

Forecasting Fund Flows in Indian Equity Mutual Funds Market using Time Series Analysis: An Empirical Investigation

Priya Malhotra^{1*} and Pankaj Sinha²

¹Ph.D. Research Scholar, Faculty of Management Studies, University of Delhi, Delhi-110007, India; priyamalhotra.17@fms.edu

²Professor, Faculty of Management Studies, University of Delhi, Delhi-110007, India; pankaj-sinha@fms.edu

Received: 30 August 2020

Accepted: 16 April 2021

Published: 15 Nov 2021

Abstract

Mutual Funds are the second most preferred financial investment option in India amongst households, corporate and private investors alike. Managed funds bring with them the expertise of fund managers along with the benefits of diversification and lower costs. The sensitivity of fund flows defines the ability of the fund manager in offering expected future returns. Mutual fund flows exhibit time series characteristics, it being financial data collected at regular intervals over a time period. This paper studies the dynamics of mutual fund flows by utilising time series regression modelling. Monthly fund flows data for a sample of 142 equity open-ended growth orientation across major marketcap categories – Large Cap, Large and Mid Cap, Multi Cap, Mid Cap, and Small Cap have been analysed using ARIMA Modelling in the R software package. Appropriate lag length and the presence of a unit root have been investigated with the help of established techniques coupled with suitable checks of robustness. Model of best fit has been used to forecast monthly fund flows for a lag length of 60. Our study leads us to two major outcomes. One, unlike many developed and emerging markets, fund flows in the chosen sample do not confirm to positive feedback trading hypothesis. This lends credible support to the absence of irrational exuberance in mutual fund investments. Second, equity-based funds in Large Cap, Large and Mid Cap, and Multi Cap category exhibit strong trend component while funds in Mid Cap and Small Cap category have a strong random component. Beginner investors can take advantage of alpha offered by fund managers possessing effective market-timing skills, an indicator of trend-investing strategy. Funds belonging to these categories are also lesser prone to market volatility in comparison to Mid Cap and Small Cap funds, being more suitable for experienced investors.

Keywords: Fund Flows, Feedback-Trading, Time Series, ARIMA Modelling, Forecasting

JEL classification: C23, G12, G23

1. Introduction

Mutual Funds are the second most preferred investment option in India among households, Corporate and Private Investors alike. Indian Mutual Fund industry comprises 44 Asset Management Companies (AMC) managing assets worth ₹23.5 Trillion. 6846 open-ended schemes and 9315 close-ended schemes exist in the mutual fund market

across all asset classes, as of January 2020. Volatility in fund flows defines the ability of fund managers to diversify the portfolio and deliver promised returns to the investors. Mutual fund flows exhibit time series characteristics of mean-reversion, presence of trend and seasonal patterns, and demonstrate structural breaks. Understanding the dynamics of fund flows is important because investment in funds influences individual units' savings, future wealth,

*Author for correspondence

and fund managers' earnings as well as their incentives. These flows, hence have a significant impact at both macro- and micro-economic levels. Past research focuses on determinants of fund flows and uses both micro and macro-level approaches. Under the micro-level approach, individual fund flows are studied and under the macro-level approach, aggregate flows in and out of the mutual fund industry have been studied (Alexakis, Niarchos, et al., 2005; Watson and Wickramanayake, 2012; Alexakis, Apostolos, et al., 2013)^{1,2,3}.

In this paper, we study the dynamics of fund flows at aggregate scheme category level across major categories (Large Cap, Large and Mid cap, Multi Cap, Mid Cap, and Small Cap) offered by open-ended equity mutual funds in India. Mutual fund flows on a monthly frequency have been analysed and it is noted that fund flows demonstrate characteristics of time series with the presence of serial correlation between the present value of scheme corpus and its own lagged values (lag length-1) (evident by Durbin Watson (DW) statistic ranging between 0.39-0.89). The study has been conducted on a very robust sample set of 142 open-ended equity schemes with at least 1100 observations per scheme category over 10 calendar year(s) from 2011-2020. Time series regression framework has been utilised to forecast fund flows at aggregate scheme category level as per Auto Regressive Integrated Moving Average (ARIMA) modelling. Relevant robustness checks have also been conducted to detect the presence of serial correlation in residuals as well as to verify the accuracy of forecasts.

Our study, to the best of our knowledge, is the first attempt to analyse mutual fund flows in a time series regression framework in the Indian context. We began the research with an attempt to understand determinants of fund flows from both scheme-specific and macro-economic factors perspectives. In the initial phase, panel-data analysis has been utilised to arrive at determinants of mutual fund flows. Macro factors namely Money supply (M3), the relative rate offered by Government securities, (Foreign Institutional Investor (FII) fund flows, and fund specific factors namely Net Asset Value (NAV), Expense Ratio, and Turnover ratio have been taken as independent variables to explain the dynamics of mutual fund flows. The results of regression analysis have been shared in Results Section 4. Our initial analysis results (evident in low R-Squared and near-zero DW Statistic) suffered from lack of accuracy and indicated the presence of serial positive auto-correlation of mutual

fund scheme corpus at time t with its one-month ($t-1$) lagged values. The results also failed the "feedback-trader" hypothesis (Warther, 1995)⁴ that "securities' returns affect mutual fund flows" (Kopsch et al., 2015)⁵. Results are consistent with studies of Fant (1999)⁶ and Zhang and Edwards (1998)⁷, who investigated the Investment Company Institute (ICI) data of monthly flows to US mutual funds and found no statistically significant relation between aggregate net monthly flows and subsequent market returns. This led us to study dynamism in fund flows in time series regression setting and results of the same were found robust to make forecasts at the scheme level for a lag length of 60 (5 years). There was no presence of auto-correlation in residuals and its variance being constant over time also excludes the presence of heteroscedasticity.

The rest of the paper is organised as follows: Section 2 deals with a review of existing literature and specifying the research problem, Section 3 explains sources of data, the methodology adopted, and research tools utilised. Section 4 details the results and findings and Section 5 concludes the study along with further scope for research.

2. Literature Review

Portfolio composition analysis and performance measurement with contemporaneous variables have been topics of considerable interest to financial researchers and analysts in India as well as globally. Persistence in mutual fund performance has been successfully explained through common factors in market returns and expense ratio under Carhart-Factor Model. Past research has also thrown light on the fact that performance persistence is not completely related to the stock-picking skill. Research on fund flows has largely been done from a performance measurement perspective, trying to understand the relationship between fund performance and inflows to fund (Chevalier and Ellison, 1997, Brbrovic, 2014)⁸. Existing literature has long recognized investors' response to fund performance and a robust, positive relationship between net flows and past performance has been well recorded (Ippolito, 1992; Sirri and Tufano, 1998)^{9,10}. An in-depth study entailing direct inquiry into patterns of inflows and outflows was conducted by Ivkovic et al. (2008)¹¹ and authors concluded that inflows are related only to "relative performance", suggesting that new flows chase the best performance as an objective while outflows are related only to "absolute" performance, a relevant bench-

mark for taxes. Existing literature also suggests a convex relationship between inflows and fund performance showing that good performance leads to high flows of new money while bad performance leads to only low redemptions (Sirri and Tufano, 1998; Geurcio and Tkac, 2002)¹². This asymmetry has been attributed often to the existence of poorly informed (unsophisticated) investors taking irrational investment decisions (Gruber, 2010)¹³.

Mutual fund flows have also been studied from various other perspectives. Some researchers have also utilised flow data to analyse different aspects of fund performance. Chen et al., (2010)¹⁴ studied fund flows connecting outflows to liquidity level of assets, finding results to be consistent with Perold and Salomon (1991)¹⁵. Research later has been extended to liquidity transformation in the asset management industry (Chernenko, and Sunderam, 2014)¹⁶. Ferreira et al., (2013)¹⁷ studied mutual funds data of 27 countries and found country-specific characteristics effective in explaining fund performance and resulting flows (Stein, 1997; Scharfstein and Stein, 2000; Johnson, 2004; Khanna and Yafeh, 2005; Gopalan et al., 2014)^{18,19,20,21,22}.

Several studies in the past have been done on determinants of mutual fund flows starting with an initial study by Martin Gruber (2010)¹³. The study answers a pertinent question as to why mutual funds and in particular active funds grow so fast, supported by empirical evidence on the rational behavior of investors in taking investment decisions. Fund flows have a significant impact at both micro- and macro-economic level (Ferson and Kim, 2012)²³. The macro approach utilises aggregate fund flows in and out of the mutual fund industry (Alexakis et al., 2005; Alexakis et al., 2013; Watson and Wickramanayake, 2012). Rakowski and Wang (2009)^{1,2,3,24} showed that past fund flows and returns, as well as the time of the month and fund characteristics, are important factors in explaining the level of daily fund flows.

There has been little research to analyse dynamism and volatility of mutual fund flows in the Indian context, this fact has been also emphasised by Oh and Parwada (2007), Khorana et al., (2005)^{25,26}. Our study makes an important contribution in this regard as we study dynamics of fund flows in a robust sample of 142 open-ended equity schemes with growth orientation across major scheme categories (Large Cap, Large and Small Cap, Multi Cap, Mid Cap, and Small Cap). We use the time series regression framework to analyse monthly net

flows data and make a forecast for chosen categories at the lag length of 60 using ARIMA modelling. We conclude the study with two important outcomes; *One*, on an average open-ended equity scheme in all categories under the sample (with a minor exception of few schemes under the Small Cap category), does not conform with the “Feedback Trader hypothesis”. Unlike the US, the Indian funds market is low in-depth in terms of volume, as also fund investments, particularly in the retail segment are done popularly as a disciplined investment (SIP route) than investments arising out of herd-mentality. Funds in the Small Cap category do reflect the “*chasing hot money*” tendency indicated by a statistically significant relationship with the previous day level of the market index. *Second*, funds across categories exhibit time series characteristics of mean reversion, presence of a trend and seasonal component, and structural breaks. Schemes in Mid Cap and Lower Cap categories, however, have more random component than schemes in Large Cap, Large and Small Cap, and Multi Cap which exhibit the presence of a strong trend component. Investors at the beginning of their investment journey can take advantage of the trend-investing skill of professional fund managers.

3. Research Methodology

This section details various sources of data consulted for conducting the study and methodology adopted for empirical investigation.

3.1 Period and Population

To conduct a time series analysis of mutual fund corpus, a period of 10 years from (January 2011 to January 2020) has been taken. The period of the study 2011-2020 coincides with a level playing field for mutual funds in Indian financial markets. The period of the study has been decided in consideration to the number of economic cycles, systematic events affecting equity markets, and data availability for respective variables. Index-linked and passively managed funds have been kept out of the purview of the study.

6846 open-ended schemes and 9315 close-ended schemes are existing across all categories – Income, Infrastructure Debt, Equity (excluding arbitrage funds), Arbitrage funds, Balanced, Liquid/Money market, Gilt, ELSS-Equity, GOLD ETF, Other ETFs, and Fund of funds

as of January 2020, ACE Mutual Fund Database. With the above-defined parameters, the scope of the present study is limited to a population set of 1022 Equity oriented schemes in the Indian Mutual Fund Industry.

3.2 Sampling

Sample mutual fund schemes have been selected based on the following criteria:

Research has been conducted on equity-based schemes existing as well as launched after January 01, 2011, to date. Open-ended schemes investing in an equity asset class with growth orientation across major categories have been considered as the sample set and analysis is conducted scheme-category-wise. Closed-ended funds have been kept out of the scope of study due to the data availability constraint. Our sample set, hence, constitutes 142 Open-Ended Market Cap schemes with growth orientation (major Market Cap categories being Large Cap, Large and Mid Cap, Multi Cap, Mid Cap, and Small Cap) out of a population set of 1022 Equity oriented schemes in Indian Mutual Funds. The sampling technique applied and sample drawn to carry out the study are mentioned in detail in Table 1 and Appendix 1.

3.3 Data Collection

Secondary data sources (both restricted and open access) have been utilised for carrying out research. ACE Mutual Fund is a dedicated database for mutual fund research covering all existing players and more than 16000 mutual fund schemes across all asset classes since inception. The data provided is free from survivorship bias and therefore, has been referred for the study. Open access databases used include Value Research and AMFI website for referencing to missing values.

3.4 Methodology

Time series analysis has been conducted on mutual fund flows' data extracted with monthly frequency. The R Software package has been utilised to conduct time series analysis and in identifying the model of best fit under ARIMA modelling. A total of 142 open-ended equity schemes with growth orientation investing across major market cap categories have been analysed under the study. A time period of 10 years from January 2011 till July 2020 across all schemes under the sample has given a robust database of more than 1100 monthly time series

observations under every scheme category (number of schemes per category x 12 months x 10 years).

The decision to conduct time series analysis and the reason behind it has been detailed well in Results and Findings section 4. The study was initiated to identify factors determining fund flows in equity-oriented open-ended mutual funds in India. We utilised a linear regression framework and then conducted data analysis by forming a balanced panel of different schemes for the period 2011-2020 and monthly frequency. Regression results indicated the presence of serial correlation in time series with the DW statistic being closer to zero (0.6-0.87). This was further confirmed by a statistically significant correlation between monthly corpus values and its own lagged values (Lagged at t-1). Collinearity diagnosis of other explanatory variables with lagged values of scheme corpus and statistically significant correlation values helped us in our decision of using a time series regression framework for the analysis. Requisite robustness checks of the ARIMA model so adopted were conducted along with heteroscedasticity test for residuals.

3.5 Variables

The study aims at the understanding determinant of mutual fund flows in equity-oriented open-ended funds in the Indian Mutual Fund market. Given below is a description of variables under the study:

- a. Dependent Variable: Monthly mutual fund flows
- b. Independent Variable: Lagged mutual fund flows (Lag-1)

3.6 Econometric Tools

Monthly mutual fund corpus data extracted from ACE Mutual Fund Database has been analysed in the R Software Package, widely used software for time-series regression. Decomposition of time series across scheme categories was conducted using R and trend patterns, as well as seasonality, were detected for further extraction of stationary time series. Augmented Dickey-Fuller Test (ADF Test 1979, 1981) was utilised to test for stationarity of time series data with the null hypothesis (H_0) being the presence of a unit root in time series. The lag length of time series data was decided based on minimum lag length criteria under Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). An appropriate time series model (AR, ARMA, MA, and ARIMA)

based on minimum values for AIC, BIC was selected across different scheme categories and used for further forecasting. Residuals of the model were then tested for the presence of any serial correlation, Auto-Correlation Function (ACF) and Partial-Auto Correlation Function (PACF) were utilised for the same. Plotting the histogram of residuals and super-imposition of normal curve along density helped us rule out any heteroscedasticity in the residuals. The accuracy of the model was verified through the calculation of the statistic Mean Percentage Error (MPE) and Root Mean Squared Error (RASE). Model of best fit has been utilised for forecasting time series with lag length 60 that is 5 years ahead.

Generalised ARIMA(p,d,q) model equation is as follows:

$$Y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

ARIMA models typically expressed like “ARIMA(p,d,q)”, where:

- P: number of preceding (“lagged”) Y values. This captures the “autoregressive” nature of ARIMA.
- D: number of times that the data have to be “differenced” to produce a stationary signal that has a constant mean over time. This captures the “integrated” nature of ARIMA.
- q: number of preceding/lagged values for the error term. This captures the “moving average” part of ARIMA.
- AR process of an order p can be written as,

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$
- MA process of order q can be written as,

$$y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Augmented Dickey-Fuller (ADF Test)

The ADF test is an augmented version of the Dickey-Fuller test equation to include a higher-order regressive process in the model. The test checks for the presence of unit root in time series. ADF test is the same as for the Dickey-Fuller test but it is applied to the model:

$$y_t = c + \beta_t + \alpha y_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \dots + \phi_p \Delta y_{t-p} + \epsilon_t$$

- $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t$, Where, : Presence of a unit root in time series

H_1 : Time series is stationary

$$AIC = -\frac{2}{N} * LL + 2 * \frac{k}{N}$$

$$BIC = -2 * LL + \log(N) * K$$

Lag-length appropriate for best-fit ARIMA model has been decided based on AIC and BIC for minimum lag length. AIC and BIC are two ways of scoring a model based on log-likelihood and complexity as illustrated below:

The model of best fit for specific scheme categories has thus been decided based on the lowest AIC and BIC values. Table 5 details the model of best-fit utilised for specific scheme categories. Measuring the strength of the trend and random component and a visual investigation of time series charts (in Annexure 1) indicates the presence of a strong trend component in monthly scheme corpus values of Large Cap and Large and Mid Cap funds, while in the case of Small Cap, Mid Cap and Multi Cap funds there is the presence of strong random component. The strength of the trend component is measured by:

$$F_t = \max\left[0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t - R_t)}\right]$$

$$(0 < F_t < 1)$$

Value of F_t nearing 0 indicates no trend and value of F_t nearing 1 indicates a strong trend

Where,

$$y_t = T_t + S_t + R_t$$

Time series of variable y at time t is an additive time series with a value of y_t being a summation of trend, seasonality, and a random component.

Box-Ljung Test

Box-Ljung Test (1979) is a diagnostic test of whether any group of autocorrelations of a time series is different from zero. It checks for “overall” randomness based on the number of lags and detects lack of fit in a time series model. The test statistic is:

$$Q_m = n(n+2) + \sum_{j=1}^m \frac{r_j^2}{n-j}$$

- Where, r_j = the accumulated sample autocorrelations, m = the time lag.

4. Results and Findings

Table 1 lists out descriptive statistics of a sample under the study. A total of 142 open-ended equity growth oriented

Table 1. Descriptive Statistics

	Population	Sample	Mean Corpus	Std. Dev Corpus	Skewness	Std. Error
Large Cap Funds	233	30	2707.0557	5035.8005	2.365	.043
Large and Mid Cap Funds	149	28	1219.7746	1293.7649	1.563	.054
Multi Cap Funds	261	35	2683.1675	4652.6106	2.648	.049
Mid Cap Funds	181	26	1848.5322	3220.6236	3.636	.051
Small Cap funds	198	23	1365.6268	2065.7705	2.084	.073

Source: Authors' calculations

Table 2. Linear Regression Model Summary

Model	R	Adj. R-Square	R-Square	Std. Error	DW
Large Cap Funds	.666a	.444	.442	3763.781	.076
Large and Mid Cap Funds	.789a	.622	.621	793.259	.089
Multi Cap Funds	.590a	.348	.347	3765.908	.037
Mid Cap Funds	.446a	.199	.198	2886.923	.039
Small Cap Funds	.522a	.272	.269	1760.934	.083

Note: a. Predictors: (Constant), NAV, G Sec 10 year Rate, Turnover ratio, Expense Ratio, M3 Money Supply, Fund Age, Forex Assets.

b. Dependent Variable: Monthly Corpus.

Source: Authors' calculations

schemes across major marketcap categories have been considered under the study. Funds across categories have a mean corpus of more than 1000 crores. Variation in fund corpus is highest in the Large Cap and Multi Cap funds category as is evident from high standard deviation statistic. Chebyshev Inequality has been utilised to remove outliers in the data¹. Missing values were generated in Stata Software. Mid Cap Funds exhibit higher skewness with majority scheme corpus values lower than mean corpus value.

Initial investigation focused on finding determinants of mutual fund corpus. We intend to test the information-response hypothesis, which states significant co-movement between macro factors (forecast of economic activity) and mutual fund flows. Both scheme-specific and macro factors were considered as explanatory variables. Poor interaction between macro factors and fund flows does

not validate a statistically significant relationship between new information about the real economy and mutual fund flows. Panel data regression results are detailed in Table 2 with Large Cap and Large and Mid Cap funds having statistically significant adjusted R-Squared Values (0.442 and 0.621). DW test statistic across all scheme categories shows values ranging between 0.37 and 0.89. DW test statistic is used to detect the presence of autocorrelation at Lag 1 in the residuals from a regression analysis. Results show values substantially lesser than 2 which indicate significant positive autocorrelation.

A step further in the elimination process, we ran collinearity diagnostics among variables with the addition of lagged values of the monthly corpus (lag length 1) as a new explanatory variable (Table 3). Pearson coefficient of correlation with values higher than 0.95 across all scheme categories indicates a very high degree of correlation between scheme corpus and its own lagged values, this further confirms the results of the DW test. Across categories, the correlation of lagged monthly corpus values with other explanatory variables was also found to be statistically significant even at a 1% confidence level. This helped us adopt time-series regression analysis

¹In probability theory, Chebyshev's inequality (also called the Bienaymé–Chebyshev inequality) guarantees that, for a wide class of probability distributions, no more than a certain fraction of values can be more than a certain distance from the mean. Specifically, no more than $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean. The theorem is widely used in the detection of significant outliers in large data sets.

Table 3. Pearson Correlation between Variables

	Monthly Corpus	Forex Assets	G Sec 10 year Rate	M3 Money Supply	Turnover Ratio	Expense Ratio	NAV	Fund age	Lagged MC (Lag1)
Large Cap Funds	1	.279**	-.169**	.283**	-.080**	-.530**	.449**	.342**	.985**
	.985**	.264**	-.160**	.270**	-.081**	-.521**	.452**	.335**	1
Large and Mid Cap funds	1	.513**	-.291**	.519**	-.114**	-.569**	.628**	.484**	.977**
	.977**	.489**	-.280**	.499**	-.113**	-.552**	.625**	.476**	1
Multi Cap Funds	1	.270**	-.160**	.272**	-.021	-.360**	.572**	.308**	.982**
	0.982	.255**	-.151**	0.259	-.021	-.350**	.569**	.302**	1
Mid Cap Funds	1	.333**	-.253**	.332**	-.031	-.497**	.388**	.204**	.980**
	.980**	.318**	-.242**	.321**	-.029	-.488**	.381**	.197**	1
Small Cap Funds	1	.527**	-.318**	.535**	-.135**	-.434**	.298**	.071*	.974**
	.974**	.503**	-.307**	.514**	-.132**	-.417**	.294**	.060*	1

Note: * Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Source: Authors' calculations

Table 4. Augmented Dickey-Fuller Test

Fund Name	N	DF Stat.	Lag-Order	P-Value	ADF Test Results
Large Cap Funds	3187	-5.6623	4	0.01	Reject Ho
Large and Mid Cap	2511	-3.1822	4	0.02133	Reject Ho
Multi Cap Funds	1995	-5.1479	4	0.01	Reject Ho
Mid Cap Funds	2257	-6.0762	4	0.01	Reject Ho
Small Cap Funds	1109	-5.2783	4	0.01	Reject Ho

Note: H_0 : A unit root is present in a time series sample.

H_1 : Time series exhibit stationarity.

Source: Authors' calculations

methodology in forecasting scheme corpus values with statistically significant results.

4.1 ADF Test Results

ADF test is more appropriate for a larger and more complicated set of time series models. With a period of 10 years and at least 1100 observations per scheme category²

²Schemes belonging to a specific investment orientation are categorised under one category, all the schemes investing in Large Cap stocks, for instance, belong to the Large Cap category.

(number of schemes in a category x 12 months x 10 years), we have utilised ADF test to detect the presence of a unit root in a time series sample. ADF test statistic with p-value < 0.05 across all scheme categories rejects the Null Hypothesis with the result that time series is stationary with characteristics of mean reversion and constant variance, this further verifies that an appropriate model can be used for timeseries forecasting (Table 4). Timeseries graphs under each category are illustrated in Annexure I.

Table 5. Model of Best Fit*

	Model Spec.	Coefficients					Model Specifications Test Results				
		ar1	ar2	ma1	ma2	ma3	sma1	Log-Likelyhood	AIC	AICC	BIC
Large Cap Funds	ARIMA (2, 0, 1)	1.3439	-0.6434	-0.7471				-831.49	1670.98	1671.41	1681.28
Large and Small Cap Funds	ARIMA (0, 0, 3) (0, 0, 1)			0.9657	0.1125	-0.1625	-0.737	-579.41	1168.82	1169.48	1181.69
Multi Cap Funds	ARIMA (0, 0, 1)			0.554			-0.6258	-628.41	1262.82	1263.08	1270.55
Mid Cap Funds	ARIMA (0, 0, 1)			0.344				-588.4	1122.81	1123.07	1130.53
Small Cap Funds	ARIMA (0, 0, 1)			0.622				-715.81	1435.62	1435.74	1440.77

Note: *R package - Auto ARIMA Function is used.

Source: Authors' own calculations

Table 6. Ljung-Box Test Results

	X-Squared	df	p-Value	Test Results
Large Cap Funds	1.2605	20	1	Fail to reject Null
Large and Small Cap Funds	18.133	20	5.79E-01	Fail to reject Null
Multi Cap Funds	23.925	20	0.2457	Fail to reject Null
Mid Cap Funds	14.285	20	0.8157	Fail to reject Null
Small Cap Funds	3.274	20	1	Fail to reject Null

Note: H_0 : Model does not show a lack of fit.

H_1 : Model does show a lack of fit and residuals are not independent.

Source: Authors' calculations

Table 7. Accuracy of the Model

	Mean Error (M.E.)	RMSE	MAE	MPE	MAPE	MASE	ACF1
Large Cap funds	-16.7055	72.039	67	43.0226	35.774	0.665	-0.1
Large and Mid Cap Funds	3.026829	89.98111	67.55	29.94375	15.611	0.50156	-0.034
Multi Cap Funds	-14.9394	152.4571	115	-683.100	925.03	0.5352	0.058
Mid Cap funds	-4.2105	75.69076	58	43.398	32.512	0.5299	0.013
Small Cap Funds	-19.1072	386.838	175	55.18671	83.4641	0.6779	-0.01

Source: Author's calculations

Table 8. Feedback-Trader Model

Model	R	R-Square	Adjusted R-Square	Std. Error of the Estimate	DW Stat.
Large Cap Funds	.278a	.077	.077	4837.8632	.029
Large and Mid Cap Funds	.516a	.267	.266	1108.1571	.047
Multi Cap Funds	.269a	.072	.072	4482.1221	.034
Mid Cap Funds	.328a	.108	.107	3042.6446	.038
Small Cap Funds	.476a	.226	.226	1817.9002	.049

Note: a. Predictors: (Constant), BSE Large Cap, BSE All Cap, BSE Mid Cap, BSE Small Cap.

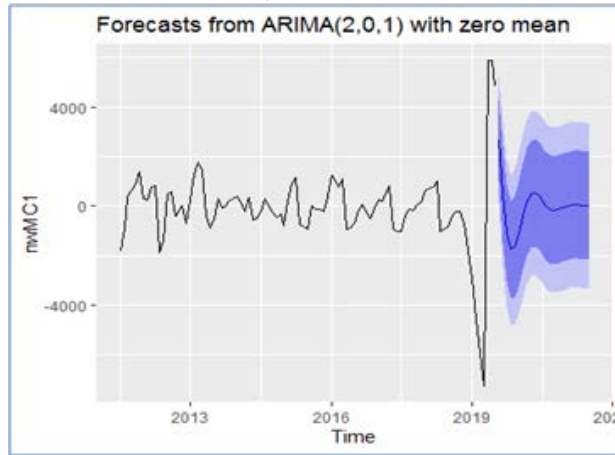
b. Dependent Variable: Monthly Corpus.

H_0 : Index returns provide positive feedback for investment in mutual fund.

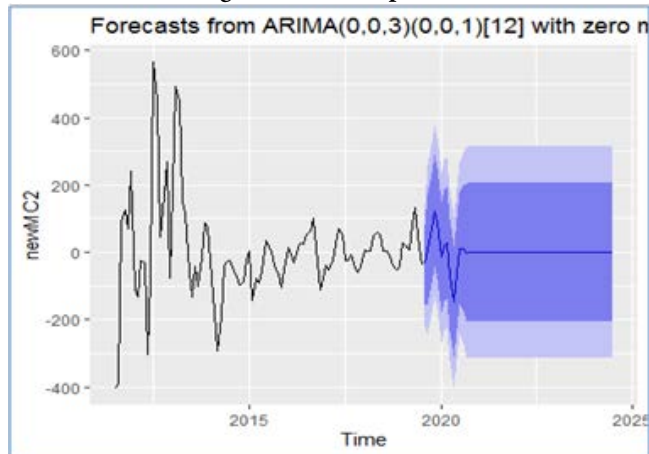
H_1 : Index returns do not provide positive feedback for investment in mutual fund.

Source: Authors' calculations

Large Cap Funds



Large and Small Cap Funds



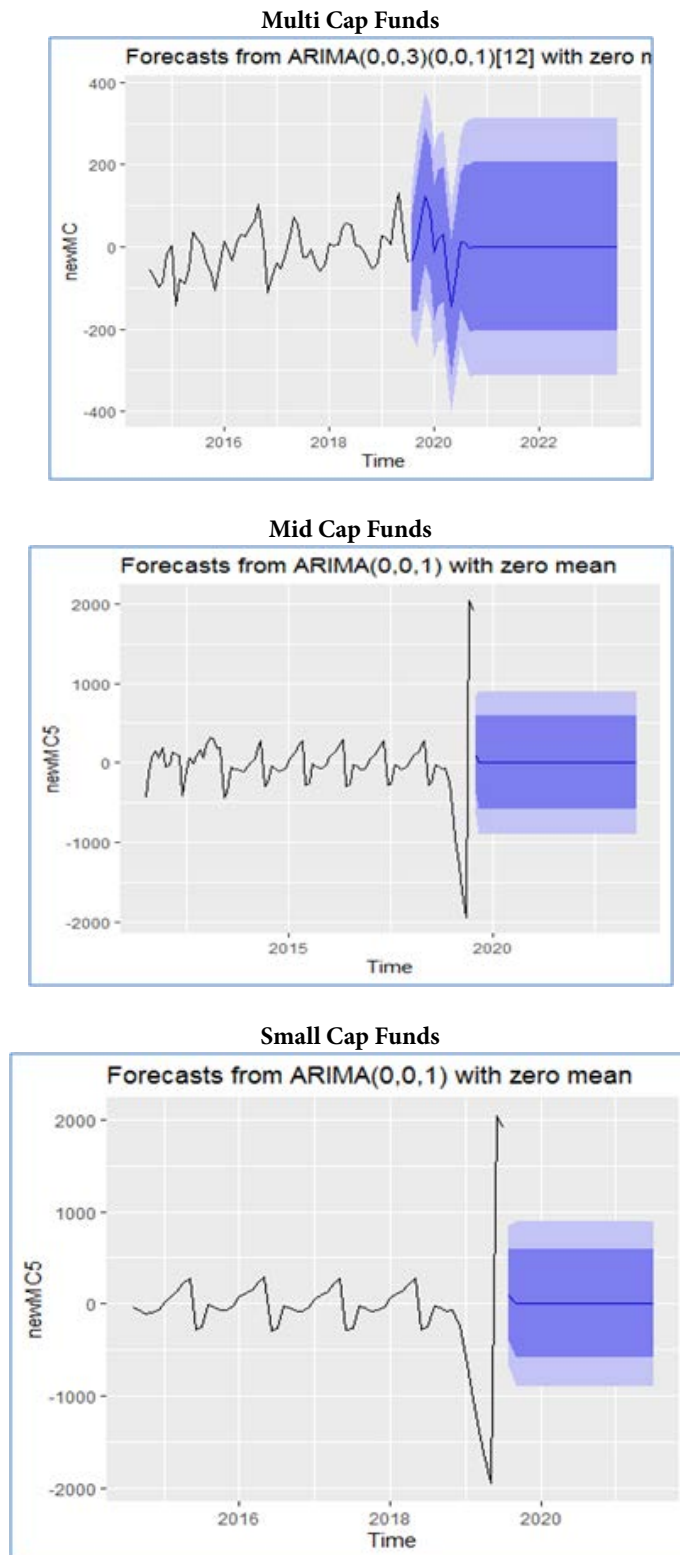


Figure 1. Forecast from ARIMA Model

The coefficients of different types of ARIMA Model fitted to specific schemes are also illustrated in Table 5 & Figure 1. Model for Large Cap funds is an ARMA(2,1) model and the rest of the models are essentially MA(1) models.

Ljung-Box test has been applied to the residuals of a fitted ARIMA model with null-hypothesis being residuals from the ARIMA model have no autocorrelation. A statistically significant p-value (>0.05) fails to reject null-hypothesis of no autocorrelation in residuals and that model can be used for forecasting timeseries (Table 6). Fitted ARIMA models have been utilised for making forecasts with a lag length of 60 (5 years). Forecast charts have been shared at the end of this section.

Results of the ARIMA Model have been further verified by running a test of the accuracy of forecasts (Table 7). Mean Absolute Percentage Error (MAPE) ranges between values 15.61-83.46 across categories. MAPE is free from deficiencies of MPE which suffers from cancelling out of negative and positive values. Forecasts have been made for the lag length of 60 and given the choppiness of financial time series data and a significantly longer period of forecast, the test of accuracy helps us in considering the model for the forecast.

The feedback trading hypothesis has been well-tested in developed markets and positive feedback trading has been proven to exist in developed as well as emerging markets (Lakonishkok et al., 1992, Bohl & Siklos, 2008; Dean & Faff, 2011; Sheih et al., 2011)^{27,28,29,30}. Feedback trading refers to a behavioral pattern, often labelled as herd mentality exhibited on part of investors to invest when markets are performing well. When a cycle of positive feedback continues for too long, investor enthusiasm can lead to irrational exuberance, which can precipitate asset bubbles that eventually lead to a market crash. We have tested the positive feedback trader hypothesis in the Indian mutual funds market. We tried to explain scheme corpus values with relevant market index (as per scheme category, e.g., BSE Small Cap index in case of Small Cap funds³, Table 8). The adjusted R-Square value (0.072-0.266) across schemes contributes lesser to the explanatory power of the regression model. The DW statistic is closer to 0 across categories (max. value 0.49). This again highlights the presence of serial correlation in scheme corpus data. Lower R-squared value across

categories (even in the case of liquid Large Cap category) indicates rejection of positive feedback trading in the Indian mutual funds market which also aligns with findings of Jank (2012)³¹.

5. Conclusion

This paper aims to study the determinants of mutual fund flows in the Indian context. Several predictors from both micro and macro perspectives were included in the analysis and two well-established and tested hypotheses – feedback trader and information-response hypothesis were included in the analysis. These hypotheses have been tested and found to hold good in developed and some emerging markets. Schemes offered across categories in the sample set failed the positive feedback trading hypothesis. The study has been fairly investigative and the presence of serial correlation as established by DW test statistic led us to the adoption of time series regression framework in analyzing mutual fund flows across scheme categories. We found the time series of mutual fund flows to be additive and the ARIMA model of best fit (as per scheme category) has been utilised to make forecasts. The ACF and PACF plot of the residuals indicates homoscedasticity or no presence of serial correlation in residuals. The forecasts have been made for the next 5 years with checks of accuracy. Our study helped us to arrive at two important conclusions. *One*, unlike studies conducted in many developed and some emerging economies, fund flows data collected on a monthly frequency for Indian open-ended growth oriented schemes reject the positive feedback trading hypothesis. *Second*, fund flows exhibit time series characteristics of mean-reversion, presence of trend and seasonality component, and structural breaks. Funds in Large Cap, Large and Mid Cap, and Multi Cap category exhibit strong trend component while funds in Mid Cap and Small Cap exhibit strong random component. Market timing and stock-selectivity are the most potent tools available in fund manager's toolkit. The superior performance generated by the fund manager based on these skills can be adequately measured by alpha. Results of the study can be utilised by investors to invest in Large Cap, Large and Mid Cap, and Multi Cap category schemes that offer consistent alpha by engaging in trend investing during market upcycle. Mid Cap and Small Cap funds invest in quality under-valued stocks, offering opportunities for significant gains. Investors, however, in the early stages of their investment journey may steer

³Other benchmarks utilised – BSE Large Cap for Large Cap, BSE 250 Large Mid Cap for Large and Mid Cap schemes, BSE Mid Cap for Mid Cap schemes.

clear of these funds as Mid Cap and Small Cap stocks are subject to undercurrents generated by market volatility. The impact of mutual funds flows on market returns and an increasing proportion of investment through direct route gives us further scope for the research.

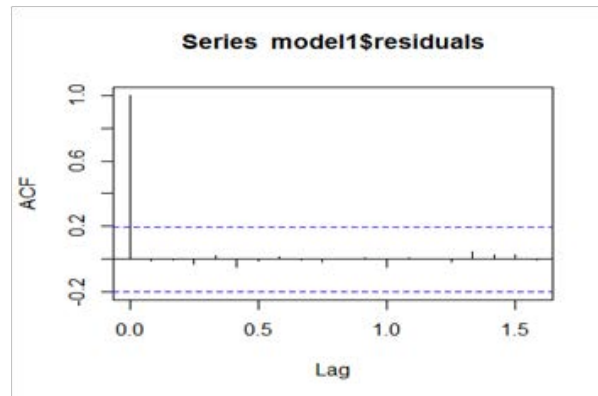
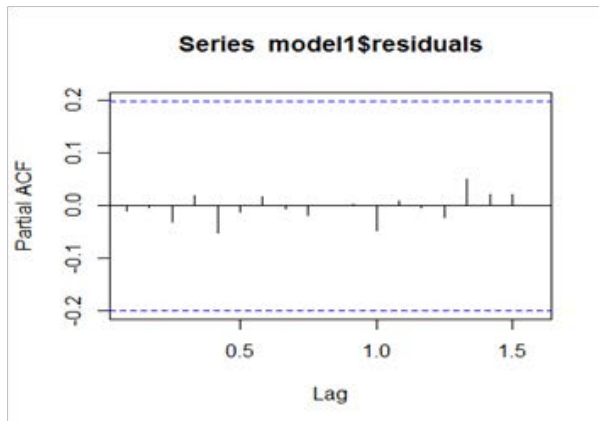
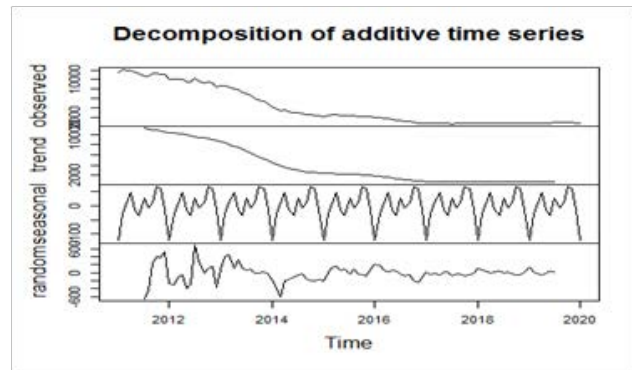
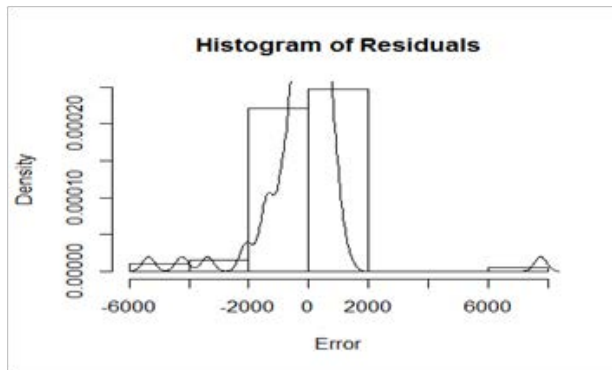
6. References

- Alexakis, C., Dasilas, A., and Grose, C. Asymmetric dynamic relations between stock prices and mutual fund units in Japan. An application of hidden cointegration technique. *International Review of Financial Analysis*. 2013; 28:1–8.
- Watson, J., and Wickramanayake, J. (2012). The relationship between aggregate managed fund flows and share market returns in Australia. *Journal of International Financial Markets, Institutions and Money*. 2012; 22(3):451–472.
- Alexakis, C., Niarchos, N., Patra, T., and Poshakwale, S. The dynamics between stock returns and mutual fund flows: empirical evidence from the Greek market. *International Review of Financial Analysis*. 2005; 14(5):559–569.
- Warther, V. A. Aggregate mutual fund flows and security returns. *Journal of Financial Economics*. 1995; 39(2–3):209–235.
- Kopsch, F., Song, H. S., and Wilhelmsson, M. Determinants of mutual fund flows. *Managerial Finance*. 2015; 41(1):10–25.
- Franklin Fant, L. Investment behavior of mutual fund shareholders: The evidence from aggregate fund flows. *Journal of Financial Markets*. 1999; 2(4):391–402.
- Zhang, X., and Edwards, F. R. Mutual Funds and Stock and Bond Market Stability. *SSRN Electronic Journal*. Published. 1998.
- Chevalier, J., and Ellison, G. Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*. 1997; 105(6):1167–1200.
- Ippolito, R. A. Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry. *The Journal of Law and Economics*, 1992; 35(1):45–70. <https://doi.org/10.1086/467244>
- Sirri, E. R., and Tufano, P. Costly Search and Mutual Fund Flows. *The Journal of Finance*. 1998; 53(5):1589–1622.
- Ivković, Z., Sialm, C., and Weisbenner, S. Portfolio Concentration and the Performance of Individual Investors. *Journal of Financial and Quantitative Analysis*. 2008; 43(3):613–655.
- Guercio, D. D., and Tkac, P. A. The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. *The Journal of Financial and Quantitative Analysis*. 2002; 37(4):523.
- Gruber, M. J. Another Puzzle: The Growth in Actively Managed Mutual Funds. *Investments and Portfolio Performance*. 2010; 117–144
- Chen, Y., Ferson, W., and Peters, H. Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics*. 2010; 98(1):72–89.
- Perold, A. F., & Salomon, R. S. The Right Amount of Assets under Management. *Financial Analysts Journal*. 1991; 47(3):31–39.
- Chernenko, S., and Sunderam, A. Frictions in Shadow Banking: Evidence from the Lending Behavior of Money Market Mutual Funds. *Review of Financial Studies*. 2010; 27(6):1717–1750.
- Ferreira, M. A., Keswani, A., Miguel, A. F., and Ramos, S. B. The Determinants of Mutual Fund Performance: A Cross-Country Study*. *Review of Finance*. 2013; 17(2):483–525.
- Stein, J. C. Internal Capital Markets and the Competition for Corporate Resources. *The Journal of Finance*. 1997; 52(1):111–133.
- Scharfstein, D. S., and Stein, J. C. The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient Investment. *The Journal of Finance*. 2000; 55(6):2537–2564.
- Johnson, W. T. Predictable Investment Horizons and Wealth Transfers among Mutual Fund Shareholders. *The Journal of Finance*. 2004; 59(5):1979–2012.
- Khanna, T., and Yafeh, Y. Business Groups and Risk Sharing around the World. *The Journal of Business*. 2005; 78(1):301–340.
- Gopalan, R., Nanda, V., and Seru, A. (2014). Internal Capital Market and Dividend Policies: Evidence from Business Groups. *Review of Financial Studies*. 2014; 27(4): 1102–1142.
- Ferson, W. E., and Kim, M. S. The factor structure of mutual fund flows. *International Journal of Portfolio Analysis and Management*. 2012; 1(2):112.
- Rakowski, D., and Wang, X. The dynamics of short-term mutual fund flows and returns: A time-series and cross-sectional investigation. *Journal of Banking & Finance*. 2009; 33(11):2102–2109.
- Oh, N. Y., and Parwada, J. T. Relations between mutual fund flows and stock market returns in Korea. *Journal of International Financial Markets, Institutions and Money*. 2007; 17(2):140–151.
- Khorana, A., Servaes, H., and Tufano, P. Explaining the size of the mutual fund industry around the world. *Journal of Financial Economics*. 2005; 78(1):145–185.
- Lakonishok, J., Shleifer, A., and Vishny, R. W. The impact of institutional trading on stock prices. *Journal of Financial Economics*, 1992; 32(1):23–43.
- Bohl, M. T., and Siklos, P. L. Empirical evidence on feedback trading in mature and emerging stock markets. *Applied Financial Economics*. 2008; 18(17):1379–1389.

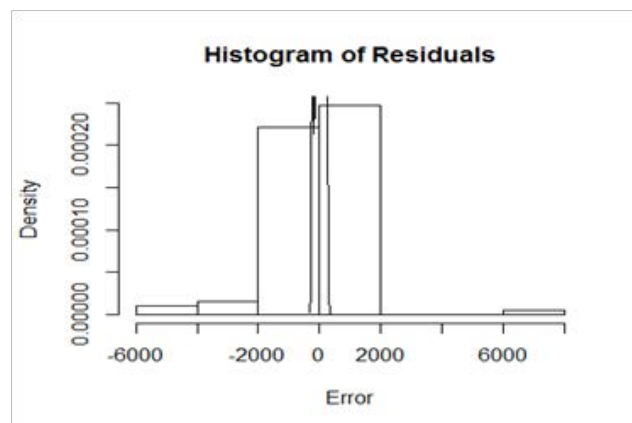
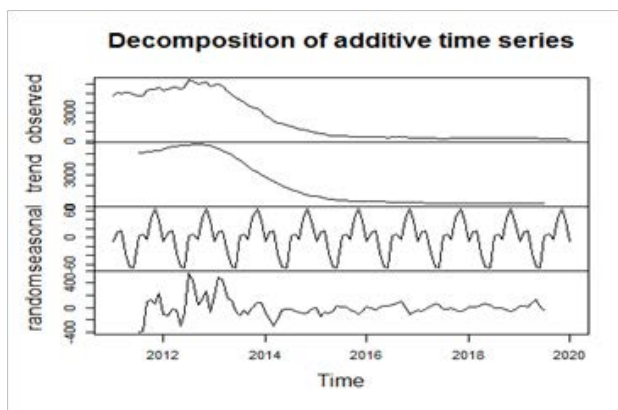
29. Dean, W. G., and Faff, R. Feedback trading and the behavioural ICAPM: multivariate evidence across international equity and bond markets. *Applied Financial Economics*. 2011; 21(22):1665–1678.
30. Shieh, M. F., Yang, T. Y., Yang, Y. T., and Lee, J. S. Evidence of herding and positive feedback trading for mutual funds in emerging Asian countries. *Quantitative Finance*. 2011; 11(3):423–435.
31. Jank, S. Mutual fund flows, expected returns, and the real economy. *Journal of Banking & Finance*. 2012; 36(11):3060–3070.
32. Chebyshev's inequality - Wikipedia

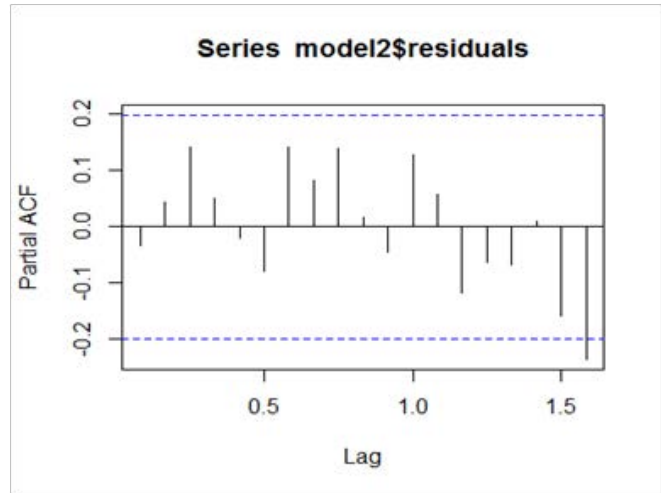
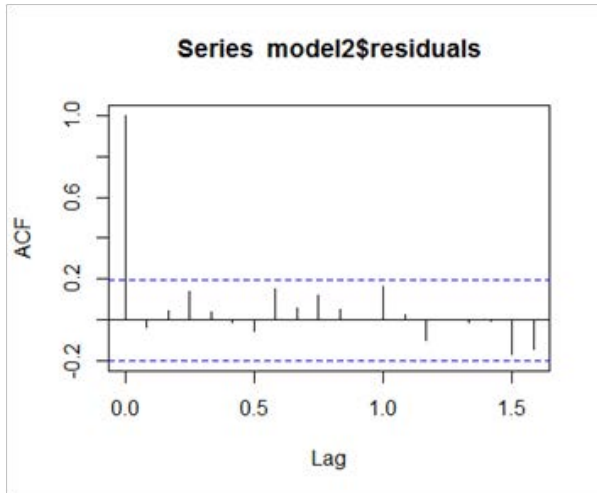
7. ANNEXURE I

Large Cap Funds

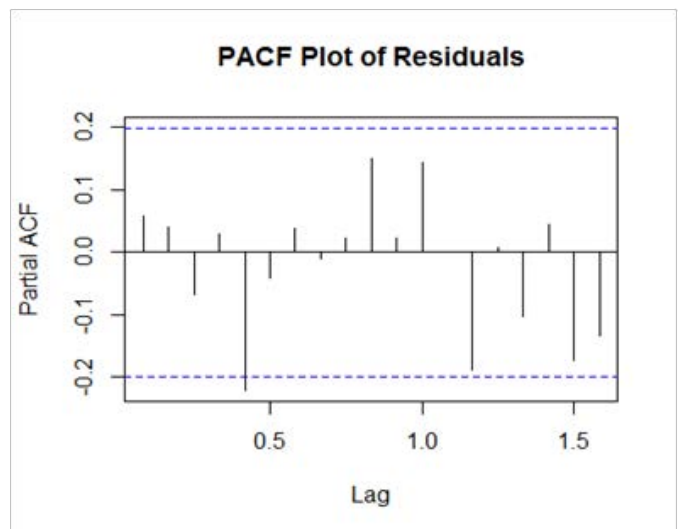
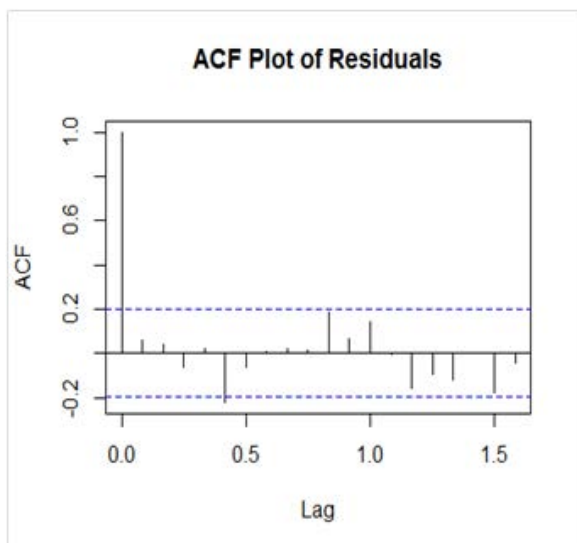
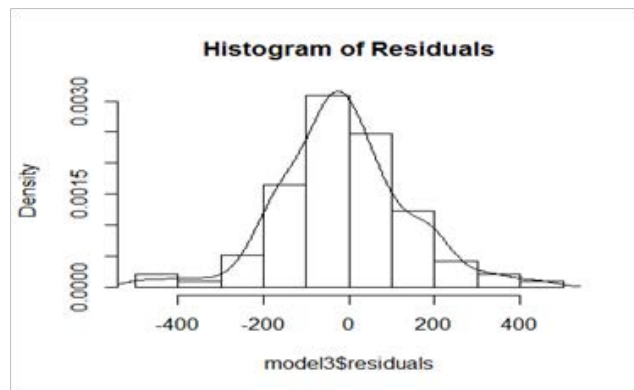
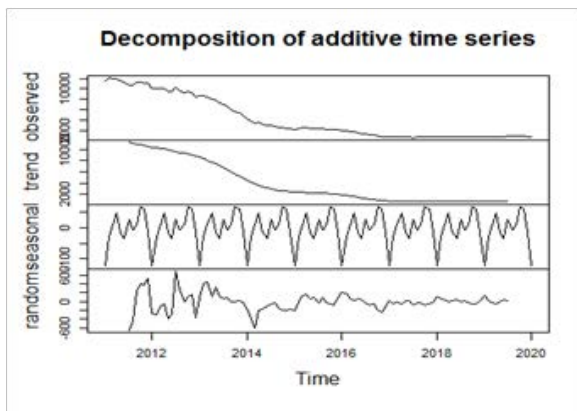


Large and Small Cap Funds

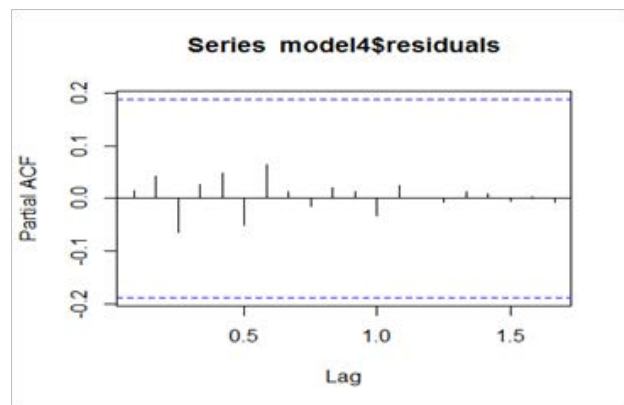
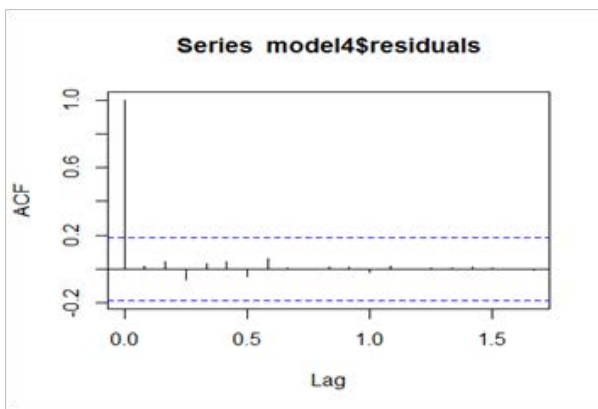
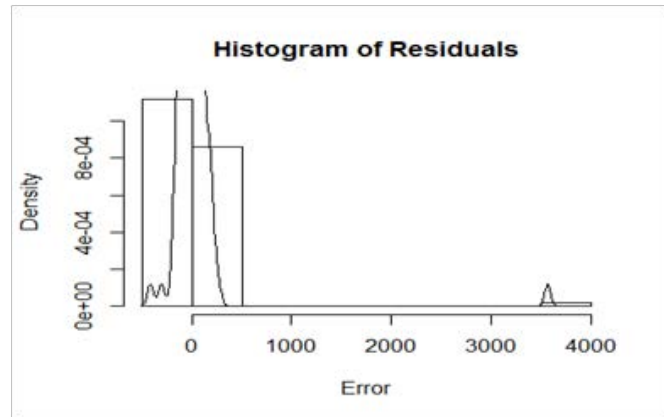
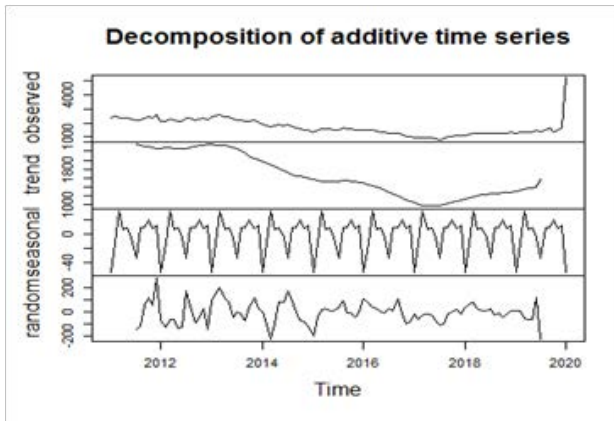




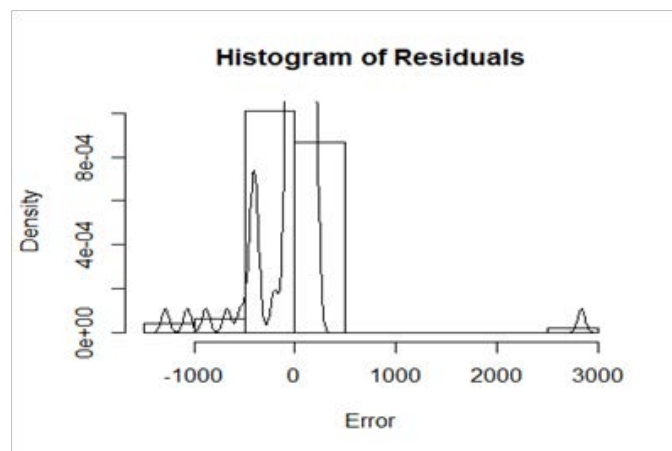
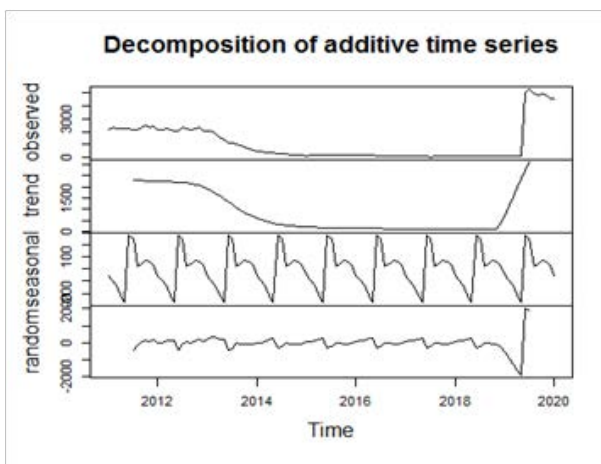
Multi Cap Funds

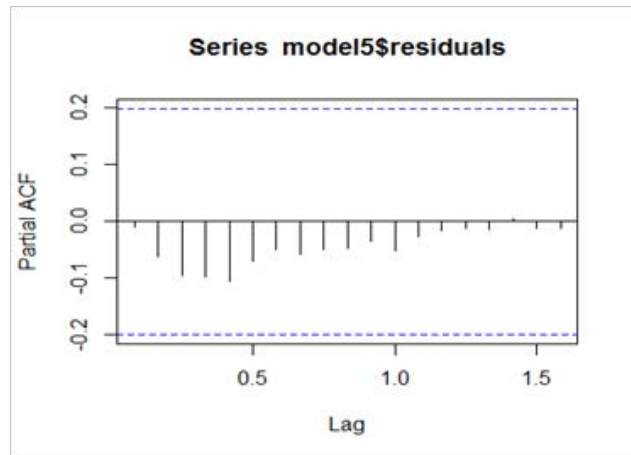
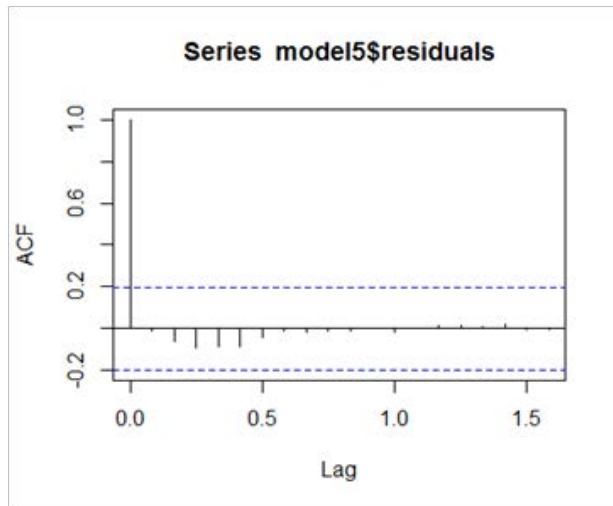


Mid Cap Funds



Small Cap Funds





ANNEXURE II

Sample mutual fund schemes have been selected based on six levels of filtration as detailed below. We arrive at a final sample of 142 open-ended equity schemes which

are standard plans with growth orientation across major marketcap categories:

Level	Filtration Criteria	Population Set	Sample Chosen	No. of Schemes
I	On basis of the asset class	Equity, Debt, Commodity, Hybrid, and Others	Equity asset class	2357
II	A category within the asset class	Contra, Dividend yield, ETFs, ELSS, Focused funds, Indexed funds, Market Cap, Sector funds, Thematic, Value funds	Market Cap category	959
III	Sub-category	Flexi Cap ⁴ (w.e.f. Jan 2021), Large Cap, Large and Mid Cap, Mid Cap, Small Cap, Multi Cap	Large Cap, Large and Mid Cap, Mid Cap, Small Cap, Multi Cap	830
IV	Type of scheme	Open-ended, Close-ended, Interval	Open-ended	654
V	Type of plan	Institutional, Regular, Super institutional, Institutional plus, Direct, Standard, Suspended	Standard plan	325
VI	Option	Bonus, Growth, Dividend	Growth	142

⁴Flexi Cap category has been recently added w.e.f. Jan 2021 to market cap category post-revised SEBI regulations for marketcap schemes

Sample Set as mentioned in Table 1, hence consists of following schemes and number of observations under each sample are mentioned in detail as under:

	Population	Schemes in Sample	No. of Years	No. of Months	No. of Observation
Large Cap Funds	233	30	9.5 ⁵	12	3187
Large and Small Cap Funds	149	28	9.5	12	2511
Multi Cap Funds	261	35	9.5	12	1995
Mid Cap Funds	181	26	9.5	12	2257
Small Cap funds	198	23	9.5	12	1109 ⁶
Total	1022	142 ⁷			

⁵The duration of the study is January 2011 till July 2020, approximately 9.5 years.

⁶The number of observations is not an exact multiple of (number of schemes*months*year) as some schemes have a later date of launch.

⁷Sample set chosen and filtration criteria have been mentioned in detail above. Our sample hence consists of at least 1100 observations per scheme category.