries the weight of 91.30% (p=0.913)), the parameter estimate is -4.067 with a t-statistics of 4.93. Since the weighting function with w=0.6220 puts 0.0517 on the first lag and 0.03979 on the second lag, I find that one percent increase of household consumer debt ratio at the current month would decrease the next month market volatility by $e^{-4.067*0.0517} - 1 \approx 0.1897$ or 18.97%. If last month's household consumer debt ratio increases by 1%, I would see the decrease of 14.94% in the stock market volatility next month. For the second component (which carries the weight of 8.7% (p=0.087)), the parameter estimate is -1.97 with a t-statistics of 2.2. Since the weighting function with w=6.32 puts higher weight on the last lags, indicating that the rising of aggregate household consumer debt ratio in the last 5 years has significant effect on stock market volatility. The result implies significant negative effect of rising household consumer debt to stock market volatility in the period of low income-inequality, which confirms the sub-hypothesis (H2a). Figure 8 illustrates the plot of the optimal weighting function for both components in the full sample. The function results in seemingly intuitive weighting patterns, a higher weight for more recent observations for the first component, however, counterintuitive weighting patterns with a lower weight for more recent observations for the second component. Higher weight for more recent observations may show the efficiency in the market during period of 1964-1985, as more distant observation information may already be captured in stock market volatility. The weighting function, however, still confirms that information regarding household consumer debt needs time lags to be captured in stock market volatility.

On the contrary, the second period (1985-2015) reflects differences effect compared to the 1964-1984 period. The parameter estimate is positive and statistically significant at 1% and

5% level. For the first components (which carries the weight of 91.01% (p=0.9101)), the parameter estimate is 1.103 with a t-statistic of 1.88. Since the weighting function with w=1.2188 places 0.0084 on the first lag and 0.0098 on the second lag, one percent increase of aggregate household consumer debt ratio at the current month would increase the next month market volatility by $e^{1.103*0.0084} - 1 \approx 0.0093$ or 0.9%. If last month's household consumer debt ratio increases by 1%, we would see the increase of 1.1% in the stock market volatility next month. For the second components (which carries the weight of 8.99% (p=0.0899)), the parameter estimate is 2.456 with a t-statistic of 2.13. Since the weighting function with w=2.007 places 0.0006 on the first lag and 0.0011 on the second lag, one percent increase of aggregate household consumer debt at the current month would increase the next month stock market volatility by $e^{2.456*0.0006} - 1 \approx$ 0.14%. If last quarter's household aggregate debt increase by 1%, we would see the increase of 0.27% in the stock market volatility next month. Figure 9 illustrates the plot of the optimal weighting function for both components in the full sample with a lower weight for more recent observations for both components.

The structural break test indicates no presence of structural breaks with 1% significant level. This is not entirely unexpected as the long span of household consumer debt ratio implies strong upward trend.

5.3. Analysis

From the result, rising aggregate household consumer debt ratio has mixed effect in different period of time. Rising household consumer debt decreases stock market volatility for the period from 1962 to 1984 while it increases stock market volatility from the period of 1985 to 2015. The effect change coincides with the sharp rise in the share of income going to household at the top of personal income distribution (starting in the early 1980s).

The result confirms the hypothesis mentioned in Section 3. Behavioral response of households whose share of income declines (bottom 95%) is considered the key explanation for the change in effect of aggregate household consumer debt to stock volatility during period of rising income inequality. In period of rising income inequality (1985-2015), rising household debt is not associated with decreasing low-income consumption rate. This fact leads to higher debt burden to low-income household and accumulating unsustainable debt. This in turn reflects fundamental uncertainty about the long-term consequences of debt accumulation, therefore, increases stock market volatility.

For the period of 1959 to 1984, rising household debt is in accordance to decreasing income share of top-income households (Figure 5). This fact leads to sustainable borrowing and consumption. The stock market responses favorably to the trend, leading to decreasing stock market volatility.

The effect finding of aggregate consumer debt ratio contributes to a substantial literature connects the relatively rapid growth of the U.S. economy during the Great Moderation to aggressive increases in household indebtedness that offset the other-wise negative impact on consumption spending of increasing income inequality (Wisman, 2013, Setterfield 2013, Barba and Pivetti 2009, Cynamom and Fazzari 2008, Palley 2002). Setterfield and Kim (2013) state that, in the presence of emulation effects in consumption behavior, rising inequality of the sort witnessed in the US since 1980 can boost growth. There are a number of literatures that argues that income inequality is inimical to fast growth, together with the past mainstream view that argues income inequality was a necessary side effect of growth and efficiency. However, the sustainability of growth by household consumer debt is questioned in this paper. The consumption growth had been lost now that the bottom 95% is no longer able to expand their balance sheets. That in turn amplifies demand drag in period of rising inequality and increasing stock market volatility.

6. CONCLUSION

From the dataset of US stock return and household consumer debt ratio for more than 50 years (1964-2015), I empirically research the effect of household consumer debt ratio in different period of income inequality. The result confirms the explanatory power (>12%) of household consumer debt to stock market volatility in the U.S. In addition, the paper's result shows that rising household consumer debt in period of rising income inequality has unfavorable effect to stock market volatility. Rising household consumer debt in period of low income inequality, however, has favorable effect to stock market volatility. The GARCH-MIDAS approach with normal mixture errors distribution provides better fit while allowing us to capture the effect of rising household consumer debt in different periods of income inequality.

The different effect of aggregate household consumer debt raises discussion to achieve a long-run stable recovery. As the financial crisis and the Great Recession have shown, the debt-led Growth generated by such strategies in rising income inequality is unsustainable in the long run. The analysis represents the trade-off for the bottom income households group to cut consumption-income ratio in period of rising inequality for its debt-income ratio to be on a sustainable path. The main contributions of our analysis with respect to the previous works can be summarized as (1) analyze the volatility driver of household consumer debt for more than 50 years of *weekly* data (from 1964 to 2015); (2) reduce the trade-off between the accuracy of the volatility measurement provided by high-frequency data of stock prices and the necessity to match it with low-frequency household consumer debt; (3) examine GARCH-MIDAS model with two-component normal mixture errors distribution; (4) empirically examine the effect of rising household consumer debt during different period of income inequality.

There are, however, several drawbacks of the study. The limitation of the study is considered with ideas for further research. First, the study only samples from the U.S. An improvement to the current research would be to enlarge the project samples by including other countries with different level of income inequality. Further research with cross-country analysis is suggested to reveal any divergence in the behavioral response of bottom income households group to rising income inequality, thus empirically account for the effect of rising household consumer debt to stock market volatility. The cross-country analysis helps to confirm the effect of rising household consumer debt during different periods of income inequality. The attention as well as availability for household data in Europe is increasing, thus provide interesting topic for under-research household finance. Other components of household debt, including mortgage debt can also be considered for further research. Second, in this paper, I use GARCH-MIDAS model with two component normal mixture distribution of the same mean. Therefore, the model may not accommodate skewness as well as asymmetric characteristics as found in the distribution of U.S. stock market returns. Future research with different mean and more than two normal mixture distributions can be considered to better capture the leptokurtic and skewed characteristics in the data. Last but not least, GARCH-MIDAS model with two component normal mixture distribution still unable to capture all of the leptokurtosis demonstrated in stock market returns. Adding more economic variables (both in terms of level and variance) may better accommodate the non-normality and asymmetric characteristics of financial time series data.

REFERENCES

Adrian, T., and J. Rosenberg (2006, revised 2008), Stock returns and volatility: Pricing the short run and long-run components of market risk, *Federal Reserve Bank of New York Staff Reports, no. 254*

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. (2014). "Regulating Financial Products: Evidence from Credit Cards."

Alper, C. E., Fendoglu, S. and Saltoglu, B. (2008), "Forecasting stock market volatilities using MIDAS regression: An application to the emerging markets", *MPRA Paper No. 7460.*

Alper, C. E., Fendoglu, S. and Saltoglu, B. (2008). Forecasting stock market volatilities using MIDAS regression: An application to the emerging markets. *MPRA Paper No. 7460.*

Andreaou, E., Ghysels, E. and Kourtellos, A. (2010). Regression models with mixed data sampling frequencies. *Journal of Econometrics*, in-press.

Andreou, E., and E. Ghysels, 2002, Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics* 17, 579–600.

Andrew Filardo (2008), Household debt, monetary policy and financial stability: still searching for a unifying model, *BIS Paper No 46*.

Bai, J., Ghysels, E. and Wright, J. (2009), "State space models and MIDAS regression", *Working Paper, NY Fed, UNC and John Hopkins.*

Belabed, C.A., T. Theobald, and T van Treeck. (2013), "Income distribution and current account imbalances", *IMK working paper 126-2013.*

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman (2010). "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment", *Quarterly Journal of Economics 125 (1): 263–305*

Blattberg, R.C. and N.J. Gonedes (1974). "A comparison of the stable and student distributions as statistical models for stock prices", *J. Bussiness, 47(2): 244-280.*

Campbell, J., and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* 31, 281–318.

Carol Alexander and Emese Lazar (2006), Normal Mixture GARCH (1,1): Applications to Exchange Rate Modelling. *Journal of Applied Econometrics, 307-336.*

Chen, Xilong, Eric Ghysels, and Eric Renault, (2009), "Semi-parametric midas regression and index models", *Working paper, UNC.*

Christelis, Dmitris, Michael Ehrmann, and Dmitris Georgarakos. (2013). "Exploring Differences in Household Debt across Euro Area Countries and the US"

Clark, P. K. (1973), "A subordinated stochastic process model with finite variance for speculative prices", *Econometrica 41(1), 135 – 155.*

Cynamon, Barry Z. and Fazzari, Steven M. (2014), Inequality, the Great Recession, and Slow Recovery. Available at SSRN: http://ssrn.com/abstract=2205524.

DellaVigna, S., and U. Malmendier. (2004). "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics 119 (2): 353–402.*

Dempster, A.P., N.M. Laird and D.B. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm". *J. the Royal Statistical Society Series B, 39: 1-38.*

Duesenberry, J.S. (1952). "Income, Saving and the Theory of Consumer Behavior". *Cambridge, MA: Harvard University Press, second printing*

Dynan, Karen (2009). "Changing Household Financial Opportunities and Economic Security", *Journal of Economic Perspectives 23 (4): 49–68.*

Dynan, Karen, and Donald Kohn. (2007). "The Rise in U.S. Household Indebtedness: Causes and Consequences."

Edelberg, Wendy (2006). "Risk-Based Pricing of Interest Rates for Consumer Loans", *Journal of Monetary Economics 53 (8): 2283–98*

Emmanuel Saez, Gabriel Zucman (2014), "Wealth Inequality in the United States since 1913", *Working Paper 20625, National Bureau of Economic Research*

Emmanuel Saez, Gabriel Zucman (2014), Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data, *Working paper 20625, National Bureau of Economic Research.*

Engle, R. F. and Rangel, J. G. (2008). "The Spline-GARCH Model for low-frequency volatility and its global macroeconomic causes", *Review of Financial Studies, Society for Financial Studies, 21(3): 1187-1222.*

Engle, Robert F. and Ghysels, Eric and Sohn, Bumjean (2008), On the Economic Sources of Stock Market Volatility. *AFA 2008 New Orleans Meetings Paper.*

Everitt, B.S. and D.J. Hand, (1981), "Finite Mixture Distributions", *Chapman and Hall.*

Fama, E.F., (1965), "The behavior of stock-market prices", *J. Bussiness, 38(1): 34-105*

Forsberg, L. and Ghysels, E. (2006), "Why do absolute returns predict volatility so well?", *Journal of Financial Econometrics, 6, 31-67.*

Frühwirth-Schnatter, S., (2006), "Finite Mixture and Markov Switching Models", *NY: Springer*

Gabaix, Xavier, and David Laibson (2006). "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics 121 (2): 505–40*

Georgarakos, Dimitris, Michael Haliassos, and Giacomo Pasini. (2014). “Household Debt and Social Interactions”, *Review of Financial Studies 27 (5): 1404–33.*

Ghysels, E. and Wang F. (2011), “Statistical inference for volatility component models”, *UNC discussion paper.*

Ghysels, E., Sinko, A. and Valkanov, R. (2006), “MIDAS regression: Further results and new directions”, *Econometric Reviews, 26, 53-90*

Ghysels, Eric, (2009), “Mixed data sampling and temporal aggregation”, *Encyclopedia of Quantitative Finance*

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2005, There is a risk-return tradeoff after all, *Journal of Financial Economics* 76, 509–548.

Gine, Xavier, Cristina Martinez, and Rafael Keenan (2014). “Financial (dis)information: Evidence from an Audit Study in Mexico.”

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), “Global optimization of statistical functions with simulated annealing”, *Journal of Econometrics, 60, 65–99*

Gridgeman, N.T., (1970), “A Comparison of Two Methods of Analysis of Mixtures of Normal Distributions”, *Technometrics, 12: 823-833.*

Gurun, Umit, Gregor Matvos, and Amit Seru. (2014). “Advertising Expensive Mortgages.”

Hacker, Jacob S. (2008). “The Great Risk Shift: The New Economic Insecurity and the Decline of the American Dream”

Heidhues, Paul, and Botond Koszegi. (2010). “Exploiting Naïvete about Self-Control in the Credit Market”, *American Economic Review 100 (5): 2279–2303*

Hossein Asgharian, Ai Jun Hou, Farrukh Javed (2013), Importance of the macroeconomic variables for variance prediction – A Garch-Midas Approach, *Working paper*

Jonathan Zinman (2014), “Household Debt: Facts, Puzzles, Theories, and Policies”, *In preparation for the Annual Review of Economics.*

Kim, C.J., and C.R. Nelson, 1999, Has the economy become more stable? a Bayesian approach based on a markov-switching model of the business cycle, *The Review of Economics and Statistics* 81, 608–616.

Kon, S.J., (1984), “Models of stock returns – a comparison”, *J. Finance, 39(1): 147-165.*

Kotze, G. L. (2007), “Forecasting inflation with high frequency asset price data”, *Working Paper. University of Stellenbosch.*

Kumhof, Ranciere and Winant (2014), “Inequality, Leverage and Crises: The Case of Endogenous Default”, *IMF Working Paper.*

Levine, A.S., R.H. Frank, and O. Dijk. (2010). “Expenditure cascades”, *available at SSRN: http://ssrn.com/abstract=1690612 or http://dx.doi.org/10.2139/ssrn.1690612*

Lindsay, B.G., (1995), “Mixture Models: Theory, Geometry and Applications”, *Hayward, CA: Institute of Mathematical Statistics.*

M. Maki (1999), The Growth of Consumer Credit and the Household Debt Service Burden. *Credit Research Centre’s 25th Anniversary Conference, Arlington, VA.*

Marglin, S.A. (1984), “Growth, Distribution, and Prices”, *Cambridge, MA: Harvard University Press.*

Mark, Yun, Jeremy (2014), Inequality, Debt Servicing and The sustainability of Steady State Growth, *Available at SSRN: http://ssrn.com/abstract=2514431 or http://dx.doi.org/10.2139/ssrn.2514431*

Mason, J.W. and A. Jayadev. (2014a), “‘Fisher dynamics’ in US household debt, 1929- 2011”, *American Economic Journal: Macroeconomics, vol. 6, no. 3, 214-34.*

McConnell, M.M., and G. Perez-Quiros, 2000, Output fluctuations in the United States: What has changed since the early 1980s?, *American Economic Review* 90(5), 1464–1476.

McLachlan, G. and D. Peel, (2000), “Finite Mixture Models”, *NY: John Wiley & Sons*

McLachlan, G.J. and K.E. Basford (1988), “Mixture Models: Inference and Applications to Clustering”, *NY: Marcel Dekker.*

Meyer, Bruce D., and James X. Sullivan. 2013. "Consumption and Income Inequality and the Great Recession." *American Economic Review*, 103(3): 178-83.

Newcomb, S., (1963). “A generalized theory of the combination of observations so as to obtain the best result, in M. Stigler, editor American Contributions to Mathematical Statistics in the Nineteenth Century”, *New York: Arno Press, 2: 343-366.*

Picard, F., (2007), “An introduction to mixture models Statistics for Systems Biology Group”, *Research Report No 7.*

Piketty, Thomas and Saez, Emmanuel (2007), Income and Wage Inequality in the United States 1913-2002; in Atkinson, A. B. and Piketty, T. (editors) Top Incomes over the Twentieth Century. A Contrast Between Continental European and English-Speaking Countries, *Oxford University Press, chapter 5*.

Praetz, P. D. (1972), “The distribution of share price changes”, *The Journal of Business 45(1), 49–55.*

Press, S. J. (1967), “A compound events model for security prices”, *Journal of Business 40, 317 – 335.*

Rinaldi, Laura and Sanchis-Arellano, Alicia (2006), Household Debt Sustainability: What Explains Household Non-Performing Loans? An Empirical Analysis. *ECB Working Paper No. 570*. Available at SSRN: http://ssrn.com/abstract=872528.

Schwert, G. W., (1989), Why does stock market volatility change over time?, *Journal of Finance* 44, 1207–1239.

Stango, Victor, and Jonathan Zinman. (2014a). “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees”, *Review of Financial Studies 27 (4): 990–1030.*

Tay, A. S. (2007), “Mixed frequencies: Stock returns as a predictor of real output growth”, *Discussion Paper, SMU.*

Titterington, D.M., A.F.M. Smith and U.E. Makov (1985), “Statistical Analysis of Finite Mixture Distributions”, *John Wiley & Sons, New York.*

Van Treeck, T and S. Sturn. (2012), “Income inequality as a cause of the Great Recession? A survey of current debates”, *ILO, Conditions of Work and Employment Branch.*

Vissing-Jørgensen, Annette (2007). “Household Finance: The Liability Side.”, *Introduction of Session Organizer at the 2007 Gerzensee European Summer Symposium.*

Zetty, Zaidi and Mohd (2012), “Mixtures of Normal Distributions: Application to Bursa Malaysia Stock Market Indices”, *World Applied Sciences Journal 16 (6): 781-790*

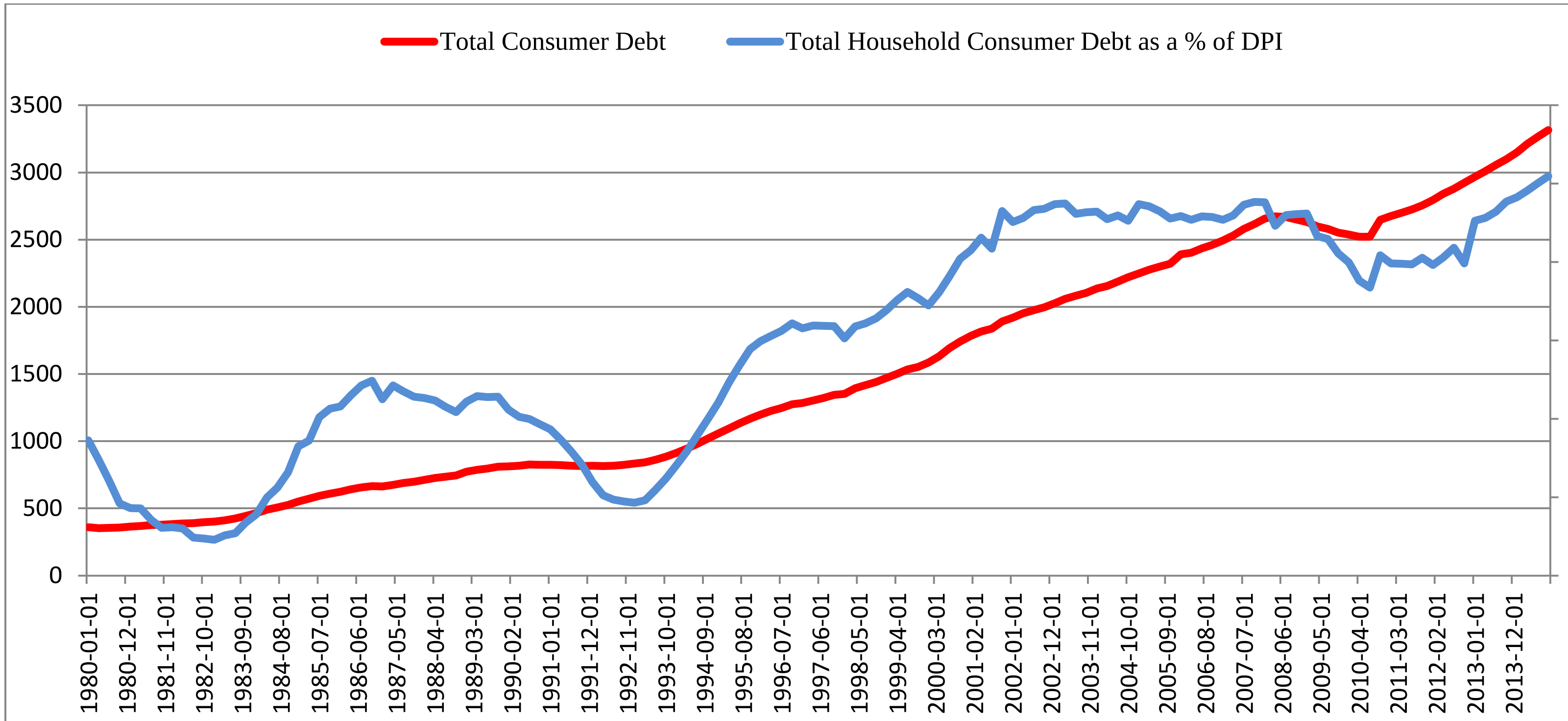


Figure 1. Total Household Consumer Debt and Total Household Consumer Debt as a Percent of Disposable Personal Income
Source: Federal Reserve Economic Data

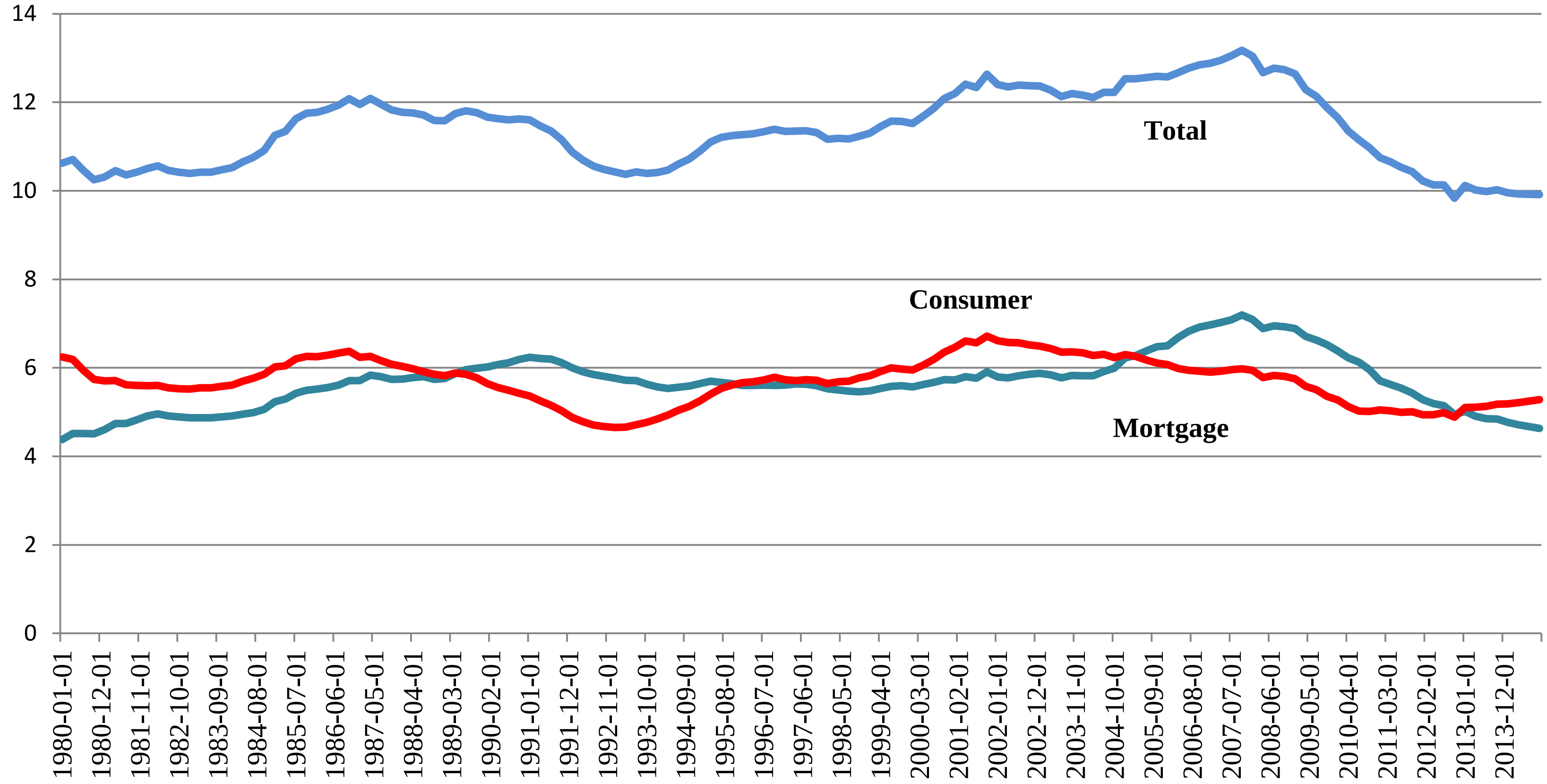


Figure 2. Total household debt service payment, mortgage debt service payment, consumer debt service payment to disposable income.
Source: Federal Reserve Economic Data

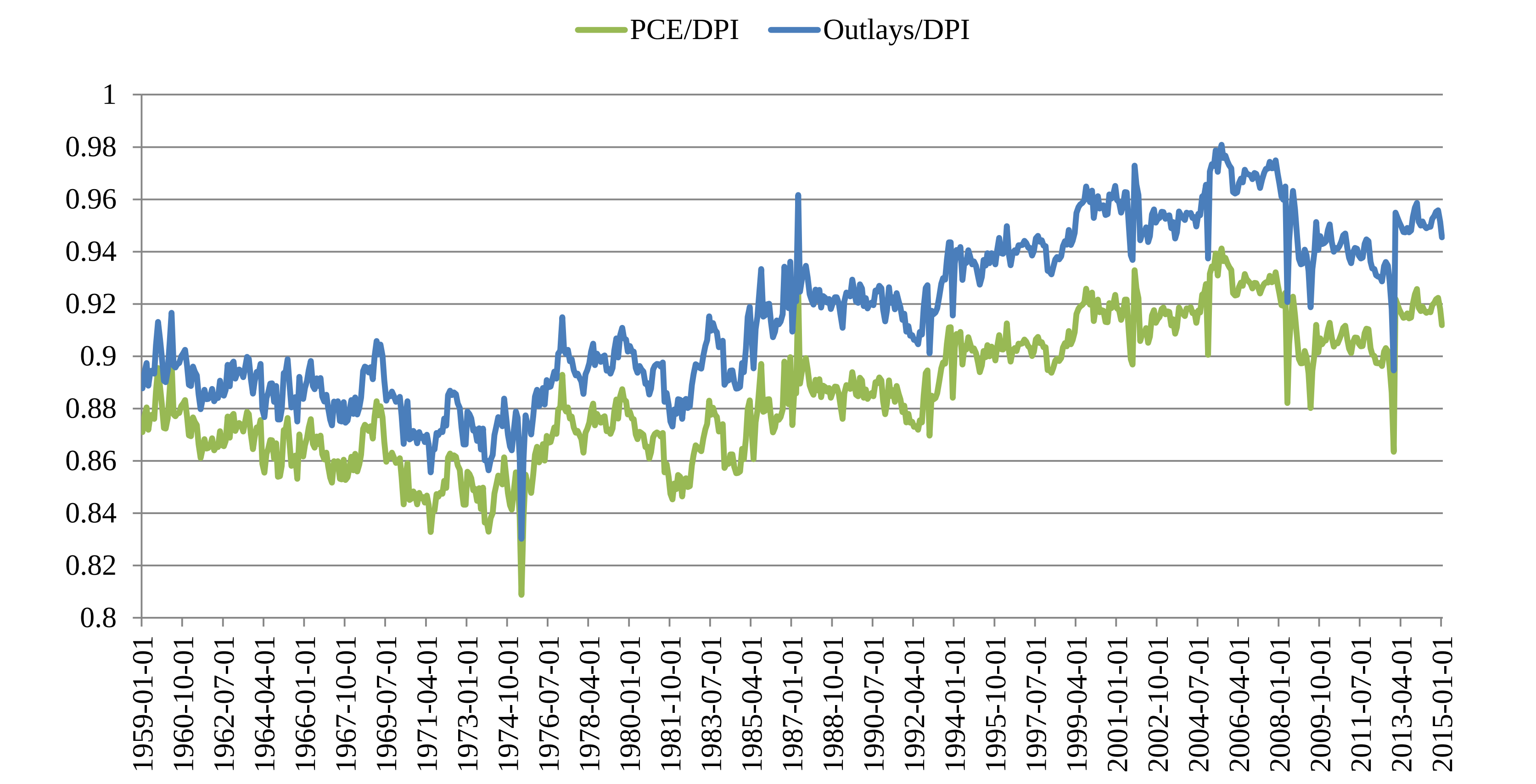


Figure 3. Personal Consumption Expenditure per Disposable Personal Income (PCE/DPI), Outlays per Disposable Personal Income (Outlays/DPI)
Source: PCE/DPI, Outlays/DPI calculated from PCE, Outlays and DPI series from US. Bureau of Economic Analysis.

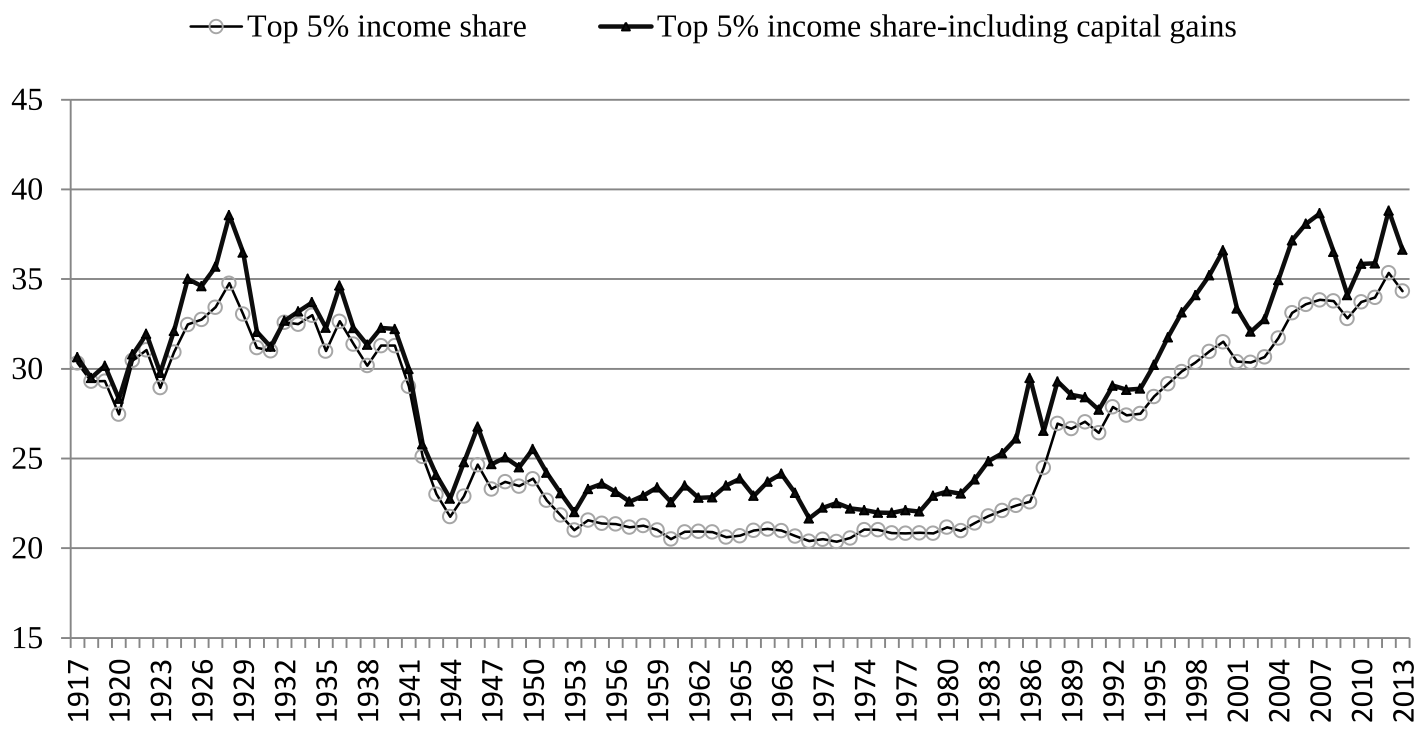


Figure 4: Top Decile Income share, 1917-2013
Source: Top Income Database

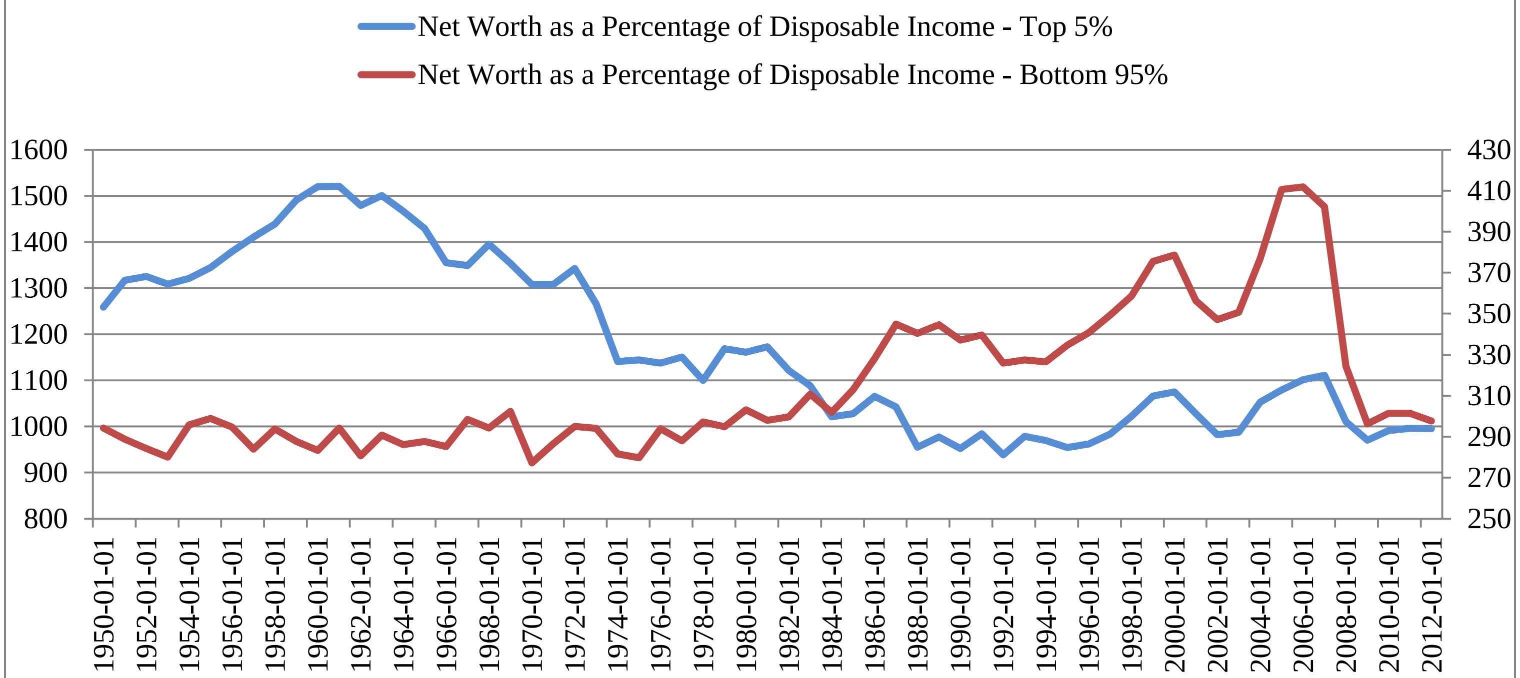


Figure 5. Household Net Worth to Disposable Income
Source: Net Woth as a Percentage of Disposable Income from FRED, Net Worth Share of Top 5% Income from SaezZucman 2014, Top 5% Income Share from Top Income Database

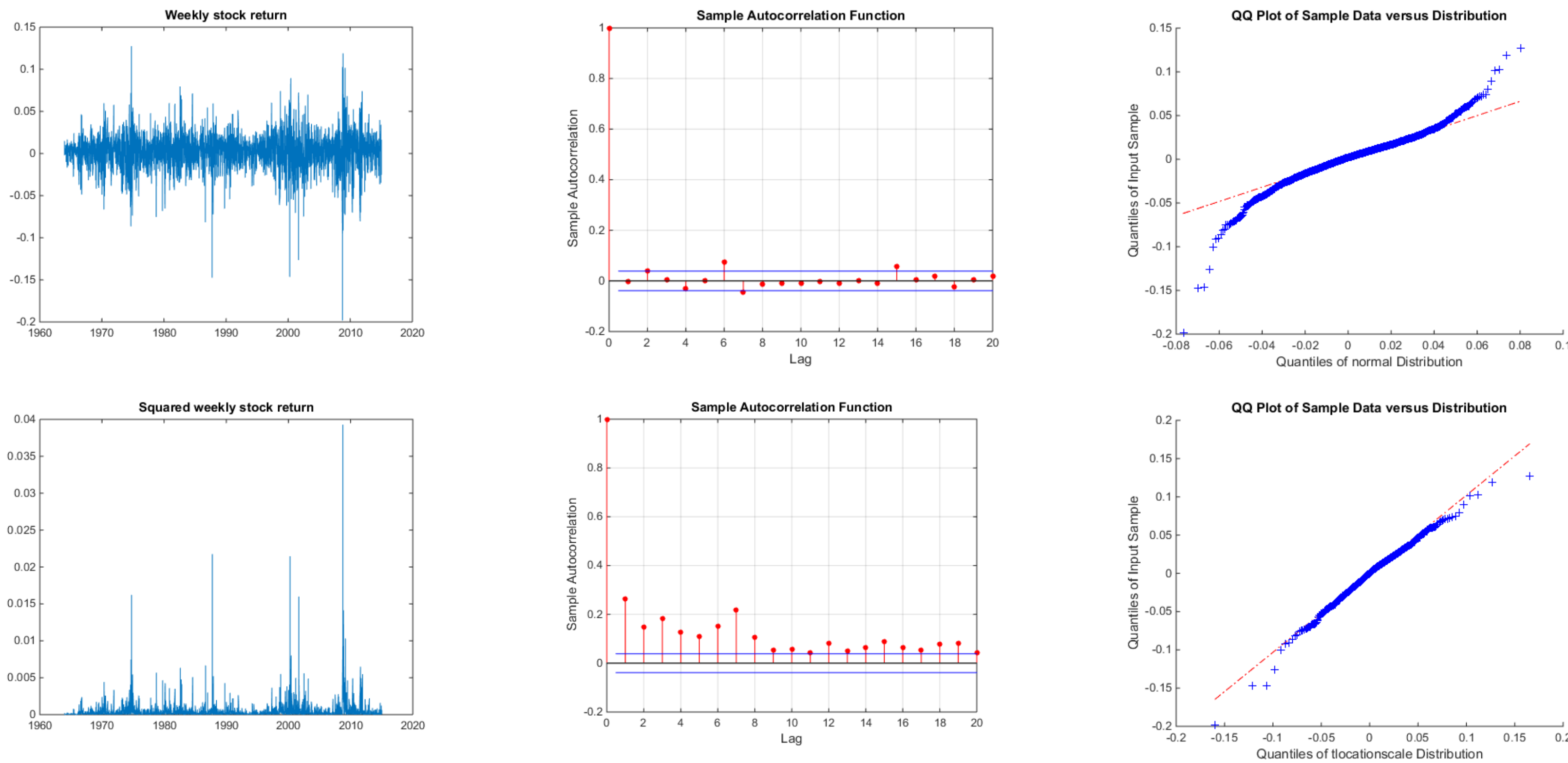


Figure 6. (Row 1) From left to right: Weekly U.S. stock return, Sample Autocorrelation Function of weekly stock return, QQ plot versus best fitted normal distribution period from 01-01-1964 to 01-01-2015.

(Row 2) From left to right: Squared weekly U.S. stock returns, Sample Partial Autocorrelation Function of squared weekly stock return, QQ plot versus best fitted t-student distribution for period from 01-01-1964 to 01-01-2015.

Data Source: Kenneth R. French

Figure 7. (Rows) From left to right, Normal Mixture GARCH-MIDAS, Normal Distribution GARCH-MIDAS, Student-t distribution GARCH-MIDAS for period from 1964 to 2015.

(Columns) From up to down: QQ plot, Sample Autocorrelation Function of z, Sample Autocorrelation Function of squared z, conditional volatility and its long-run component, optimal weight function and errors distribution.

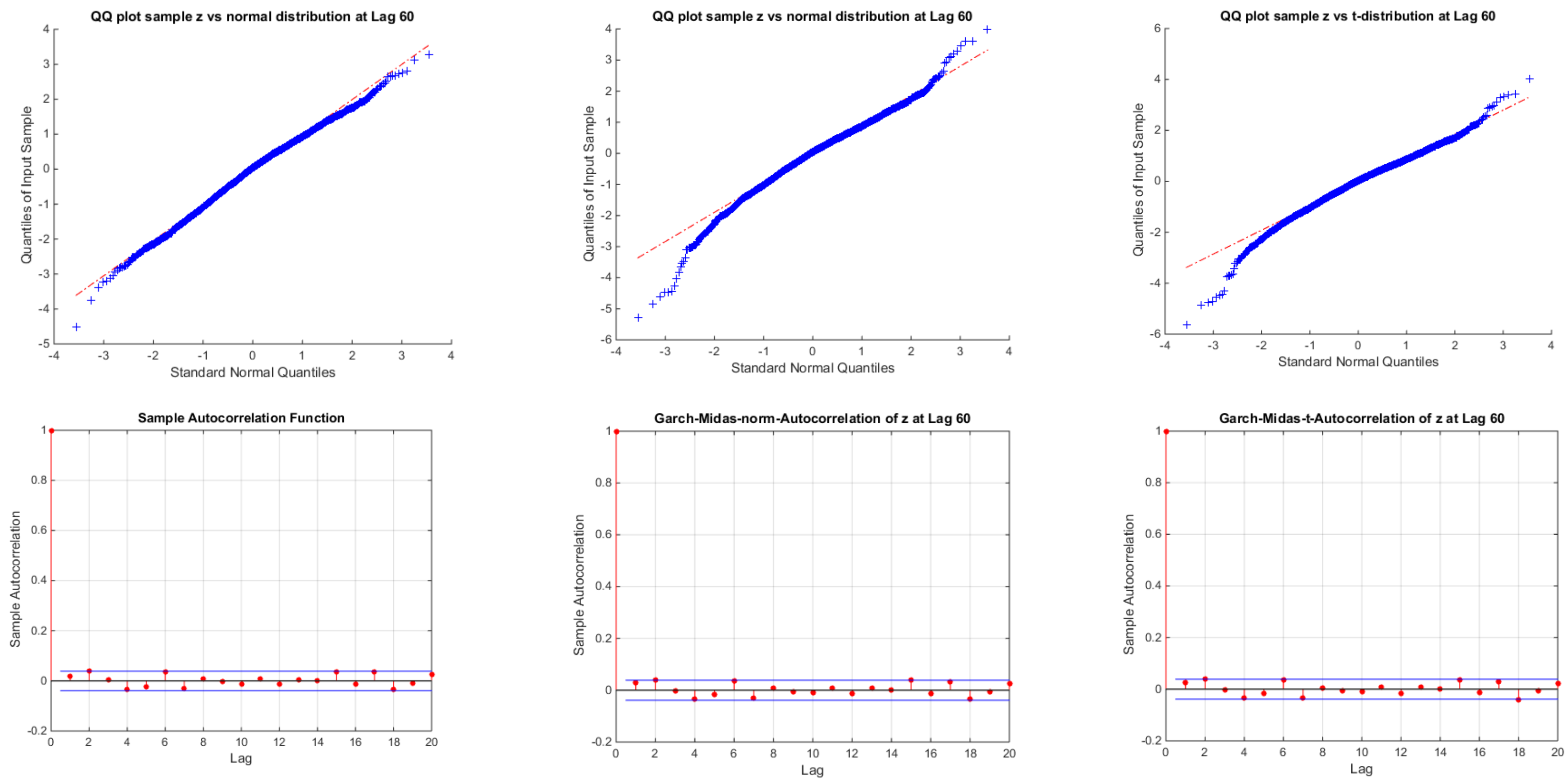


Figure continues to next page…

Continue from Figure 7

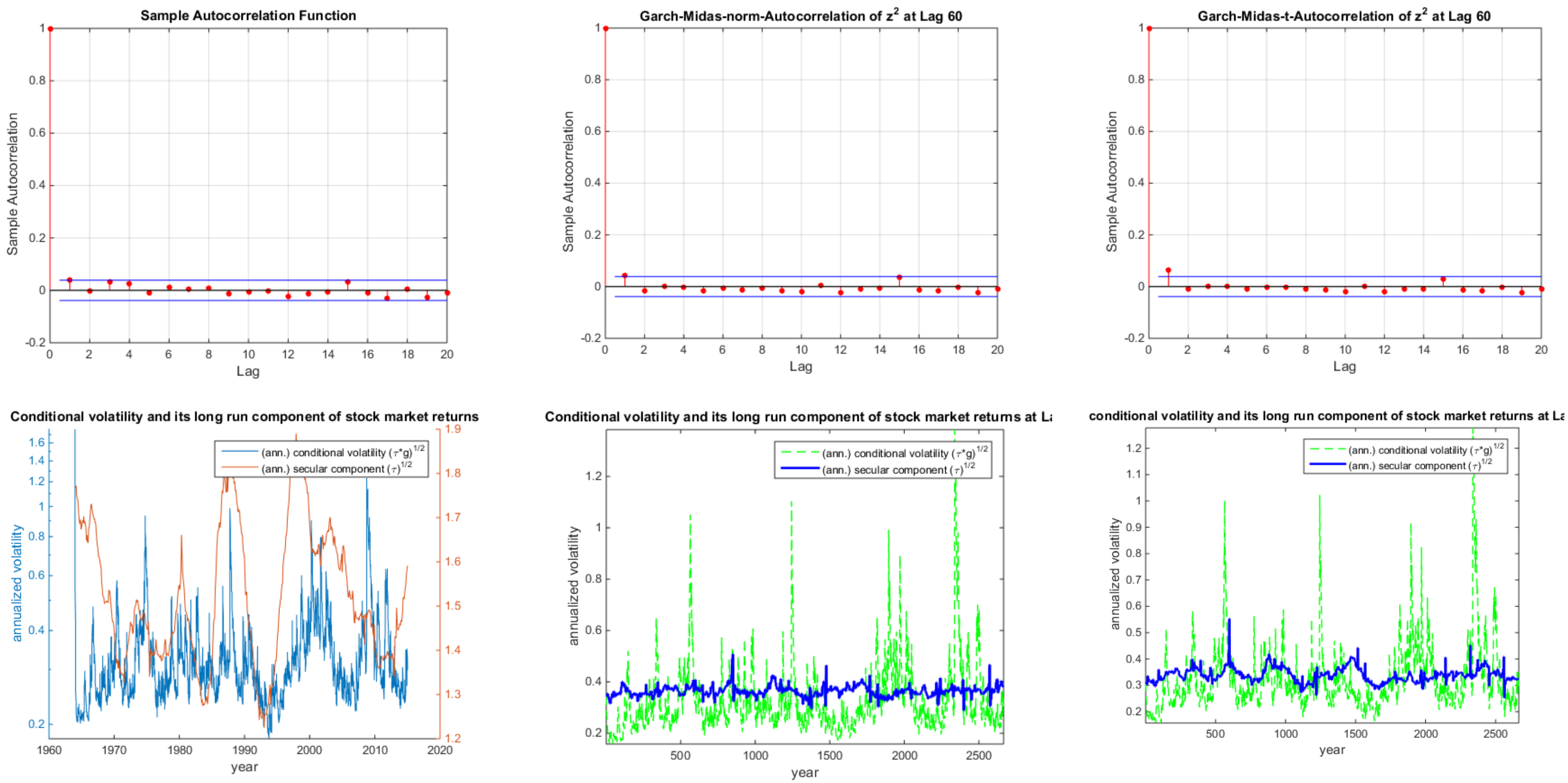


Figure continues to next page…

Continue from Figure 7

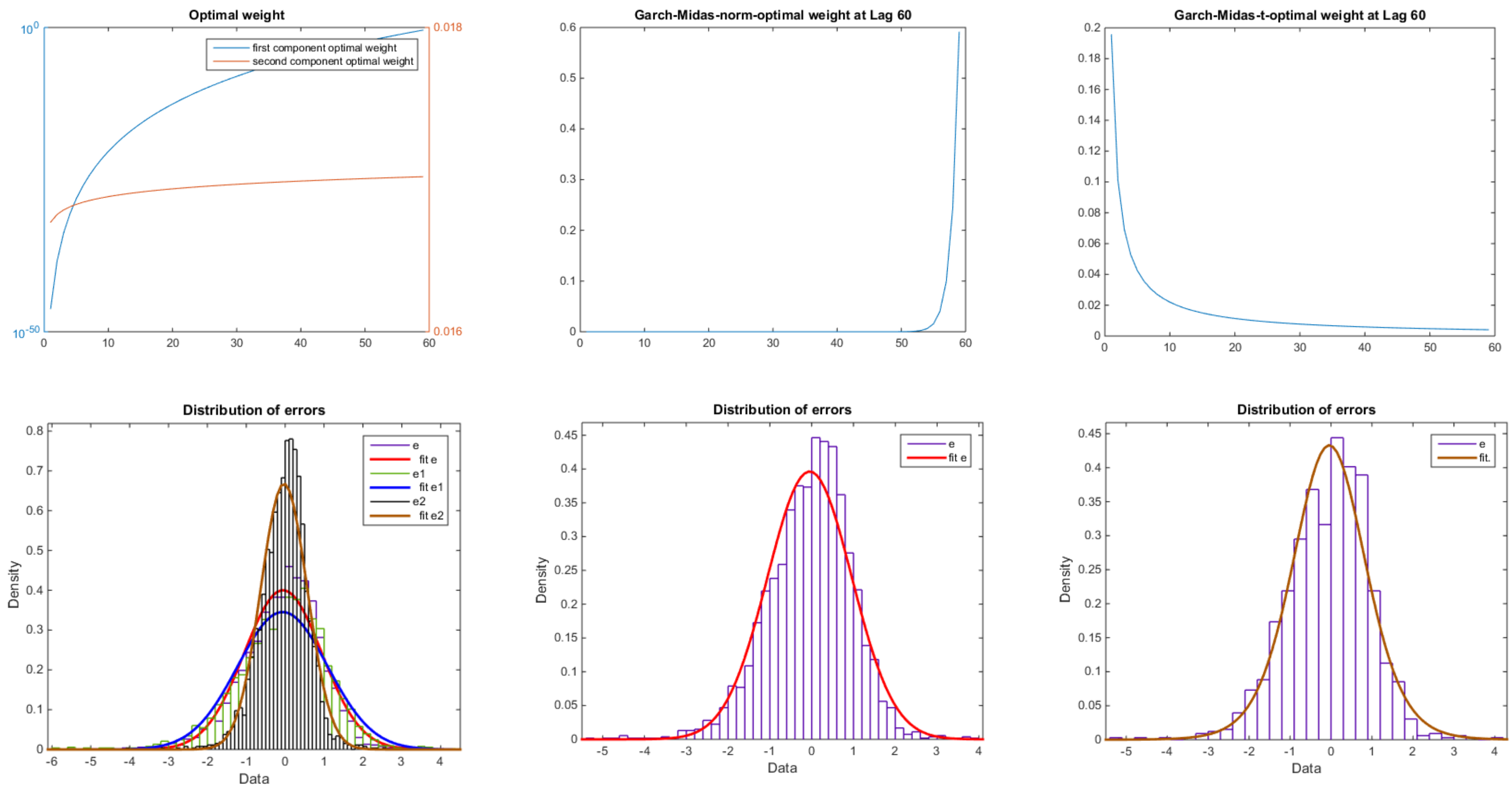


Figure 8. (Rows) From left to right, Normal Mixture GARCH-MIDAS, Normal Distribution GARCH-MIDAS, Student-t distribution GARCH-MIDAS for period from 01-01-1964 to 01-12-1984 (dd-mm-yy)

(Columns) From up to down: QQ plot, Sample Autocorrelation Function of z, Sample Autocorrelation Function of squared z, conditional volatility and its long-run component, optimal weight function and errors distribution for period from 01-01-1964 to 01-12-1984 (dd-mm-yy)

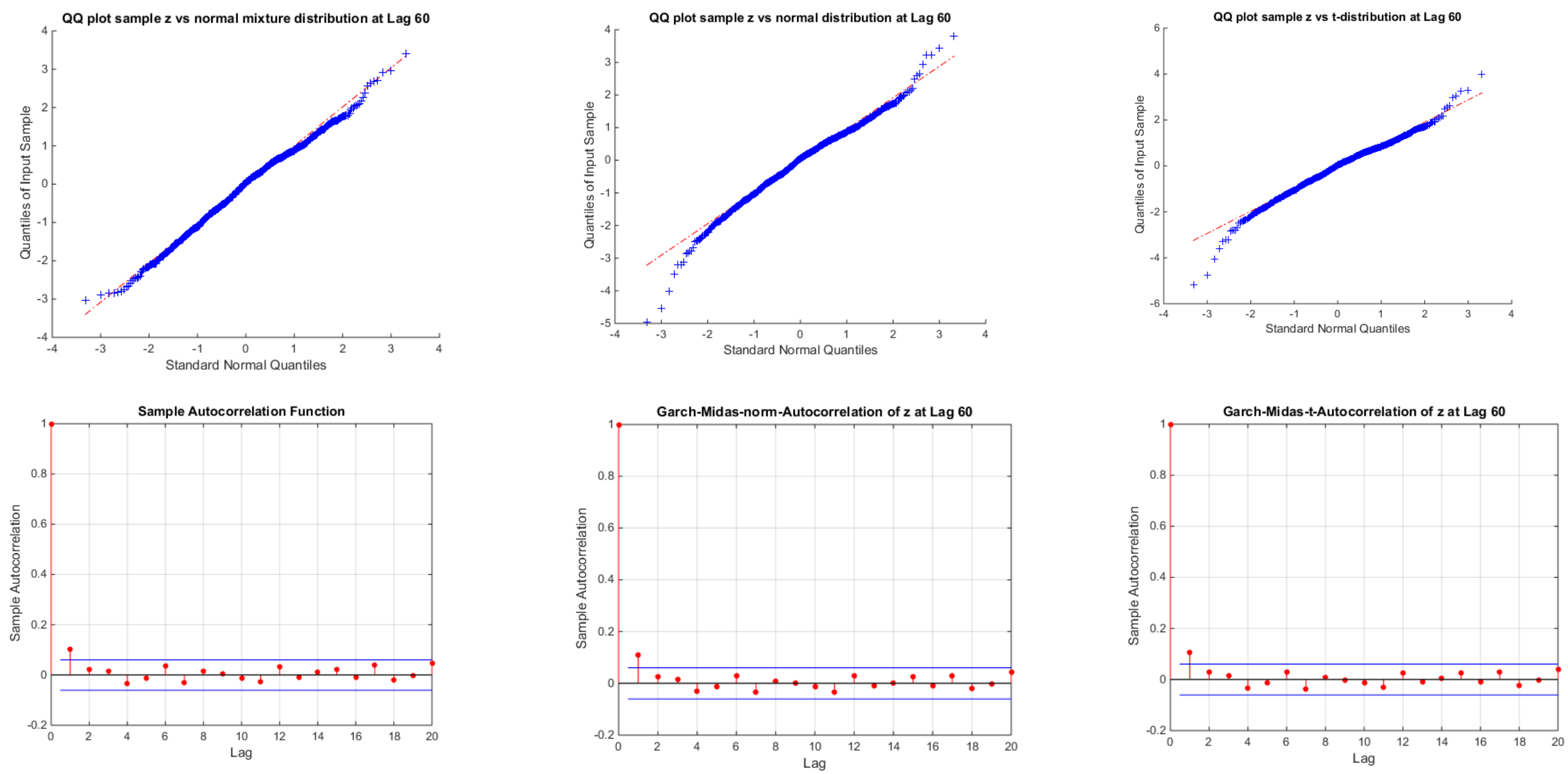
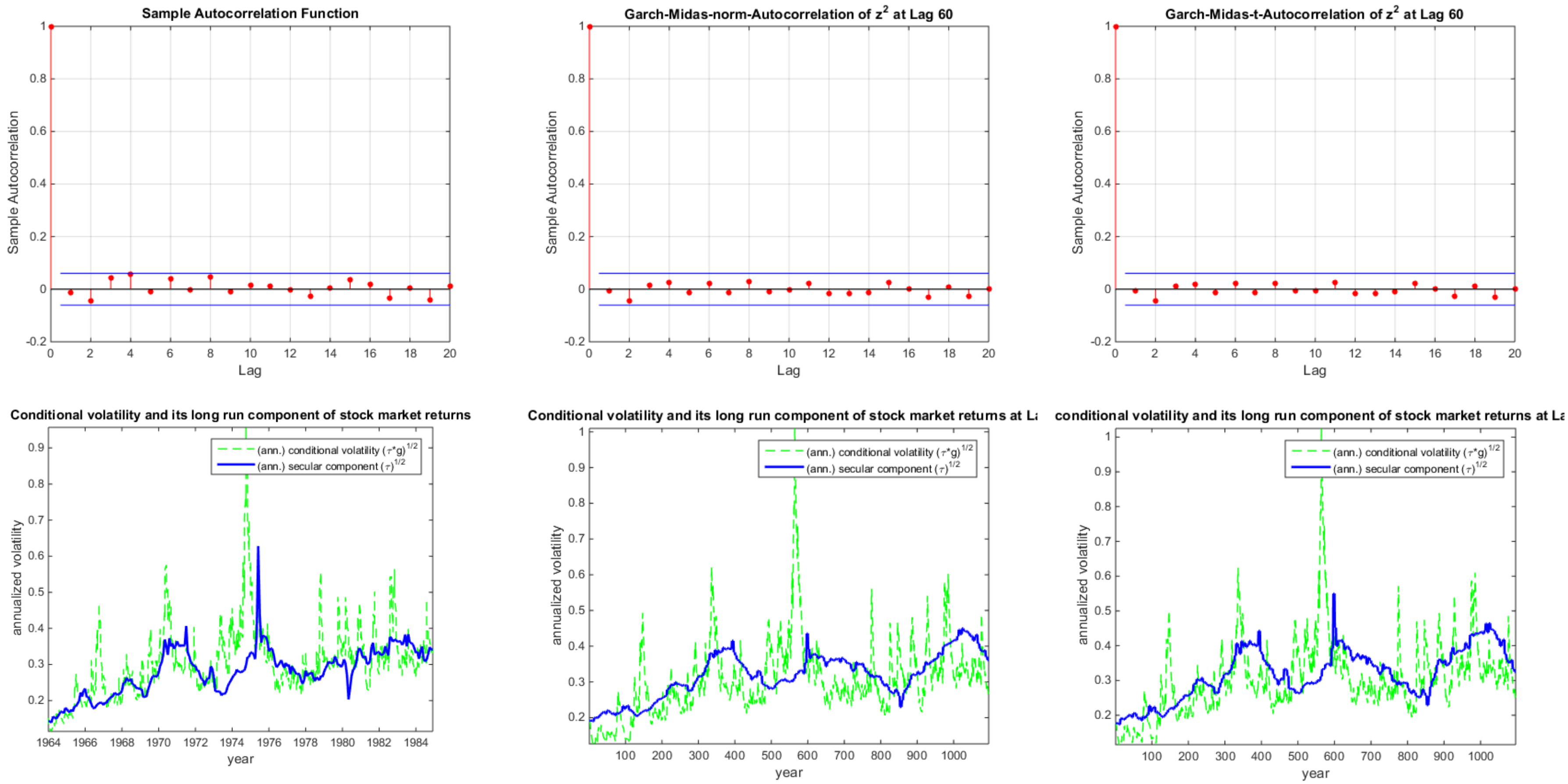


Figure continues to next page…

Continue from Figure 8



Continue from Figure 8

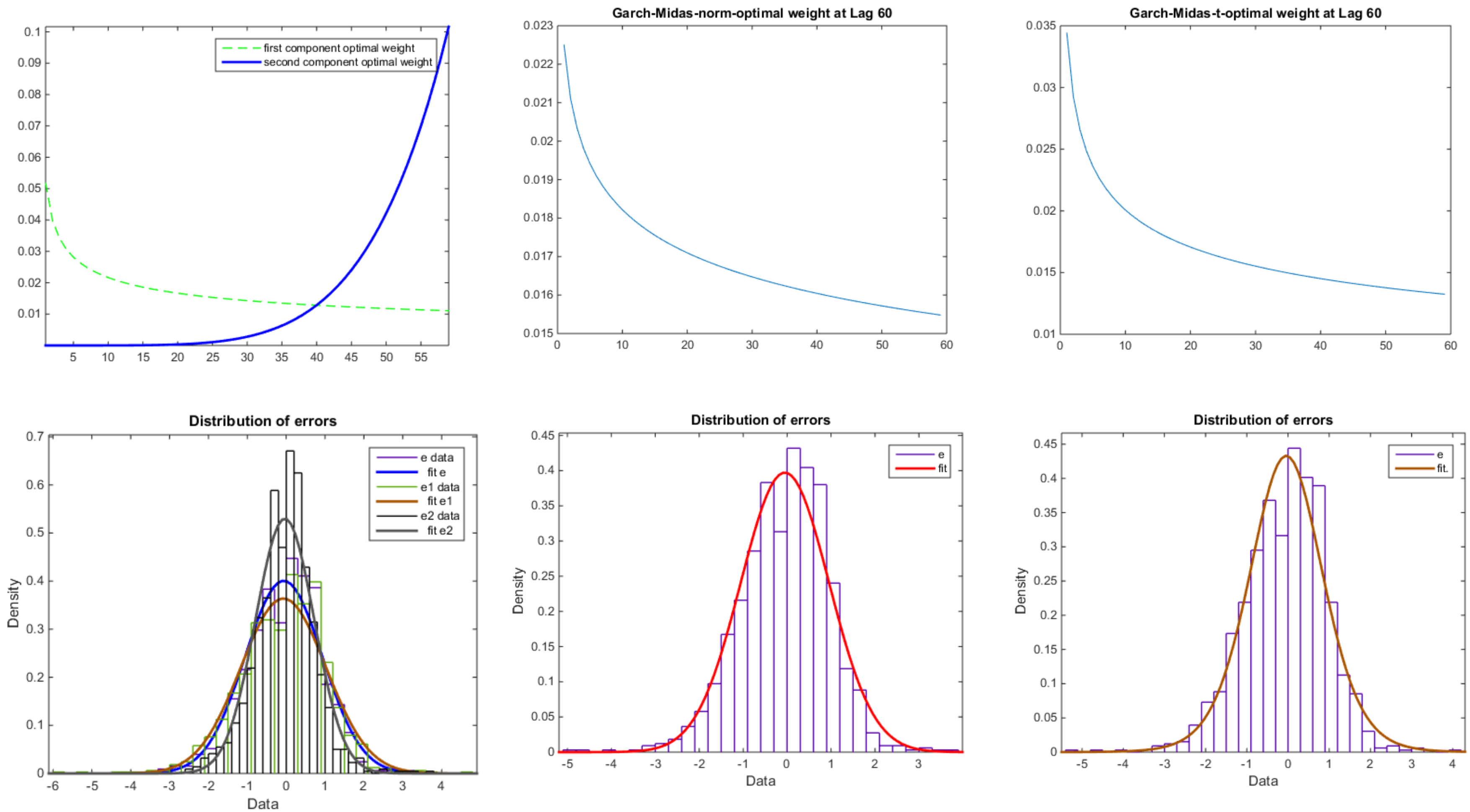


Figure 9. (Rows) From left to right: Normal Mixture GARCH-MIDAS, Normal Distribution GARCH-MIDAS, Student-t distribution GARCH-MIDAS for period from 01-01-1985 to 01-01-2015 (dd-mm-yy).

(Columns) From up to down: QQ plot, Sample Autocorrelation Function of z, Sample Autocorrelation Function of squared z, conditional volatility and its long-run component, optimal weight function and errors distribution for period from 01-01-1985 to 01-01-2015 (dd-mm-yy).

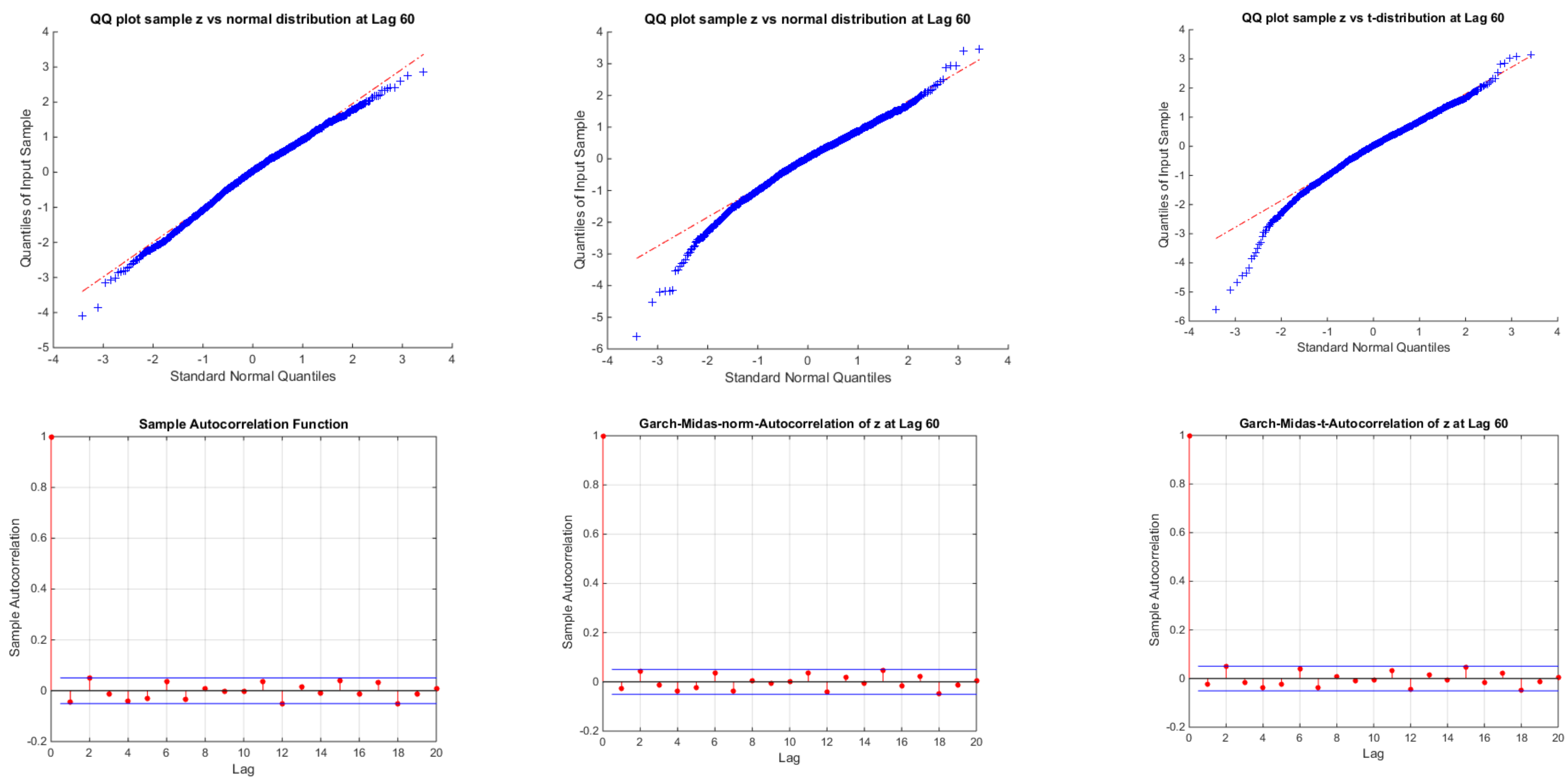
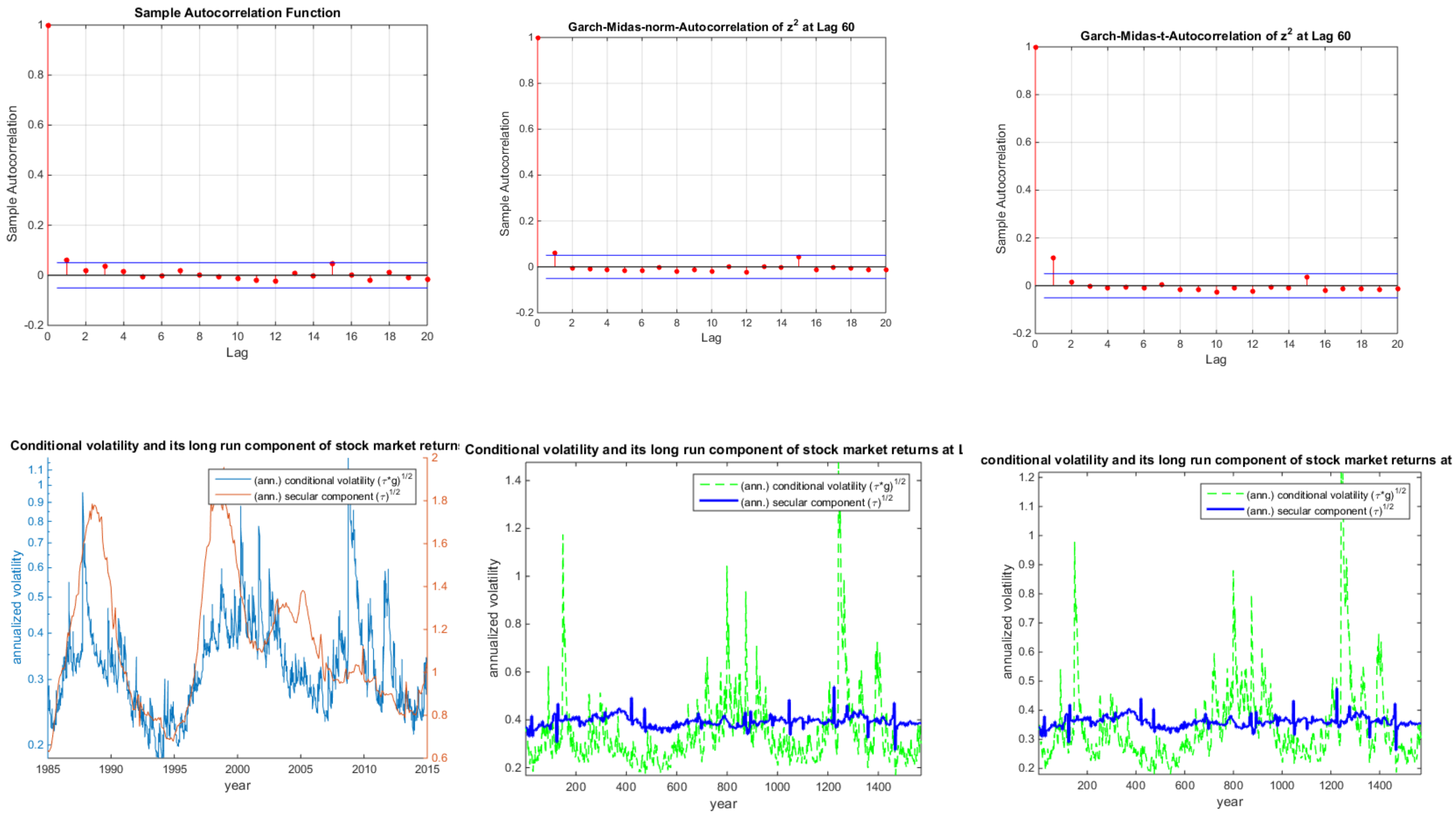


Figure continues to next page….

Continue from Figure 9



Continue from Figure 9

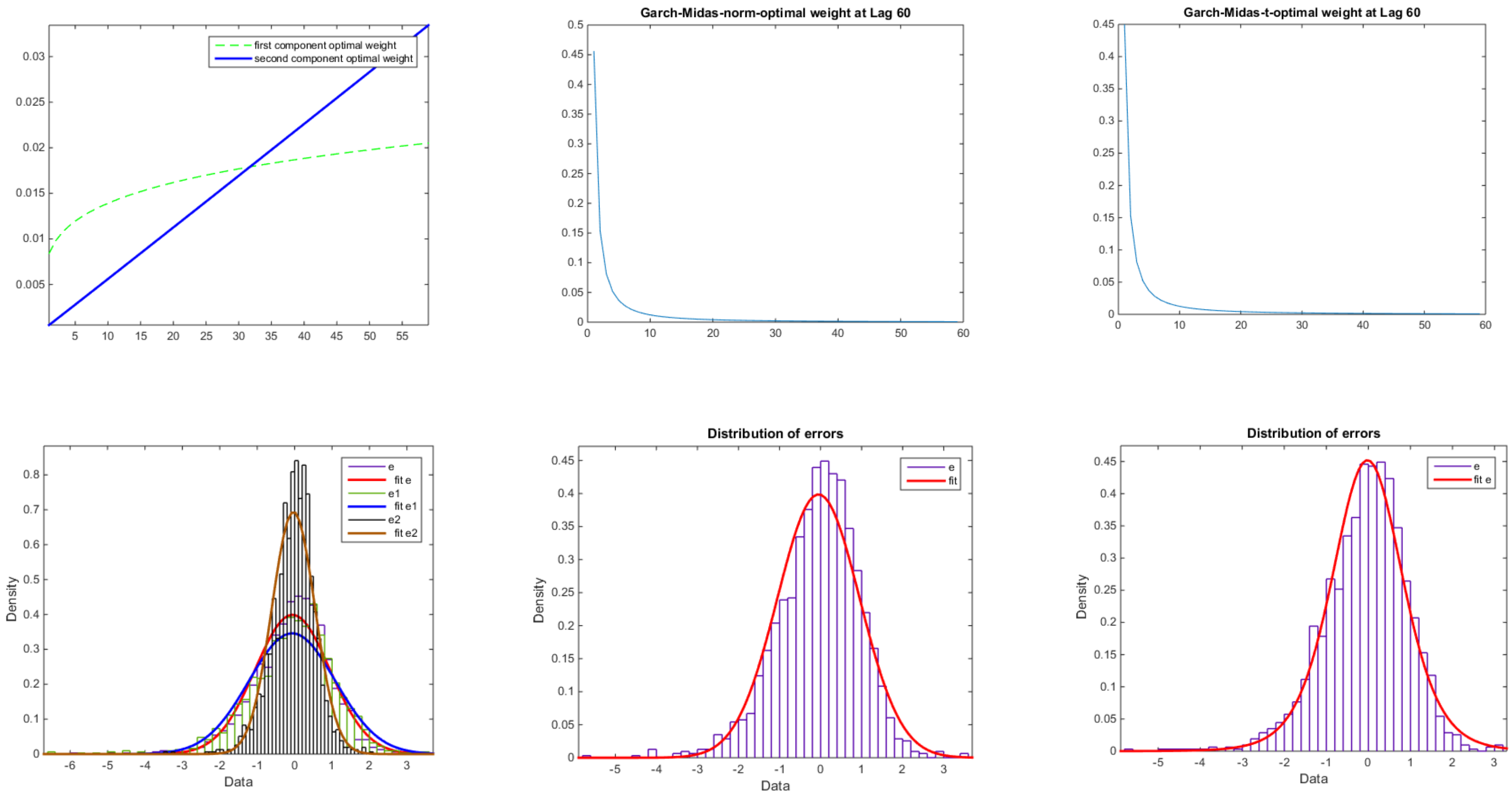


Table 01: Parameter Estimates for GARCH-MIDAS

GARCH-MIDAS models with various specifications are fitted via QMLE. For all specifications, 60 lags are taken to model $\log\tau_t$. θ are rescaled by multiplication of 10^{-2} to make the level repesented in percentage unit. The model specification has different interpretations for GARCH-MIDAS model with different errors distribution. The numbers in the parenthesis are t-stats (***p<0.01, **p<0.05, *p<0.1).

Model	Period	Component	μ	α	β	m	θ	w	v	p	LLF/BIC	MSE/ MAE	KS Stat/ LBQ	GARCH-LLF/BIC
GARCH-MIDAS NM	1964-2015	1	0.00	0.07	0.89	(8.79)	(0.30)	26.84		0.89	(6,752.98)	1.00	0.09	
			8.64	5.34	44.76	(30.29)	(1.41)*	1.09		20.95	(5.03)	0.77	0.08	
		2		0.47	0.52	(2.55)	1.21	1.00						
				3.10	3.41	(0.61)	1.44*	2.07						
GARCH-MIDAS Norm	1964-2015		0.00	0.15	0.82	(7.53)	(0.22)**	52.80			(6,703.28)	1.02	0.00	6,699.4
			8.08	7.91	37.28	(44.09)	(1.92)	1.49			(5.03)	0.78	0.10	(5.01)
GARCH-MIDAS Student-t	1964-2015		0.00	0.12	0.84	(7.65)	(0.62)*	0.05	8.39		(6,748.40)	1.02	0.00	
			8.98	6.27	34.82	(45.37)	(1.59)	0.14	7.43		(5.06)	0.78	0.09	

Table continues to next page …

Continue from Table 01

Model	**Period**	**Component**	**μ**	**α**	**β**	**m**	**θ**	**w**	**v**	**p**	**LLF/BIC**	**MSE/ MAE**	**KS Stat/ LBQ**	**GARCH-LLF/BIC**
GARCH-MIDAS NM	1964-1984	1	0.00	0.13	0.79	(8.28)	(4.07)***	0.62		0.91	(2,853.30)	0.99	0.11	
			5.49	4.82	21.88	(48.69)	(4.93)	4.98		22.19	(5.13)	0.78	0.23	
		2		0.02	0.71	(7.10)	(1.97)**	6.32		0.09				
				0.09	0.82	(12.72)	(2.20)	1.84						
	1985-2015	1	0.00	0.06	0.90	(8.56)	1.10**	1.22		0.91	(3,920.98)	1.00	0.39	
			7.20	4.02	37.65	(46.81)	1.88	3.22		20.87	(4.94)	0.76	0.06	
		2		0.67	0.32	(3.24)	2.46**	2.01					0.74	
				3.37	1.63	(1.02)	2.13	1.72		0.09				
GARCH-MIDAS Norm	1964-1984		0.00	0.14	0.81	(7.80)	(3.12)***	0.91			(2,830.64)	1.01	0.02	2,821.5
			4.72	5.47	26.11	(48.40)	(4.60)	4.53			(5.16)	0.79	0.22	(5.12)
	1985-2015		0.00	0.17	0.79	(7.37)	(0.28)*	(0.57)			(3,883.27)	1.00	0.02	3,880.5
			6.88	5.93	22.37	(31.37)	(1.55)	(0.99)			(4.94)	0.77	0.17	(4.92)
GARCH-MIDAS student-t	1964-1984		0.00	0.14	0.80	(7.81)	(3.54)***	0.77	9.77		(2,844.81)	1.02	0.00	
			5.53	4.99	22.77	(40.88)	(4.69)	4.82	4.24		(5.18)	0.79	0.25	
	1985-2015		0.00	0.11	0.86	(7.55)	(0.25)	(0.56)	7.20		(3,915.94)	1.01	0.00	
			7.70	3.87	23.87	(31.48)	(1.26)	(0.79)	6.17		(4.98)	0.76	0.14	

Continue from Table 01

Test statistics

LLF	Optimal log-likelihood function value
BIC	Bayesian Information Criterion
MSE	Mean squared errors
MAE	Mean absolute errors
KS stat	Kolmogorov-Smirnoff (KS) statistic (p-value)
LBQ	Ljung-Box Q-test statistic (p-value)
GARCH-LLF	Optimal log-likelihood function value for GARCH (1,1)
GARCH BIC	Bayesian Information Criterion for GARCH (1,1)

APPENDIX

Appendix A. Ljung-Box Test

The Ljung-Box Q-test is a "portmanteau" test that assesses the null hypothesis that a series of residuals exhibits no autocorrelation for a fixed number of lags L, against the alternative that some autocorrelation coefficient ρ(k), k = 1, ..., L, is nonzero.

The test statistic is

$$Q = T(T+2)\sum_{k=1}^{L}\left(\frac{\rho(k)^2}{(T-K)}\right)$$

Where T is the sample size

L is the number of autocorrelation lags

$\rho(k)$ is the sample autocorrelation at lag k.

Under the null hypothesis, the asymptotic distribution of Q is chi-squared with L degrees of freedom

Appendix B. Engle ARCH test

Engle's ARCH test assesses the null hypothesis that a series of residuals (rt) exhibits no conditional heteroscedasticity (ARCH effects), against the alternative that an ARCH(L) model describes the series.

The ARCH(L) model has the following form:

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_L r_{t-L}^2 + e_t$$

Where there is at least one $\alpha_j \neq 0; j = 0, \ldots, L$

The test statistics is the Lagrange multiplier statistic TR^2, where T is the sample size and R^2 is the coefficient of determination from fitting ARCH(L) model for a number of lags (L) via regression.

Under the null hypothesis, the asymptotic distribution of the test statistic is the chi-square with L degrees of freedom.