

Master in Quantitative Economics and Finance

**Factor models for optimized**

**option trading strategies**

Dimitrakopoulos Konstantinos Ntario

13-610-027

supervised by

Prof. Ph.D. Juan-Pablo Ortega

May 2017

Acknowledgments

**Table of Contents**

1. Introduction 8

1.1. Motivation 8

1.2. Literature Review 9

1.3. Research Objective 10

2. Data Acquisition - Description 12

2.1. Positioning Data 12

2.2. Option Data 14

3. Methodology 15

3.1. Data Preparation 15

3.2. Data Features 16

3.3. Investment Framework 19

3.4. BASE Model 21

3.5. AR Model 22

3.6. GARCH Model 25

3.7. Asset Selection 27

4. Results 28

4.1. Overview 28

4.2. BASE Model 29

4.3. AR Model 30

4.4. GARCH Model 31

4.5. Portfolio 32

4.5.1. Biases 33

4.5.2. Preliminary Results 34

4.5.3. Portfolio vs. Model Benchmarks 36

4.5.4. Portfolio vs. Market Benchmarks 38

5. Conclusions 40

6. Bibliography 43

7. Appendix 46

7.1. CFTC Current Reporting Levels 46

7.2. Performance Metrics 46

7.3. US 1-Year Bond Yield Data / Calculation 47

7.4. Market Benchmarks 47

**Table of Figures**

Figure 2‑1 – CoT Report Format 12

Figure 2‑2 – TFF Report Format 13

Figure 3‑1 – Positioning and Returns for Asset Managers in USDJPY 15

Figure 3‑2 – Probability Density Function for USDJPY (AM), USDCAD (De), AUDUSD (Com), XAUUSD (NonComm) 16

Figure 3‑3 – ACF / PACF on Positioning Returns of Asset Managers / Dealers in USDJPY / EURUSD 17

Figure 3‑4 ACF / PACF on Positioning Returns of Commercials / Non-Commercials in GBPUSD / USDCHF 18

Figure 3‑5 – Option Pricing Plots / Parameters 20

Figure 3‑6 – Duration of Options under Proposed Investment Process 20

Figure 3‑7 – BASE Model Signals on USDCAD using Asset Managers’ Positioning 21

Figure 3‑8 - Positioning Returns and AR(1) Fit for Dealers in NZDUSD 22

Figure 3‑9 - Positioning Filter on USDJPY for Dealers’ Positioning 23

Figure 3‑10 – Volatility and GARCH Fit for Asset Managers in AUDUSD 26

Figure 4‑1 – BASE Model – Cum. Returns (per asset) 29

Figure 4‑2 – AR Model – Cum. Returns (per asset) 30

Figure 4‑3 - GARCH Model – Cum. Returns (per asset) 31

Figure 4‑4 – LAM Portfolio – Cum. Returns 35

Figure 4‑5 – LAM Portfolio – Returns Distribution / Seasonality 35

Figure 4‑6 – LAM Portfolio vs. Model Benchmarks – Cum. Returns 37

Figure 4‑7 – Model Benchmarks – Returns Distributions 37

Figure 4‑8 – LAM Portfolio vs. Market Benchmarks – Cum. Returns & Correlations 39

Figure 4‑9 – Market Benchmarks – Returns Distributions 39

**List of Tables**

Table 3‑1 – Augmented Dickey-Fuller test for a random selection of assets 17

Table 4‑1 – BASE Model Performance Metrics (weekly basis) 29

Table 4‑2 – AR Model Performance Metrics (weekly basis) 30

Table 4‑3 - GARCH Model Performance Metrics (weekly basis) 31

Table 4‑4 –LAM Portfolio Performance Metrics (weekly basis) 35

Table 4‑5 – LAM Portfolio Asset Selection Information 35

Table 4‑6 - LAM Portfolio vs. Model Benchmarks Performance Metrics (weekly basis) 36

Table 4‑7 – LAM Portfolio vs. Market Benchmarks Performance Metrics (weekly basis) 38

**Abbreviations**

CFTC - U.S. Commodities Futures Trading Commission

TFF - Traders in Financial Futures

CoT - Commitment of Traders

ARIMA - Autoregressive Integrated Moving Average

ARMA - Autoregressive Moving Average

AR - Autoregressive Model

GARCH - Generalized Autoregressive Conditional Heteroscedasticity

ADF - Augmented Dickey-Fuller

ACF - Autocorrelation Function

PACF - Partial Autocorrelation Function

WPL - Weekly Profit Loss Function

NP - Net Positioning

NP% - Net Positioning Returns (week-over-week)

**Disclaimer**

# Introduction

## Motivation

In today’s computer-driven world, investment professionals have access to a vast array of sources from which they derive informed investment decisions. The later come in two basic forms: discretionary trades and systematic trades. The first are based on intuition, experience and research, while the later on mechanical setups that generate trading signals. Although distinctly different, both methods aim to implement one, or more, forms of analysis to determine market direction. Price prediction has been an area of ongoing research for a long time, and, since the late 1990’s, the construction and use of market sentiment indicators for empirical analysis has also become mainstream.

The price of an asset is determined by the buying / selling activity undertaken by the market participants at any point in time. It is therefore of interest to determine if insights in future price movements can be obtained from this activity. A detailed analysis of such data could, among others, reveal how specific participants view market fundamentals and their perception of future conditions. Indeed, while technical indicators, news and research represent the cornerstone of the investment pyramid, professionals have always been interested on the positioning of their likes in the markets. As a simplification, knowing that an established and reputable investor is on the same side of a trade can provide, even to a small degree, confidence. It is therefore no wonder that an emerging method of tracking price behaviour is to collect and analyse the market participants’ activity.

One of the principal challenges researchers face in this area is the limited amount of data available. Since investor performances at the trade level are scarcely available to the public, research has been exhausted[[1]](#footnote-1) on asset classes where this problem does not apply (e.g. mutual funds[[2]](#footnote-2)). Interestingly, a set of high quality figures that, although carefully monitored by market professionals, appear to have remained largely untapped by the research community are the U.S. futures positioning databases. These U.S.-wide data are regularly updated by the *U.S. Commodities Futures Trading Commission* (CFTC) in the *Commitments of Traders* (COT) and *Traders in Financial Futures* (TFF) reports, and provide insights on the aggregate positioning of investor sub-categories. Combining this information with market data can provide a novel understanding of market dynamics that are pertinent to academic research as well. In the next section, relevant literature across the use of sentiment indicators, such as market positioning, to forecast future prices is presented.

## Literature Review

Studies supporting the semi-strong *Efficient Market Hypothesis* (EMH) have been around since the 1990’s. (Siegel, 1992) found correlations between the shift in investors’ sentiment indicators and stock price movements before the market crash in October 1987[[3]](#footnote-3), while (Brunnermeier & Nagel, 2004) determined that in another stock market collapse, the technology bubble burst in March 2000, hedge funds had in a short period of time reversed their exposure to the technology sector from long to short, benefiting from the subsequent crash.

Much of the more recent literature focuses on measuring investor sentiment and how this is priced into risk by augmenting the default finance model that assumes EMH (Friedman, 1953) (Fama, 1965). For instance, (Brown & Cliff, 2004) analysed the impact of investor sentiment on asset returns with the use of VAR models, while (Berger & Turtle, 2012) employed a multi-factor model to investigate the cross-sectional returns of stocks against investor sentiment.

As previously mentioned, the lack of data has constrained research on an important proxy of investor sentiment, the actual positioning of the investors. (Stewart, 1949) undertook the first extensive analysis of trader positions by analysing the trades of 9’000 investors; his findings indicate that 75% of small investors lost money[[4]](#footnote-4). Since in the futures markets profits and losses must balance, he deducted there had to exist one or more other groups of investors who were on the other side of these trades, and made substantial profits.

Research undertaken using the actual CFTC data appears to be limited to a handful of academics. Among the first was (Houthakker, 1957) who examined the aggregate positioning of large hedgers, large speculators and small traders in the cotton, corn and wheat futures markets. His research concluded that large traders (i.e. large hedgers and speculators) possessed superior forecasting abilities, as opposed to small traders[[5]](#footnote-5). The lack of any statistical significance test however made (Chang, 1985) challenge these results; using the HM non-parametric test[[6]](#footnote-6) to check the timing abilities of the different investors, he determined that large hedgers were consistent losers, and attributed speculators’ profits in certain markets to risk premia[[7]](#footnote-7), and not to forecasting skills. Two years later, the research of (Hartzmark M. , 1987) contradicted Chang’s results; his analysis suggested that speculative traders earn negative or zero profits, while large hedgers are the most profitable[[8]](#footnote-8). In a follow-up study (Hartzmark M. , 1991) found that traders with consistent forecasting abilities are no more than one would expect due to chance, in other words that “luck determines trader performance”[[9]](#footnote-9). The extend of results contradicting each other clearly evidences the widely-recognized issues of analysing aggregate information (Holderness, 2016) (Hsiao, 2005).

More recently, (Buchananan, Hodges, & Theis, 2001) made use of the CoT reports to prove that prior to large price moves large speculators successfully took their largest positions. In what can however be considered to be the most comprehensive use of these datasets to date, (Wang, 2001) produced preliminary evidence that future commodity prices can be forecasted using the previous reported positioning. Interestingly, he determined that increased speculator and hedgers positions correlated well with trend continuation and reversal patterns in the price of the underlying commodity prices[[10]](#footnote-10).

What appears to be particularly interesting throughout the limited research undertaken using the CFTC data, is the lack of any literature on forecasting asset classes other than commodities[[11]](#footnote-11). To the extent of this review, only (Wang C. , 2003) appears to have looked into non-commodity assets. Using the same approach applied to his earlier study (Wang, 2001), but on the S&P500 futures contracts this time, he obtained exactly the same results: speculators’ positioning constitutes a trend continuation indicator, while hedgers’ positioning a price reversal indicator.

## Research Objective

While research has been conducted in a similar field, known as Sentiment Analysis and Models (Notter, 2013), and literature exists on forecasting commodity prices using the CoT datasets (Wang C. , 2003), the current paper aims to present a novel approach to utilizing these market positioning data (CoT and TFF) to produce a consistent currency and precious metals option trading model. In more detail the extent to which the results of (Wang C. , 2003) hold across these asset classes will be investigated, and an investment portfolio will be presented. The research objective is therefore to determine whether a successful investment portfolio can be constructed on the positioning data disclosed in the CFTC reports.

In order to accommodate the particularities of the CFTC data, a custom investment process has been devised in this paper (Section 3.3). In short, the developed models yield signals on a weekly basis, and performance is reproduced by selling 1-week, 20 delta strike options.

# Data Acquisition - Description

## Positioning Data

The CFTC overseas all trading activity in the U.S. futures market, and requires from all of its futures exchange clearing members and commission merchants, as well as foreign brokers, to report daily positions that meet specific reporting levels. Although the exact requirements vary depending on the market, the latest levels cover approximately 70%-90% of total open interest in each product (CFTC, 2006) (see Appendix 7.1). The CFTC has been compiling since 1962 the CoT report[[12]](#footnote-12), an aggregation of this data per market in long and short positions, and delivers them in two classifications: “futures only” and “futures and options" (Figure 2‑1), and two main broad participant categories:

1. *Commercials* *(COMM)* - a trading entity that is commercially "...engaged in business activities hedged by the use of the futures or option markets".
2. *Non-Commercials* *(NCOMM)*- remaining participants.

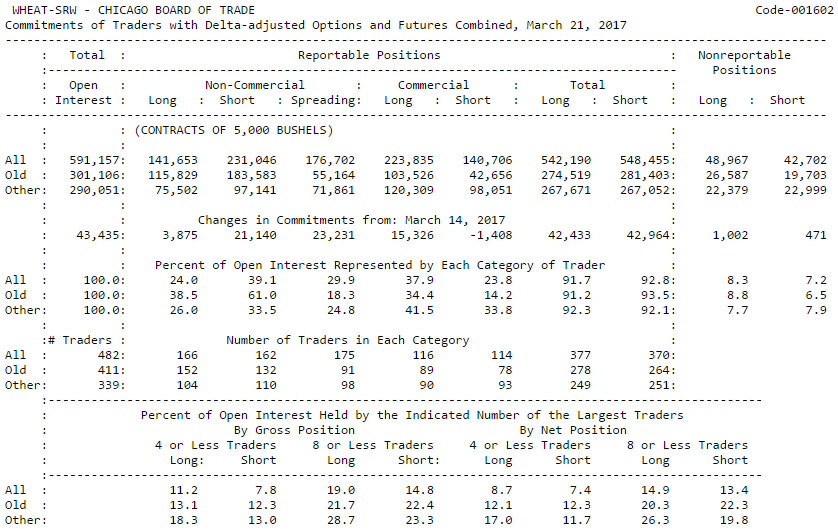


Figure ‑ – CoT Report Format

Since these data however separate reportable traders in very broad categories, in 2010 the CFTC announced the delivery of an additional weekly report. The later, termed the Traders in Financial Futures (TFF) report[[13]](#footnote-13), separates large trades in the financial markets in "sell-side" (Dealer / Intermediary) and "buy-side" (Asset Manager / Institutional, Leverage Funds, Other Reportable) participants:

1. *Dealer / Intermediary (DE)* - market participants that earn commissions from selling financial products, capturing the bid/ask spread and accommodating their clients. These include large institutions such as banks and securities' dealers that tend to run matched books, or offset their risk across clients and markets.
2. *Asset Manager / Institutional* *(AM)* - this segment covers institutional investors, such as mutual and pension funds, insurance companies, endowments etc.
3. *Leverage Funds* *(LM)*- these are various types of asset managers and hedge funds whose strategies may involve, among others, opening outright positions or arbitrage within or between markets.
4. *Other Reportables* - market participants not covered in the three categories above; traders in this category use markets mostly to hedge business risk. Examples include central banks, corporate treasuries, credit unions, mortgage originators, etc.

The CFTC backfilled the TFF report from June 1st, 2006, which is since delivered in a consistent format (Figure 2‑2) along with the CoT report.

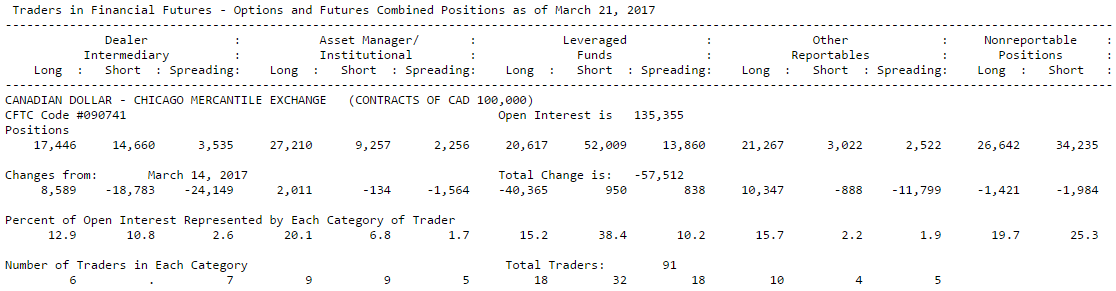


Figure ‑ – TFF Report Format

Since the present report will make use almost exclusively of the TFF / CoT data, certain limitations of the later are worth mentioning. The classification of the participants in any of the presented categories is conducted using a strict guideline, but as the CFTC states, judgement on the part of the Commission staff is sometimes required (e.g. particular classification due to the predominant business activity of the participant). Additionally, the Commission classifies the participants, and not their trading activities. For instance, hedge funds are known to fall under the "Leverage Funds" category, but not all of their activity can always be described as speculative (e.g. using the foreign exchange futures market to hedge exposure to an OTC core position). Finally, a multifunctional organization with more than one trading entities may have each entity classified separately in each market. It can therefore be the case that an asset manager with two trading entities has part of its positions in a market described as commercial (under the first entity), and the rest as non-commercial positions (under the second entity) (CFTC, 2006).

For the purpose of this report, the historical weekly CoT and TFF raw data have been collected and compiled in uniform databases for the period 01.06.2006-27.12.2016. While the data includes more than 600 instruments, and spans across numerous main and secondary markets[[14]](#footnote-14), the number of instruments with complete weekly positioning information is more limited. Particularly complete are currency, treasury and index data; this, in combination with the lack of any literature analysing these data, further strengthens the case of this thesis.

## Option Data

The CFTC reports are released to the public every Friday at the close of US markets. Making use of the latest positioning information, the investment process, described in detail in Section 3.3, stipulates trades must be placed every Monday at GMT 14:00[[15]](#footnote-15). In more detail, if one, or more, signals are generated, 1-week, 20 delta strike options on the underlying asset are sold. To that end, historical call/put premiums have been collected from Bloomberg for every Monday since June 1st, 2006, in line with the first available TFF data[[16]](#footnote-16).

The use of Bloomberg option data provides numerous advantages. The platform calculates historical option premia by using a number of market variables at the requested time and date (e.g. spot price, instrument volatility etc.). Furthermore, the data produced feature, except for the premium, also the settlement status of the priced option at expiry, along with its resulting profit/loss. This is particularly convenient since no additional calculations need to be performed to account for profit/losses in different currencies (e.g. trading non-USD denominated instruments), spot prices at expiry and public holidays. These features allowed the creation of a uniform cross-asset dollar-denominated database of call/put option premia based on a 1 million USD notional.

# Methodology

## Data Preparation

As outlined in Section 2.1, complete historical positioning databases were compiled from the aggregated CoT and TFF reports. Since these databases however contain the long/short positioning of each group of participants in absolute terms (i.e. number of contracts held), the data were transformed so as to indicate net participant positioning as a factor of total number of contracts held. To that end, the following transformation was applied:

,where is the net percentage positioning of a participant in a market. As an example, if Asset Managers positioning features 150’000 contracts long and 65’000 contracts short, the net positioning of the group is 85’000 contracts long on a total of 215’000 contracts, or 39.5%. The main advantage of this approach is that it creates a uniform positioning metric across market participants (Figure 3‑1). This transformation was applied on both the CoT and TFF datasets, yielding net positioning information for all market participants.

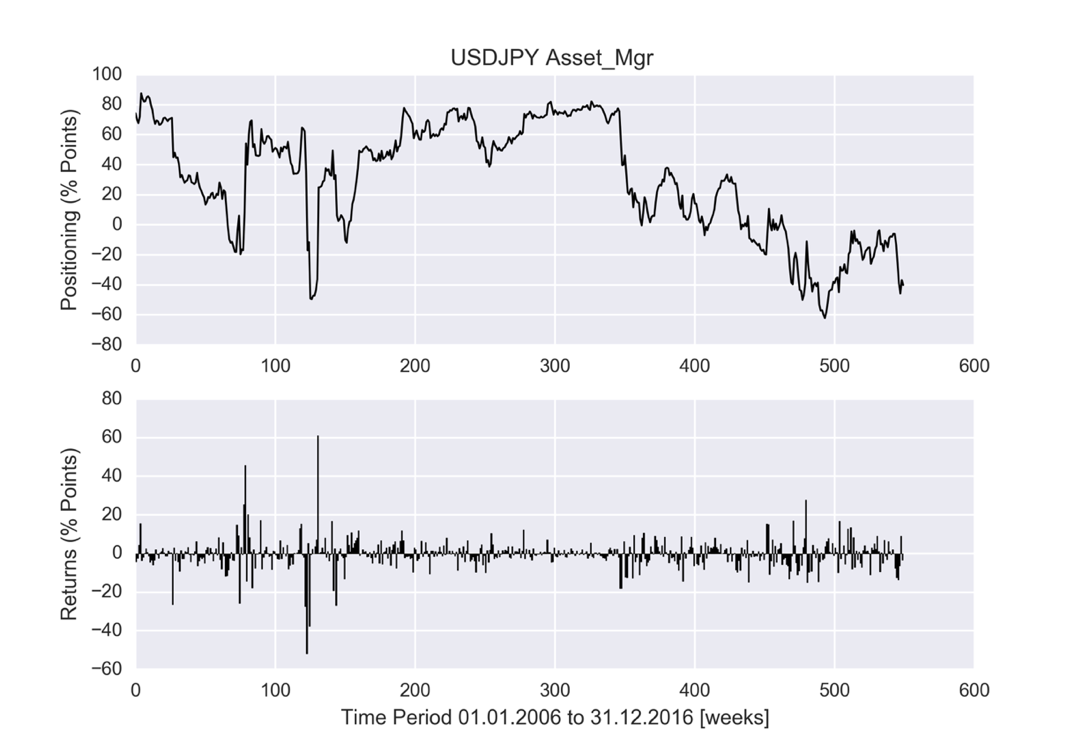
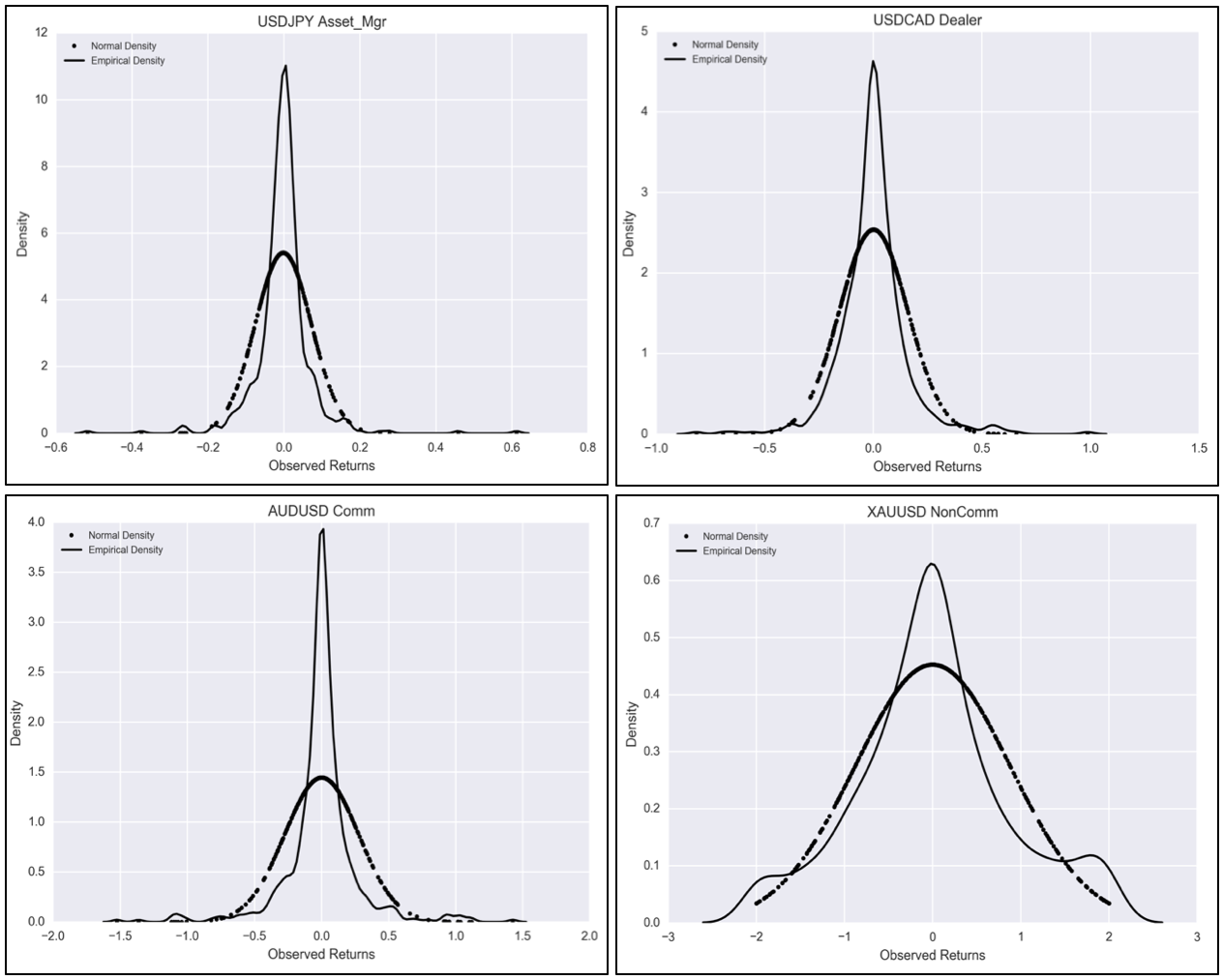


Figure ‑ – Positioning and Returns for Asset Managers in USDJPY

## Data Features

Since investment decisions will be based on participants’ positioning and week-over-week changes, the statistical properties of these data are of interest.

In line with Figure 3‑2, most assets’ empirical density appears to have higher peaks around their means and often fatter tails, than that of their corresponding normal distributions. In other words, the empirical density function is taller and skinnier, but with a wider support, than the normal density, making the normality assumption questionable.

Figure ‑ – Probability Density Function for USDJPY (AM), USDCAD (De), AUDUSD (Com), XAUUSD (NonComm)

Stationarity is another feature of the time series that needs to be investigated. This is of particular importance, since it lays at the core of any potential predictive power deriving from the CFTC data. One of the most commonly used unit root tests that provide information on the potential stationarity, or not, of a time series, is the Augmented Dickey-Fuller (ADF) test[[17]](#footnote-17). The test was carried out on all available time series data; the majority[[18]](#footnote-18) of time series feature results similar to the ones presented in Table 3‑1. The test statistic is in most cases smaller than the 1% critical value, implying the null hypothesis (H0) can be rejected with a significance level of less than 1%[[19]](#footnote-19). Therefore, the vast majority of the processes are deemed to not have a unit root, yielding stationary, or time-independent, time series.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ADF Test | AUDUSD | XAUUSD | USDJPY | GBPUSD | USDCAD |
| Commercials | Non-Commercials | Asset Managers | Dealers | Leverage Money |
| Test Statistic | -1.12E+01 | -1.08E+01 | -9.85E+00 | -8.75E+00 | -1.36E+01 |
| p-value | 1.98E-20 | 1.71E-19 | 4.39E-17 | 2.93E-14 | 1.80E-25 |
| # Lags Used | 1.10E+01 | 1.60E+01 | 7.00E+00 | 1.00E+01 | 3.00E+00 |
| # Observations | 5.60E+02 | 5.55E+02 | 5.41E+02 | 5.38E+02 | 5.45E+02 |
| Critical Value (5%) | -2.87E+00 | -2.87E+00 | -2.87E+00 | -2.87E+00 | -2.87E+00 |
| Critical Value (1%) | -3.44E+00 | -3.44E+00 | -3.44E+00 | -3.44E+00 | -3.44E+00 |
| Critical Value (10%) | -2.57E+00 | -2.57E+00 | -2.57E+00 | -2.57E+00 | -2.57E+00 |

Table ‑ – Augmented Dickey-Fuller test for a random selection of assets

With questionable normality and likely stationarity of the time series, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)[[20]](#footnote-20) of the processes is investigated next to determine the randomness of the data.

**ACF/PACF with 95% Confidence Intervals**

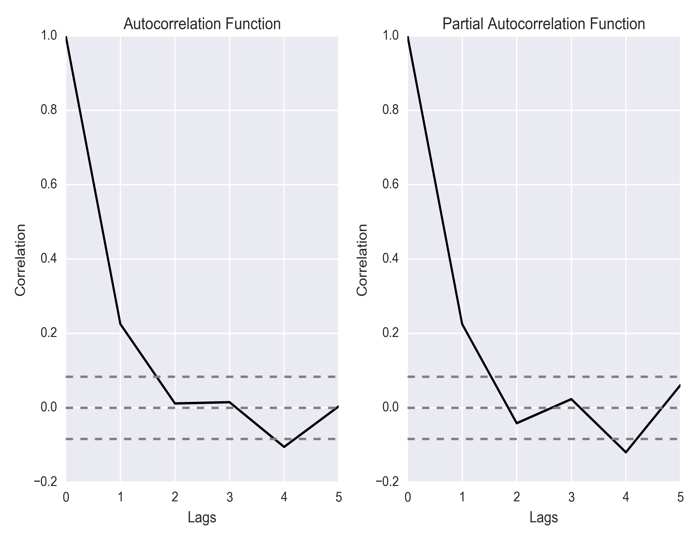
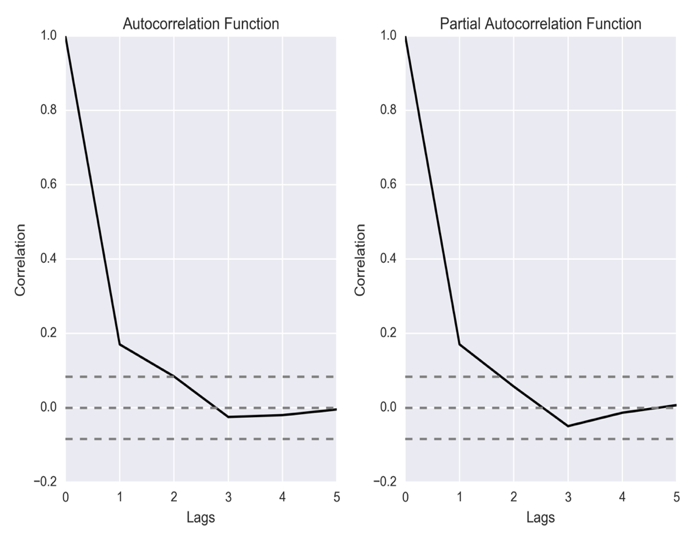


Figure ‑ – ACF / PACF on Positioning Returns of Asset Managers / Dealers in USDJPY / EURUSD

Calculating the ACF/PACF, as seen in Figure 3‑3, for all the available time series, it was determined the majority of assets of the TFF exhibit strong positive autocorrelation with regards to their previous values (i.e. lag of 1); interestingly the result is diametrically the opposite for the data of the CoT (Figure 3‑4). While for the TFF datasets the autocorrelation with a one-period-lag is substantially positive, for the CoT time series the autocorrelation with a one-period-lag is instead considerably negative. In both cases the autocorrelations are high in absolute terms, and well above the 95% confidence intervals (dashed lines in Figure 3‑3 and Figure 3‑4).

**ACF/PACF with 95% Confidence Intervals**

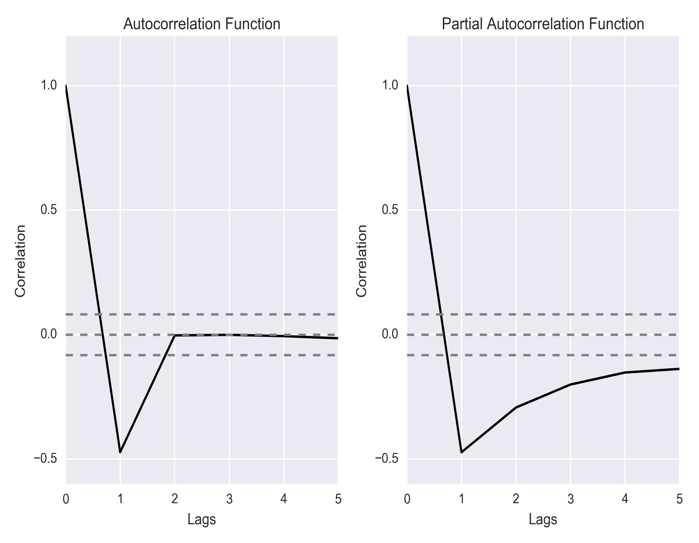
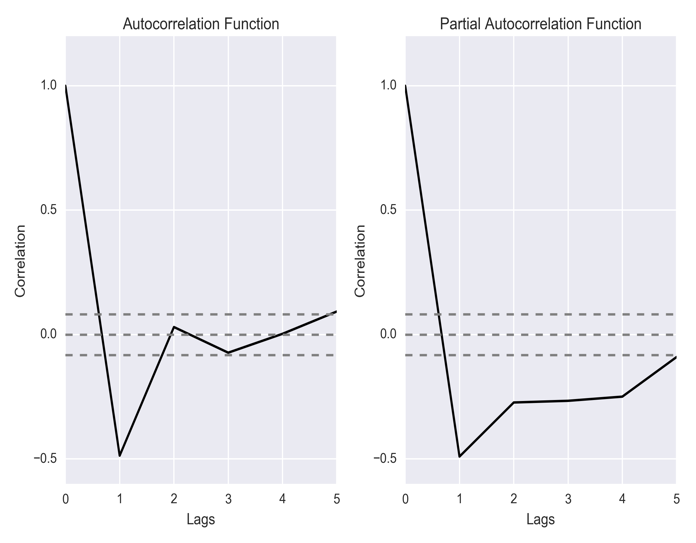


Figure ‑ ACF / PACF on Positioning Returns of Commercials / Non-Commercials in GBPUSD / USDCHF

The tests conducted determined questionable normality, likely stationarity and non-randomness of the time series. Based on these conclusions, a statistical model can be built to exploit these features within the scope of a sound investment process.

## Investment Framework

The investment processes implemented within the context of the literature examined are fairly simple. They mainly consist of monitoring the positioning data released and taking cash positions in the market by using the 1-period properties of the time series presented in Section 3.2. For instance, (Upperman, 2006) looks out for extremes in the positioning of wheat contracts, and, knowing the negative 1-period autocorrelation of the time series, attempts to catch price reversals using cash instruments, such as the futures’ market. Furthermore, it appears researched strategies concentrate on interpreting the CFTC datasets with a subsequent long investment time-horizon; proposed trades are therefore kept open for a number of weeks or months.

In this report, a new framework within which the proposed strategies will operate is presented. Instead of establishing cash positions, the potential of using options is examined; this alternative provides the investor with a range of advantages, such as potential for increased leverage, numerous parameterization options, reduced margin requirements and simple exit possibilities. As discussed in Section 2.1, positioning data are released from the CFTC on a weekly basis. The plan brought forward is therefore to analyse the data and create processes that yield one-week-long trading signals, economically quantified by selling out-of-the-market options. The potential profit hence lies within the premium collected on stipulating the option, and a subsequent non-exercise at expiry. As an example, if the analysis yields a SELL signal on a specific week for a market, an out-of-the-money, 1-week-long call option is sold; the premium is collected and after one week the option either expires worthless or in the money. Since the option data collected (Section 2.2) provide final settlement profit/loss of the options, the net Weekly Profit Loss function can be defined as follows:

The primary factors that influence option pricing are common across most option pricing formulae: the underlying’s price at the time of taking out the option, the strike price, the time until expiration, volatility, interest rates and dividends. Since this paper focuses on currencies and commodities, dividends do not apply, while interest rates are pre-set by the interest differentials between the underlyings of the asset (e.g. Euro and US Dollar for EURUSD). The volatility used to price options, known as implied volatility, is also pre-set since it represents the volatility the market, or counterparty, expects for the underlying asset over the life of the option. Therefore, the main parameters an investor is allowed to customize according to his needs are the strike, the notional and the expiration date/time of the option (Figure 3‑5).

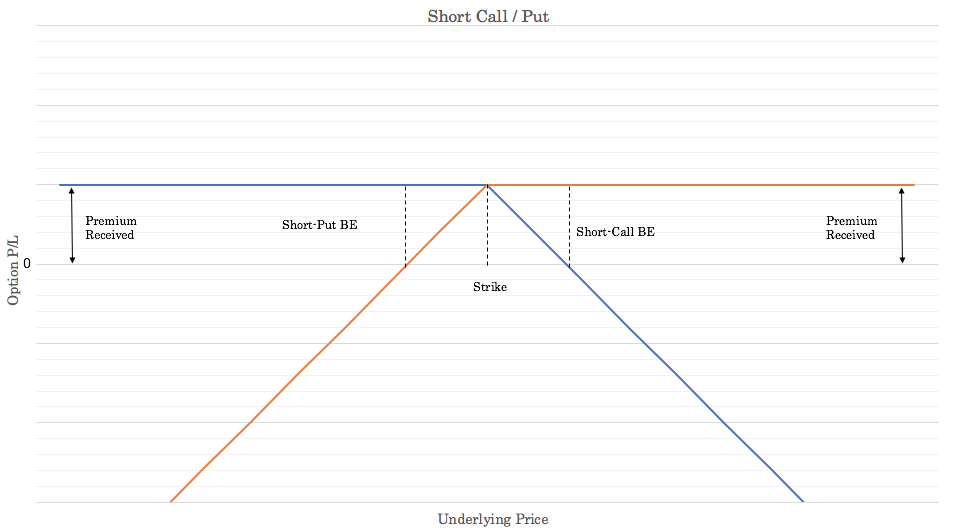


Figure ‑ – Option Pricing Plots / Parameters

As previously mentioned, the proposed strategy consists of selling call or put options depending on the signal generated. The options to be traded are all out of the money, with a strike of 20-Delta[[21]](#footnote-21), and a duration of one week. For the purposes of this paper options were priced on Mondays, at GMT 14:00, and hence expire the following Monday at GMT 14:00, when possibly the next option is traded (Figure 3‑6).



Figure ‑ – Duration of Options under Proposed Investment Process

For all tasks relating to data analysis, model development as well as portfolio construction and optimization, the computational packages MATLAB and Python are utilized. Both packages feature a plethora of statistical and economics tools allowing for the cross-validation of the results produced[[22]](#footnote-22).

## BASE Model

To determine the preliminary predictive capabilities of the CFTC data, a BASE Model was developed (Figure 2‑1).

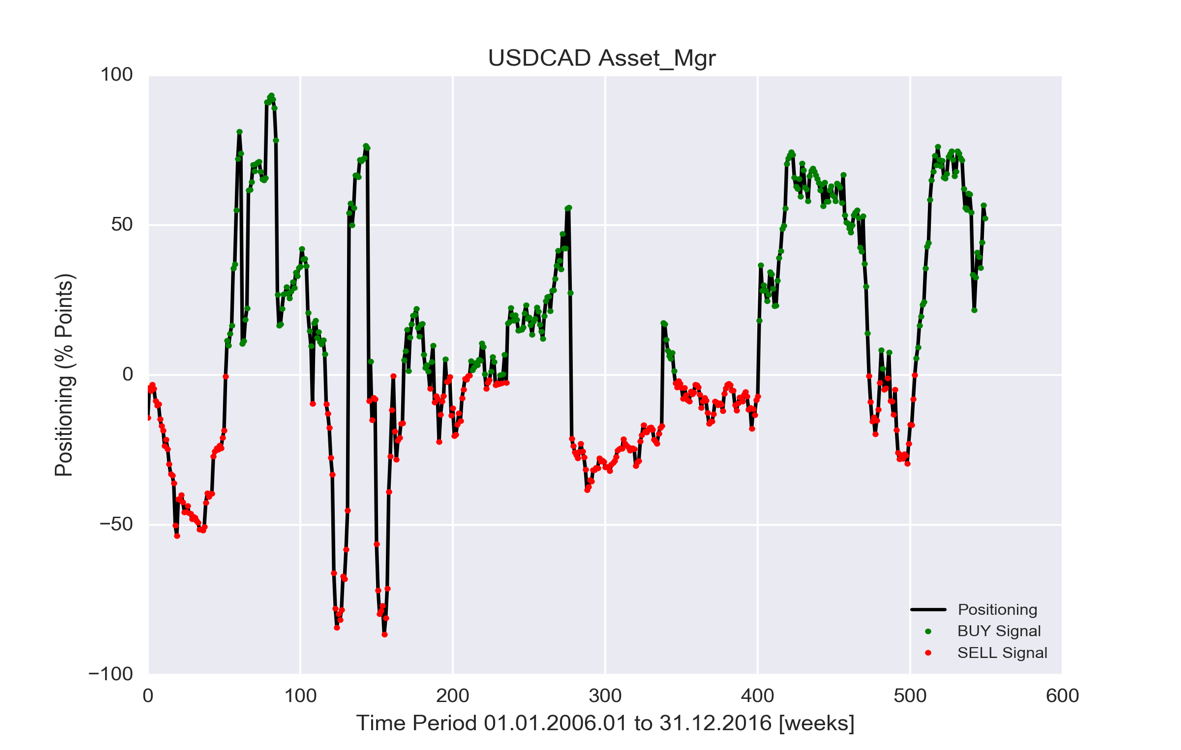


Figure ‑ – BASE Model Signals on USDCAD using Asset Managers’ Positioning

The process has a simplistic framework which monitors the positioning of market participants in an asset and attempts to be on the same side of the trade as them[[23]](#footnote-23). Therefore, if for instance Commercials’ positioning is net positive (i.e. long) at any point in time, the system will produce a LONG signal for the next period, selling, as a result, the put option. The model can hence be summarised as follows:

,and the Weekly Profit Loss function is hence:

The process looks only at the absolute positioning of a participant, completely disregarding its’ actual scale or relation to the overall market dynamics. Furthermore, one of the shortcomings of the system is its constant presence in the markets. Nonetheless, the BASE model constitutes a good benchmark for the AR and GARCH models presented in Sections 3.5 and 3.6.

## AR Model

There exist many models used for time series forecasting and monitoring, however three very broad classes are most often used. These are the autoregressive (AR) models, the integrated (I) models and the moving average (MA) models. These processes are often combined together to produce new models, such as the ARMA model and the ARIMA model. The most commonly employed model for time series analysis is the autoregressive process (AR), a difference equation determined by random variables, whose distribution is the key component to model time series.

Based on the time series properties determined in Section 3.2[[24]](#footnote-24) the first order autoregressive equation is examined. The *pth* order autoregressive time series (also known as ) is defined as follows:

,where and are assumed to be uncorrelated random variables (i.e. ). Therefore, the process is simply a first order linear difference equation described by:

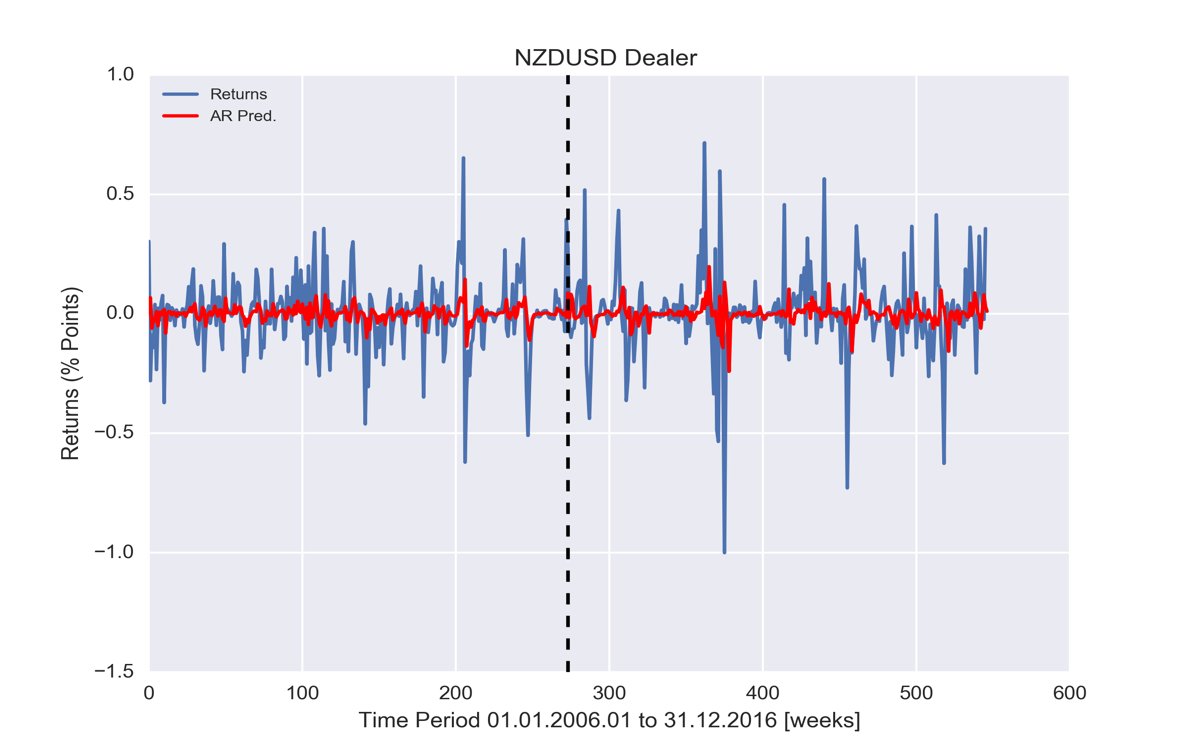
**

Figure ‑ - Positioning Returns and AR(1) Fit for Dealers in NZDUSD

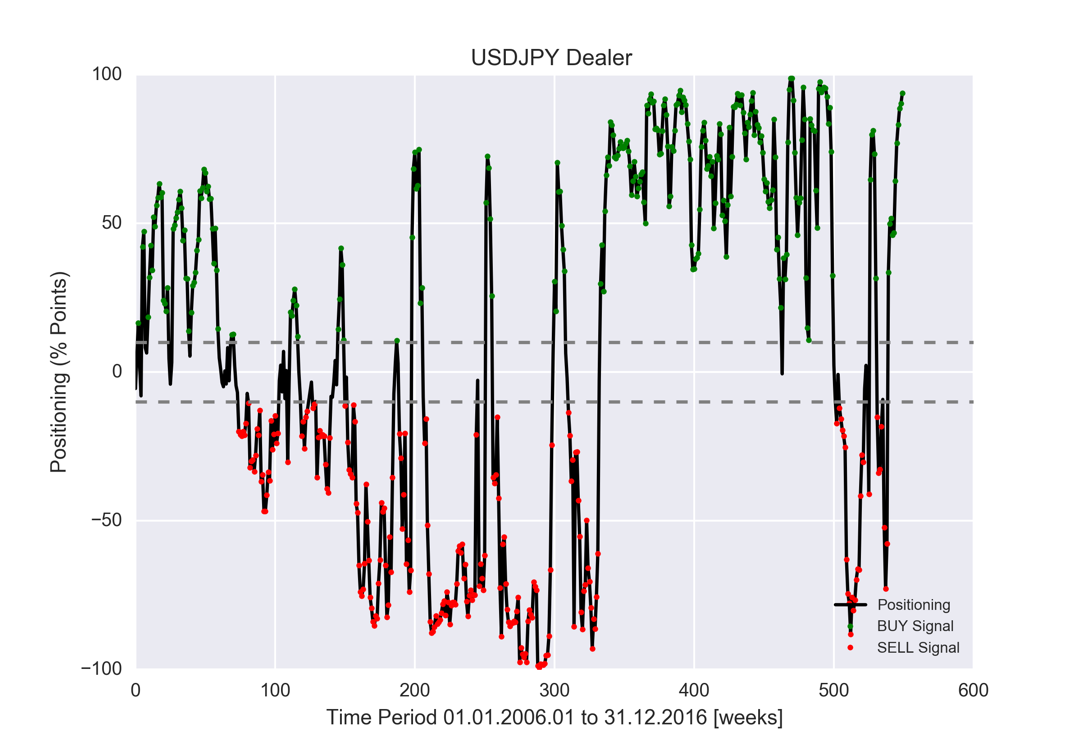


Figure ‑ - Positioning Filter on USDJPY for Dealers’ Positioning

Although the core investment approach remains the same as in the BASE model, for the model the forecasted positioning returns are employed to produce a trade signal. Therefore, the forecasted return is:

,and the resulting forecasted next period positioning is:

In order to yield significant model parameters, the AR model is optimized and fitted on half of each time series and then re-fitted at every data point going forward to create the next period forecast. Additionally, a minimum positioning threshold is put in place to filter out net positions and an optimization is further developed to yield an optimal weight between the forecast and the latest positioning data point. In more detail:

1. *Filter* – Analysis of the time series evidences that if the net positioning of a market participant is in the range of , the likelihood of a successful signal reduces substantially[[25]](#footnote-26). Therefore, any signals generated based on a net positioning in this range are disregarded (e.g. Figure 3‑9).
2. *Weight Optimization* – The premise here is to produce a consistent positioning prediction using a combination of the last period data and the forecast. The aim of the optimization process is therefore to yield an optimal weight between the two that maximizes the strategy’s equity at the end of the in-sample data[[26]](#footnote-27):

,where:

* – the positioning of the market in the previous period
* – the positioning predicted for the next period
* *N* – sample size

,and subject, as per point *1.*, to the resulting combined absolute prediction being bigger than 10% (i.e.).

To summarize, the designed method optimizes the weight (*w*) between the latest positioning data point and the forecast to maximize the equity at the end of the in-sample data. In the process, the optimized absolute positioning forecasts that are smaller than 10% are disregarded, yielding a neutral signal, during which the model takes no positions in the market.

## GARCH Model

Having developed a positioning forecast model (Section 3.5), the alternative of a volatility based strategy is investigated. The idea behind this is to use the forecasted volatility produced by a GARCH process to filter out trades from the BASE model described in Section 3.4. The premise is therefore that large positioning reversals (i.e. high volatility events), that result in losing trades, can be forecasted through such a model and excluded from the BASE trading strategy.

The motivation behind studying conditional heteroscedasticity in financial time series is that of volatility of returns; this is of particular importance in finance since it is a notion highly synonymous to risk (Bollerslev, 1986). Let us first investigate the concept of *heteroscedasticity* and then discuss the *conditional* part.

Given a set of random variables, the collection is said to be heteroscedastic if there are certain subsets of variables within the larger set that have a different variance from the remaining variables. This could, for instance, be the case in a non-stationary time series that exhibits trending or seasonality effects; in such scenarios, it is possible for the variance to increase with this trend or these seasonality effects.

The correlation between an increase in variance and future further increase in variance occurs for many reasons and is a highly-researched topic. For instance, in the case of downside portfolio protection insurance used by long–only funds, a substantial one-period drop in the equities in the portfolios can trigger automated risk management sell orders, which in turn further supresses the price of equities. Such occurrences, as well as many other forms of volatility, lead to heteroscedasticity that is serially correlated, and thus conditional on periods of higher variance; such series are known as conditional heteroscedastic.

While volatility clustering, this phenomenon of periods of relatively high and/or low volatility, is a seemingly common attribute of market data, no universally accepted explanation of it exists (Alexander, 2008). One set of models that attempt to model these clusterings are the GARCH processes. Given a time series , described at each instance by , where is discrete white noise with zero mean and unit variance, GARCH methods attempt to model these volatility clusters by defining as:

,where and are parameters of the model. Therefore, a GARCH(1,1) model is defined as :

,where , . Furthermore, in order for the next period forecast of variance to be a blend of the last period forecast and squared return we require:

For the purposes of this paper, the positioning returns are fed in the GARCH(1,1) model to make a forecast for the next period volatility . Once again, the model is fitted on the in-sample data (i.e. first half of the time series length) and re-fitted at every data point before predicting the next period volatility on the out-of-sample data (i.e. second half of the time series) (Figure 3‑10). To remove trades for which a high volatility has been forecasted, a volatility filter is employed. In more detail the realized variance of the last four period returns is calculated, converted to volatility , and compared to the forecasted next period volatility; if the forecasted volatility is higher than the realized volatility, the trade is skipped. Therefore, the model can be summarized as follows:

,where:

* – the previous period positioning
* – the next period variance forecasted by the GARCH(1,1) model
* – the realized variance of the last four data points

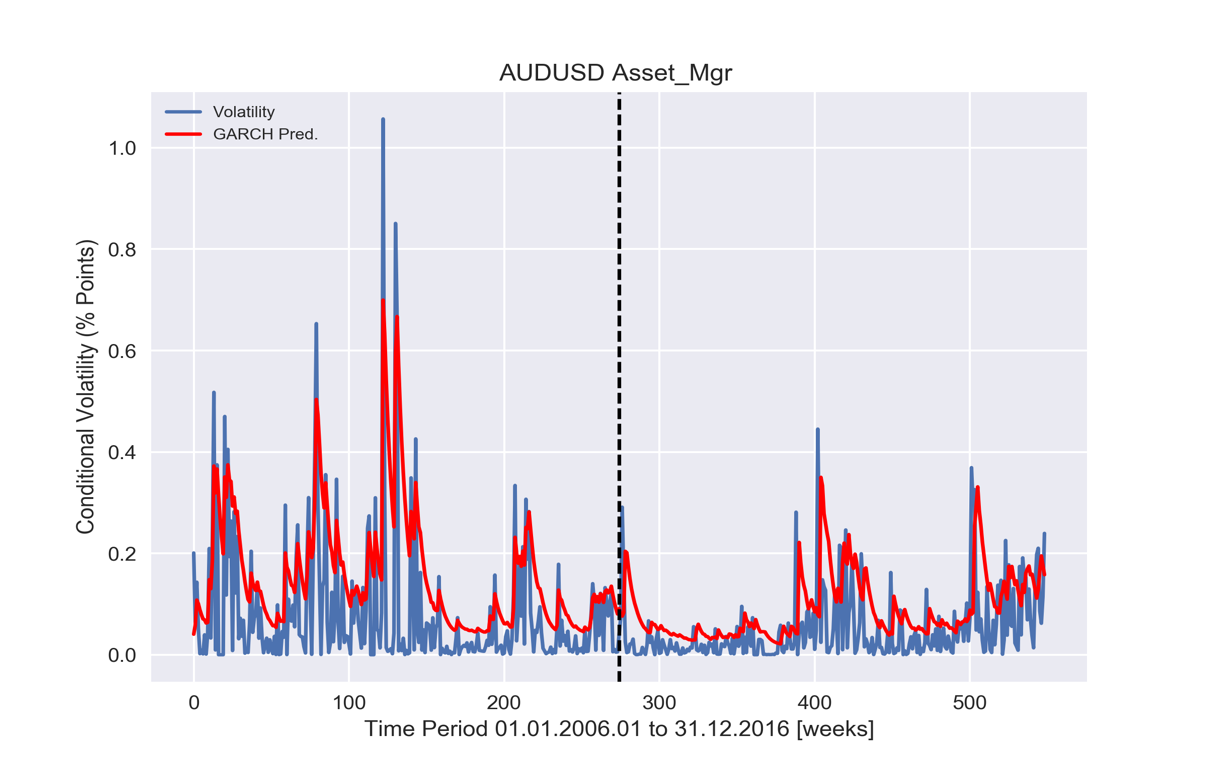


Figure ‑ – Volatility and GARCH Fit for Asset Managers in AUDUSD

## Asset Selection

Given the number of assets available in the CFTC reports and the diversity of market participants, constructing an optimal portfolio for the presented strategies is no trivial task. While several selection models exist in the literature that apply limitations to the number of assets and the weights in a portfolio, the Limited Asset Markowitz Model is investigated here.

The classical Mean-Variance portfolio optimization process (Markowitz, 1952) aims to determine the fractions of a given capital to be invested in each asset belonging to a pool of assets. The objective function of this model is to minimize the risk of the return of the entire portfolio (i.e. variance), while restricting the expected return of the portfolio to reach a specified value. In more detail, assuming available assets, each asset with an expected return , and the covariance of the returns of assets and , the classical MV model is:

,where is the required level of portfolio return. Next, two constraints are added, one limiting the number of assets that can be held in the portfolio to , and one that the quantity allocated to each asset must be constrained within a defined interval . Applying these two constraints yields the Limited Asset Markowitz (LAM) model:

subject to:

,where .

For the purposes of this paper, the Limited Asset Markowitz Model was implemented using a Lasso regression model (Yan, 2009) using the Python programming language. The process was designed to select 5 assets, each with a weight in the range and to conduct a re-balancing of the portfolio every year. Once the LAM model yields the preferred assets and allocations for the past 72 weeks, portfolio returns are calculated for the following year, at the end of which a new re-balancing is conducted. In more detail, the later occurs every 52 weeks (i.e. one year), with a lookback of 72 weeks (i.e. one year and a half); this allows for a variable selection of assets over the strategy’s lifecycle, while also enabling the model to capture short-term positive performances of market participants.

# Results

## Overview

Following the implementation of the three models presented in Section 3 (BASE, AR and GARCH Models), the generated signals are matched against the option pricing database (Section 2.2), and the resulting participant, asset-specific, equity lines and performance metrics are derived. Given the 22 complete time series analysed and the number of participants available[[27]](#footnote-28), a total of 110 equity lines are generated. The Limited Asset Markowitz Model (Section3.7) subsequently analyses parts of these time series and yields a selection of assets to be traded for the subsequent year. Following the construction of the portfolio index, it is benchmarked against two of its variations[[28]](#footnote-29). Finally, a comparison of the generated portfolio against a basket of equity, currency/commodity and fixed income indices is carried out.

The performance of the produced equities[[29]](#footnote-30) are computed against an initial AUM of USD 1 million; this corresponds to a leverage of 1, an extremely low value when considering all instruments lay in the currencies and precious market space where the industry’s (e.g. asset managers, hedge funds etc.) leverage factors tend to rather be in that 3-10 area (Mirabile, 2016). The performance metrics calculated are the Sharpe Ratio, for which a risk-free rate of +0.68%[[30]](#footnote-31) was assumed, the Max. Drawdown (Appendix 7.2) and a number of commonly used ratios (e.g. Profit Factor, Hit Ratio, Cumulative Returns etc.). Given the number of produced equities, a random selection of 5 assets[[31]](#footnote-32) is presented in Sections 4.2, 4.3 and 4.4. to showcase the respective properties of each model.

## BASE Model

As described in Section 3.4, the BASE Model constitutes the foundation of the predictive power of the CFTC data and acts as the benchmark for the AR Model and GARCH Model. In short, the BASE Model is always in the market attempting to generate positive returns from being on the same side as the participants.

The main performance metrics of the randomly selected assets are displayed in Table 4‑1, while the respective indexed cumulative returns are presented in Figure 4‑1. Except from the overall positive performance of the curves (except the Euro), the metrics indicate in all cases high Hit Ratios[[32]](#footnote-33); this, in combination with the meagre Profit Factors[[33]](#footnote-34), indicate that a large number of positive, albeit small in profit, trades are counterbalanced by a small number, albeit large in loss, trades. Furthermore, the Max. Drawdowns appear to be substantial (in the 10% area), a negative attribute compounded by similar or even lower absolute cumulative returns. To summarise, with the exception of the Japanese Yen whose performance stands out, the remaining assets’ equities feature relatively high volatilities, as reflected by their Sharpe Ratios.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Asset** | **GBPUSD** | **USDCAD** | **USDJPY** | **AUDUSD** | **EURUSD** |
| **Participant** | **Commercial** | **Asset Manager** | **Dealer** | **NonCommercials** | **Leveverage Money** |
| Cum. Performance | +11.6% | +13.6% | +23.1% | +2.7% | -0.2% |
| Profit Factor | 1.15 | 1.28 | 1.49 | 1.04 | 1.00 |
| Hit Ratio | 81.2% | 86.0% | 86.9% | 82.4% | 82.2% |
| Max. DrawDown | -9.53% | -10.06% | -3.90% | -9.96% | -11.91% |
| Sharpe Ratio | 0.32 | 0.46 | 0.81 | 0.09 | 0.01 |

Table ‑ – BASE Model Performance Metrics (weekly basis)

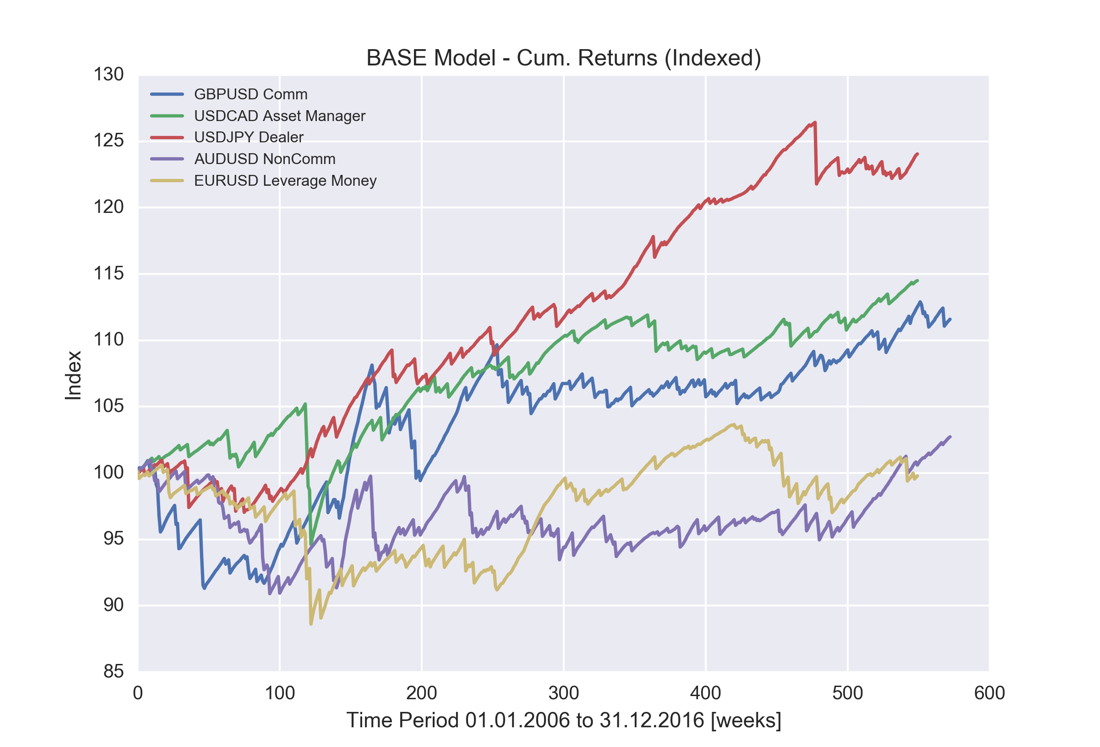


Figure ‑ – BASE Model – Cum. Returns (per asset)

## AR Model

Moving on to the AR Model, as presented in Section 3.5, this model constitutes of an AR(1) process combined with a filter and an optimization between the positioning return of the previous period and the forecasted value. It is important to note this optimization is conducted only once, on the in-sample data, and employed in the out-of-sample series.

The AR Model equity lines (Figure 4‑2) appear to be comparable to the ones of the BASE Model (Figure 4‑2), while the performance metrics indicate an overall increase in the Profit Factors and Cumulative Returns. Additionally, Max. DrawDowns are lower, or in line with the BASE Model, while Sharpe Ratios appear to improve, largely because of reduced volatility. The later evidences that the optimisation conducted and introduction of neutral periods[[34]](#footnote-35) in the AR Model via the filter have beneficial effects. The only real underperformer appears to be the British Pound, whose metrics are substantially inferior to the BASE Model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Asset** | **GBPUSD** | **USDCAD** | **USDJPY** | **AUDUSD** | **EURUSD** |
| **Participant** | **Commercial** | **Asset Manager** | **Dealer** | **NonCommercials** | **Leveverage Money** |
| Cum. Returns | +2.3% | +14.5% | +20.5% | +9.6% | +2.5% |
| Profit Factor | 1.09 | 1.45 | 1.52 | 1.16 | 1.05 |
| Hit Ratio | 84.3% | 86.9% | 87.3% | 83.2% | 81.8% |
| Max. DrawDown | -4.72% | -10.17% | -4.00% | -9.20% | -9.61% |
| Sharpe Ratio | 0.12 | 0.56 | 0.78 | 0.31 | 0.10 |

Table ‑ – AR Model Performance Metrics (weekly basis)

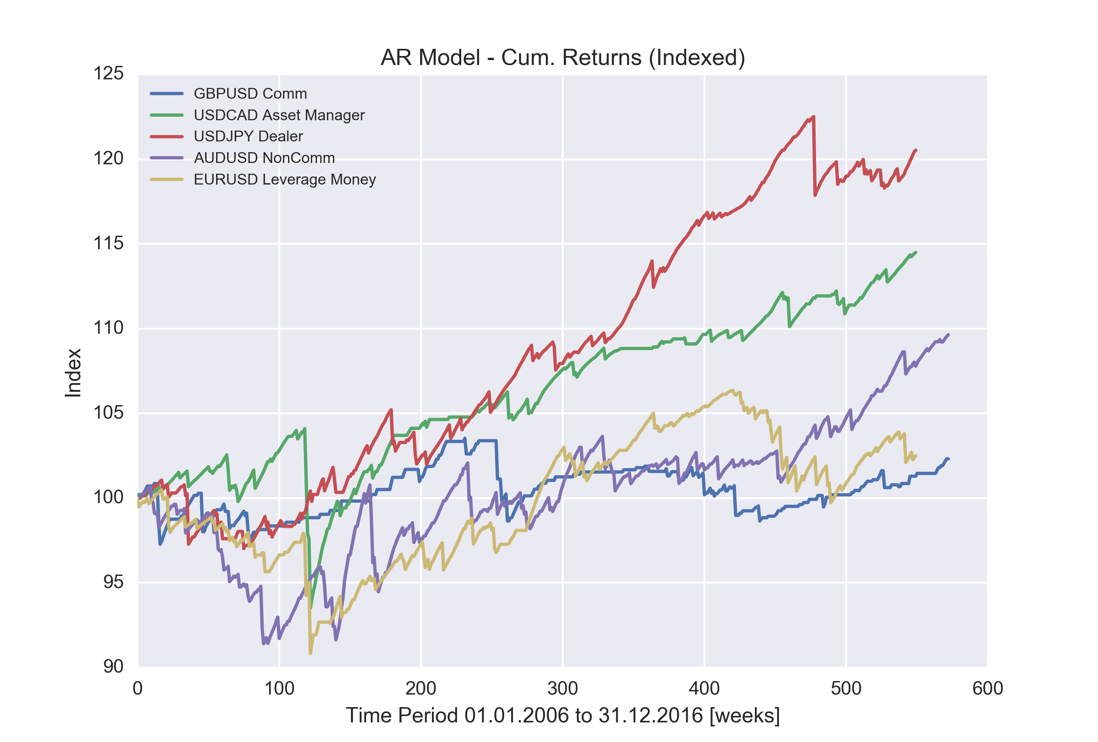


Figure ‑ – AR Model – Cum. Returns (per asset)

## GARCH Model

The GARCH model introduced in Section 3.6 is based on the premise that large positioning reversals, and subsequent losses from open trades, can be avoided through the forecasting of the series volatility. In more detail, the trades whose forecasted volatility is above the historical norm are excluded.

The performance metrics of the GARCH Model (Table 4‑3) do not display substantial divergences from the two other models. While the equities and metrics of British Pound and the Japanese Yen appear to outperform both the BASE and the AR Model[[35]](#footnote-36), the Canadian Dollar and the Australian Dollar being marginally worse off. Regarding presence in the market, the volatility filter excluded, on average, of the trades, less than the filter in the AR Model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Asset** | **GBPUSD** | **USDCAD** | **USDJPY** | **AUDUSD** | **EURUSD** |
| **Participant** | **Commercial** | **Asset Manager** | **Dealer** | **NonCommercials** | **Leveverage Money** |
| Cum. Returns | +12.4% | +12.4% | +23.0% | +0.3% | +0.7% |
| Profit Factor | 1.18 | 1.29 | 1.54 | 1.11 | 1.01 |
| Hit Ratio | 82.9% | 86.9% | 87.3% | 83.0% | 82.6% |
| Max. DrawDown | -9.37% | -10.07% | -3.99% | -7.92% | -12.02% |
| Sharpe Ratio | 0.36 | 0.44 | 0.85 | 0.02 | 0.04 |

Table ‑ - GARCH Model Performance Metrics (weekly basis)

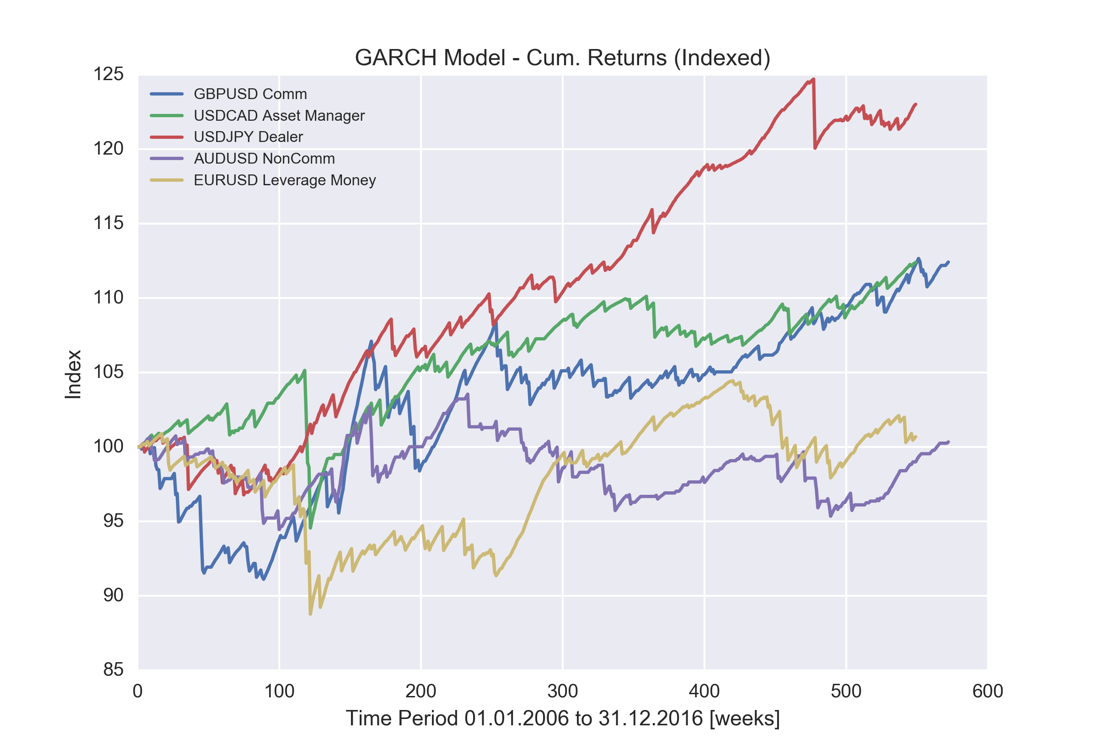


Figure ‑ - GARCH Model – Cum. Returns (per asset)

## Portfolio

The results presented in Sections 4.2-4.4 evidence that no single model is best. Depending on the asset and market participant examined, either the BASE, AR or GARCH model yield a better investable opportunity; for this reason an asset selection process, as the LAM Model (Section 3.7), is deemed necessary. This claim is further reinforced by the universe of signals available (i.e. 110 signals across 22 assets), making it unfeasible to allocate capital to each one of them using modern portfolio theory (e.g. Mean-Variance analysis).

In this section, a review of potential biases of the selection model is initially conducted. Subsequently, an overview of the portfolio produced, composition and performance is presented, followed by a comparison against two benchmark versions of the LAM Model. In more detail, *Benchmark 1* consists of the equities selected by the LAM Model but with equal allocations, while *Benchmark 2* is a portfolio produced once again, by the LAM Model but composed only of equities derived from the BASE Model. Therefore, the notation in this section is as follows:

* *LAM Portfolio* – the portfolio the LAM Model yields with a full investment universe, and the optimal allocations.
* *Benchmark 1* – the portfolio the LAM Model yields with a full investment universe but with equal allocations (i.e. equally weighted-portfolio).
* *Benchmark 2* – the portfolio the LAM Model yields with an investment universe restricted to the BASE equities, and the optimal allocations.

Finally, the produced portfolio is set side by side with a number of market indices, to determine its correlation and performance vis-à-vis equities as well as managed fixed-income and systematic FX/PM programs.

## Biases

There exist numerous biases that could affect the performance of the selection process / optimization undertaken in this paper. Since the majority of these biases tend to inflate the resulting performance, rather than detract from it, particular attention needs to be placed on identifying and mitigating them. The major biases are optimisation bias, look-ahead bias and survivorship bias, and in the context of this paper they can be summarized and dealt with as follows:

* *Optimisation Bias* – also known as “curve fitting” or “data-snooping bias”, optimisation bias consists of the optimisation or addition of trading parameters resulting in an attractive performance on the backtest data. This type of bias is however not likely to affect the LAM Model. Since the in-sample data are utilized to select the assets and respective allocations for the next period (i.e. out-of-sample) data, no optimisation takes place at the portfolio level and the resulting equity curve.
* *Look-Ahead Bias* – this type of bias is introduced into a strategy when future data are accidentally included at a point of the simulation where these would have actually not been available. The use of a cognitive computing process such as the LAM Model, in combination with the use of checks in the last window fed to the process ensure such bias cannot occur.
* *Survivorship Bias* – a particularly dangerous phenomenon whereby a strategy is tested on incomplete datasets that do not cover the full investment universe that may have been chosen at a particular point in time. During the data collection stage of the project, it was ensured the range of investment vehicles made available to the LAM Model are complete; given the nature of the analysed assets (currencies, precious metals, indices), as opposed to, for instance, stocks that may get de-listed, this type of bias can be excluded.

## Preliminary Results

After determining all of the biases described in Section 4.5.1 have been ruled out or appropriately mitigated, the performance of the produced LAM Portfolio is examined[[36]](#footnote-37). The metrics of the strategy (Table 4‑4) indicate a cumulative performance of +31.8%, implying an annualized average return of +3.06%; the Sharpe Ratio of 1.05, in combination with the Risk-Free rate employed of 0.68% suggests the annualized volatility stands at 2.3%. The Hit Ratio[[37]](#footnote-38) is 78.2%, below the averages observed in Sections 4.2-4.4, while the Max. DrawDown is -5.7%, a reasonably low value considering the overall performance of the portfolio.

In Figure 4‑4 the indexed performance of the portfolio, as well as the re-balancing periods, can be seen. Of interest, the portfolio appears to have experienced only one losing year (out of 10), while the larger moves appear to occur at the beginning or mid of each year. Regarding the composition of the portfolio, a range of 41 different vehicles has been invested-in throughout the lifecycle of the strategy. In more detail, a clear preference for the AR Model is observed (46.3% of equities selected derived from AR Model) and for the market participants present in the CoT report (Commercials and Non-Commercials). Furthermore, asset-wise, the Japanese Yen and the Swiss Franc appear to be the most invested-in instruments, while gold and the New Zealand Dollar the less ones.

Of interest, the distribution of the portfolio returns (Figure 4‑5) confirms an evident, previously examined (Section 4.2), feature of the investment strategy; the process consists of selling options, therefore the portfolio features numerous small profitable trades (in the 0.0% – 0.3% area, from options that expire out of the money or marginally in the money yielding overall positive trades), and a small number of large losers (options that expire in the money yielding losing trades). Indeed, only 7 weeks are observed with a performance in the range -0.7% to -1.7%; these trades correspond to 76.7% of the total losing weeks, underlying the extent of the potential losses of this portfolio. Finally, strong seasonality effects appear to be in place (Figure 4‑5), with the largest part of the returns generated in the months between October and February, while March is the worst performing month of the year[[38]](#footnote-39).

|  |  |
| --- | --- |
| **LAM Portfolio** | |
| Cum. Returns | +31.8% |
| Profit Factor | 1.58 |
| Hit Ratio | 78.2% |
| Max. DrawDown | -5.69% |
| Sharpe Ratio | 1.05 |
| Leverage | 1 |

Table ‑ –LAM Portfolio Performance Metrics (weekly basis)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Preference** |  | **Asset** | **Preference** |  | **Participant** | **Preference** |
| BASE | 24.4% |  | USDJPY | 19.5% |  | NonCommercial | 26.8% |
| AR | 46.3% |  | USDCHF | 17.1% |  | Commercial | 24.4% |
| GARCH | 29.3% |  | AUDUSD | 14.6% |  | Asset Manager | 22% |

Table ‑ – LAM Portfolio Asset Selection Information

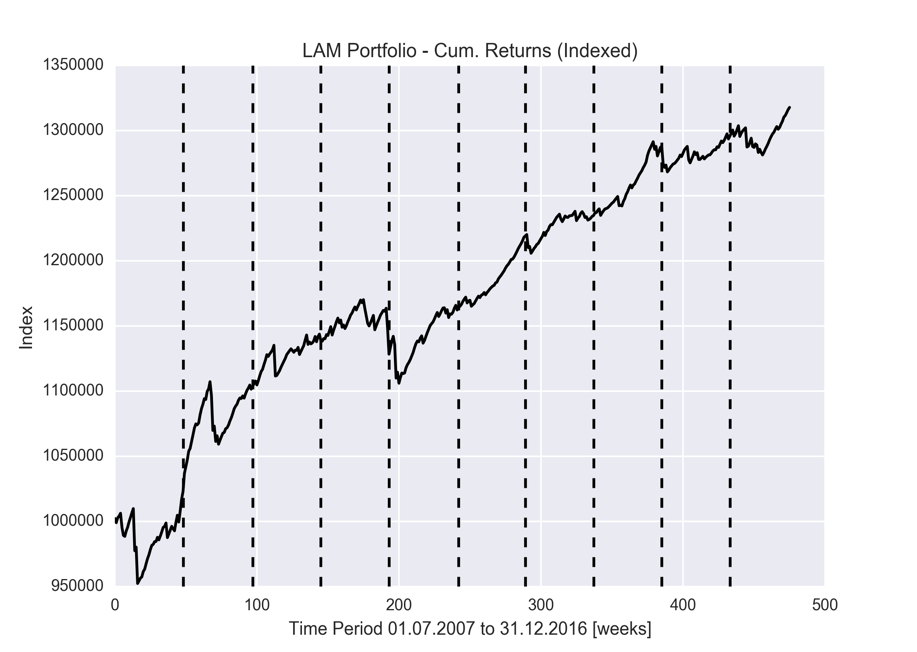


Figure ‑ – LAM Portfolio – Cum. Returns

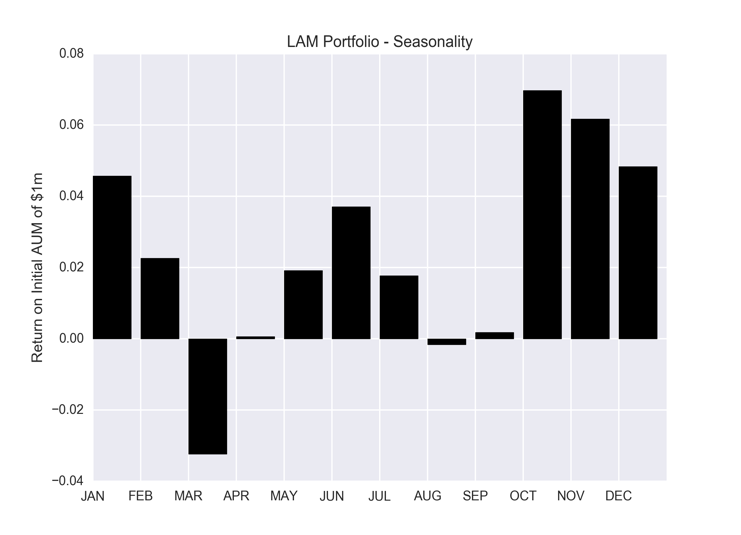
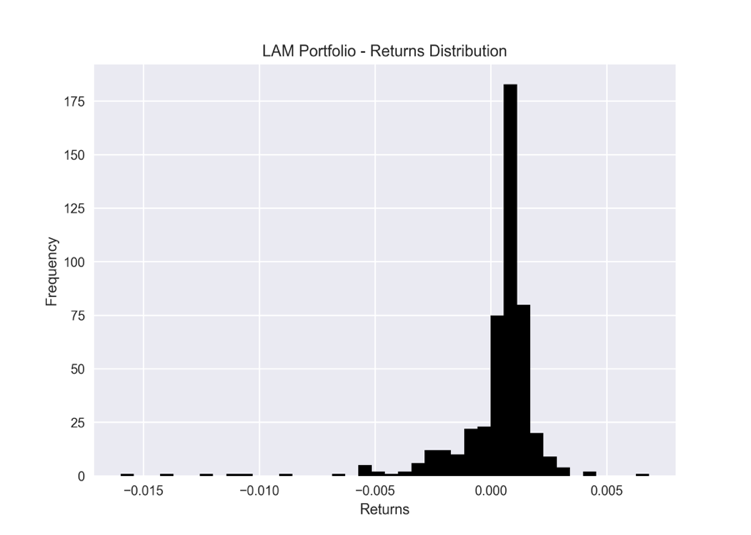


Figure ‑ – LAM Portfolio – Returns Distribution / Seasonality

## Portfolio vs. Model Benchmarks

In order to assess the strength of the portfolio and the robustness of the optimization process undertaken through the LAM Model, two benchmarks are presented. *Benchmark 1* consists of the same assets selected by the core portfolio presented in Section 4.5.2, but with equal allocations across them. The motivation behind this lies within extensive literature which suggests that “naïve” diversification (i.e. holding asset weights equal) is hard to beat[[39]](#footnote-40). *Benchmark 2* on the other hand, is identical to the core *LAM Portfolio*, the only difference being it is composed of strategies derived from the BASE Model alone[[40]](#footnote-41). The objective in this case is therefore to check for the strength of the AR and GARCH models, against the BASE model in a portfolio environment.

The “too-hard-to-beat” statement regarding the equally weighted portfolio appears to be proved by the resulting performances (Figure 4‑6). Comparing the *LAM Portfolio* with *Benchmark 1*, the latter appears to outperform in terms of Cumulative Returns, but is more volatile; this is confirmed by the higher Sharpe Ratio of the *LAM Portfolio*. Furthermore, the higher Profit Factor and lower Max. DrawDown of the latter indicates the allocations made by the *LAM Portfolio* to its constituents is superior. *Benchmark 2*, composed of BASE Model-derived equities, has an entirely different performance, and appears to lag substantially the other two models. Except for the marginally positive performance, *Benchmark 2* features a lower Profit Factor, lower Hit Ratio, lower Sharpe Ratio and higher Max. DrawDown.

To summarize, both *Benchmark 1* and *Benchmark 2* are shown to feature fatter tails in their returns distributions (Figure 4‑5 and Figure 4‑7), while lagging in overall performance to the *LAM Portfolio*. The optimization process devised therefore achieves its main objective of minimizing the portfolio’s volatility for the assets it selects, while yielding an attractive return. Furthermore, the weakness of a BASE Model-only portfolio against a larger universe of econometrically-derived models is substantiated.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **LAM Portfolio** | **Benchmark 1** | **Benchmark 2** |
| Cum. Returns | +31.8% | +32.6% | +1.5% |
| Profit Factor | 1.58 | 1.48 | 1.02 |
| Hit Ratio | 78.2% | 78.2% | 71.4% |
| Max. DrawDown | -5.69% | -8.32% | -12.75% |
| Sharpe Ratio | 1.05 | 0.85 | 0.06 |
| Leverage | 1 | | |

Table ‑ - LAM Portfolio vs. Model Benchmarks Performance Metrics (weekly basis)

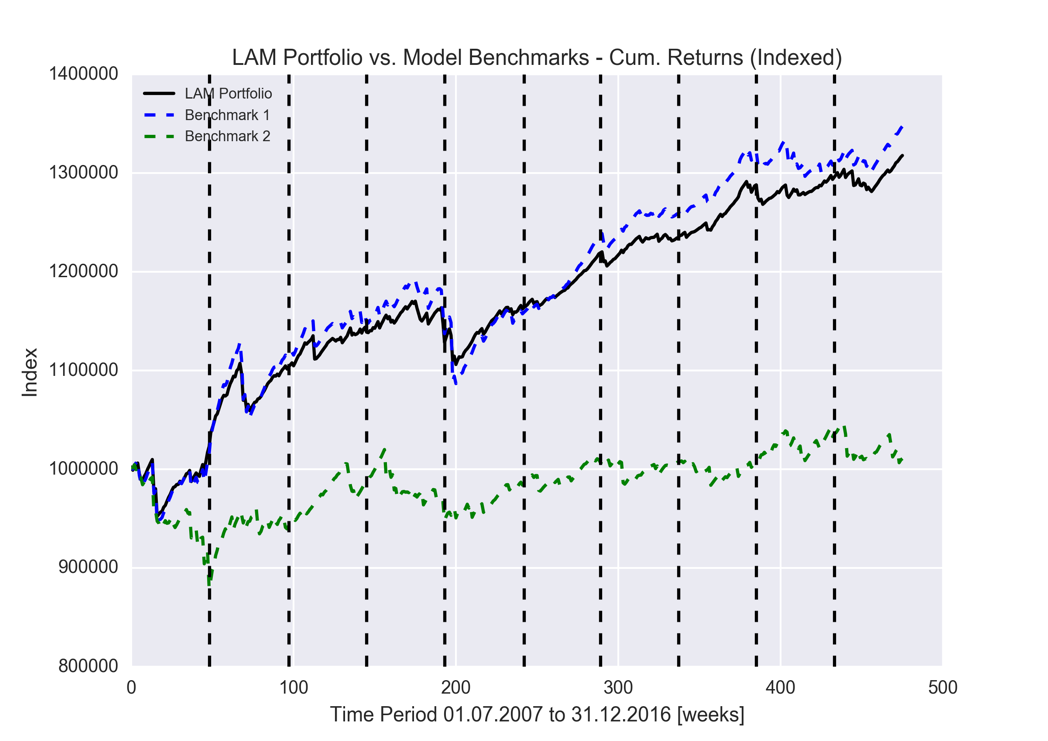


Figure ‑ – LAM Portfolio vs. Model Benchmarks – Cum. Returns

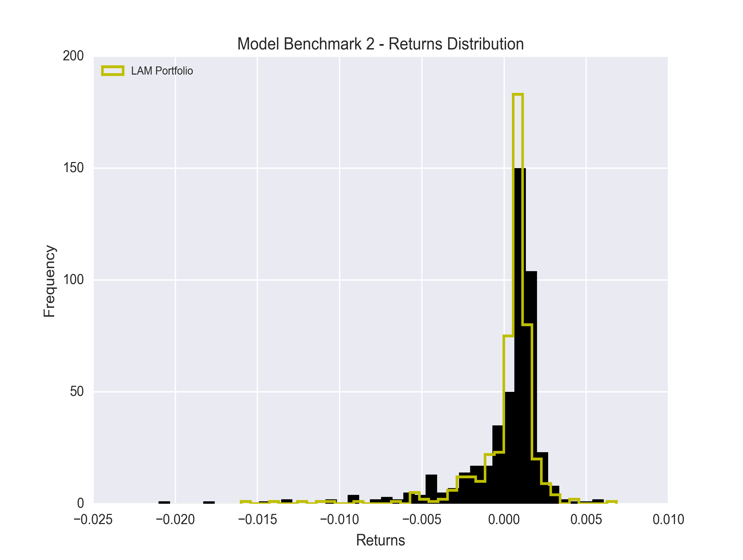
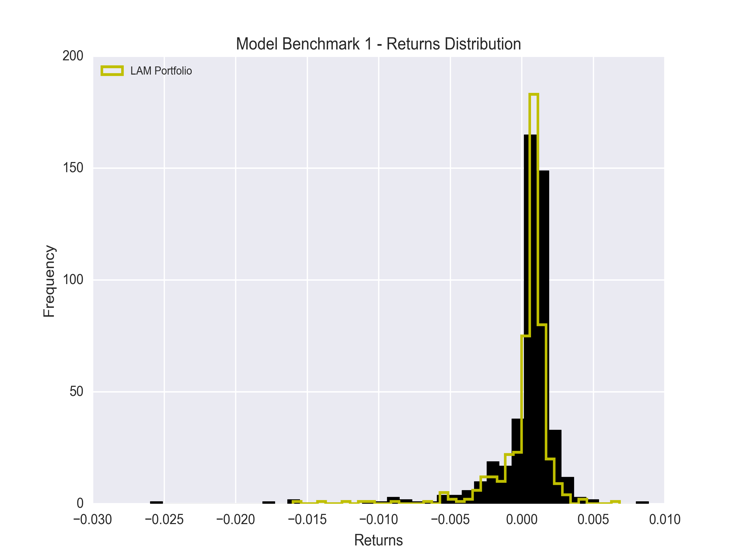


Figure ‑ – Model Benchmarks – Returns Distributions

## Portfolio vs. Market Benchmarks

Of particular interest when assessing the potential of a strategy is its performance against other investable sectors. For this purpose, a selection of four widely-employed indices is brought forward; *BARCCURR* for trading advisors in the currencies/commodities space, *BARCSYST* for systematic portfolios, *BARCBTOP* for the managed bond industry, and *MSCI World* for the equities sector (Appendix 7.4)[[41]](#footnote-42).

The equity lines presented in Figure 4‑8 indicate the *LAM Portfolio* features superior returns compared to all the benchmarks. Furthermore, the Profit Factor and Hit Ratio outperform the benchmarks (Table 4‑7), while only the *BARCCURR* index appears to benefit from a higher Sharpe Ratio, due to its lower Max. DrawDown and volatility, but lags in Cumulative Performance. As expected, the *MSCI World* index features a substantial drawdown of more than 50%, a contraction due to the 2007-2008 global financial crisis[[42]](#footnote-43); it has since more than doubled in absolute terms, albeit with a very high volatility.

To summarize, the *LAM Portfolio* does not appear to be significantly correlated to any of the presented benchmarks (Table 4‑7). Furthermore, with regards to the returns’ distributions and with the exception of the *BARCCURR* Index, all the benchmarks feature, as expected fatter tails. All in all, the *LAM Portfolio* features superior returns and better performance metrics than the examined counterparties.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LAM Portfolio** | **BARCCURR** | **BARCSYST** | **BARCBTOP** | **MSCI World** |
| Cum. Returns | +31.8% | +24.9% | +16.2% | +13.8% | +19.4% |
| Profit Factor | 2.29 | 2.13 | 1.20 | 1.16 | 1.12 |
| Hit Ratio | 75.2% | 61.7% | 50.0% | 53.3% | 53.3% |
| Max. DrawDown | -5.11% | -3.21% | -11.82% | -10.24% | -50.79% |
| Sharpe Ratio | 2.09 | 2.20 | 0.61 | 0.53 | 0.42 |
| Leverage | 1 | | | | |

Table ‑ – LAM Portfolio vs. Market Benchmarks Performance Metrics (weekly basis)

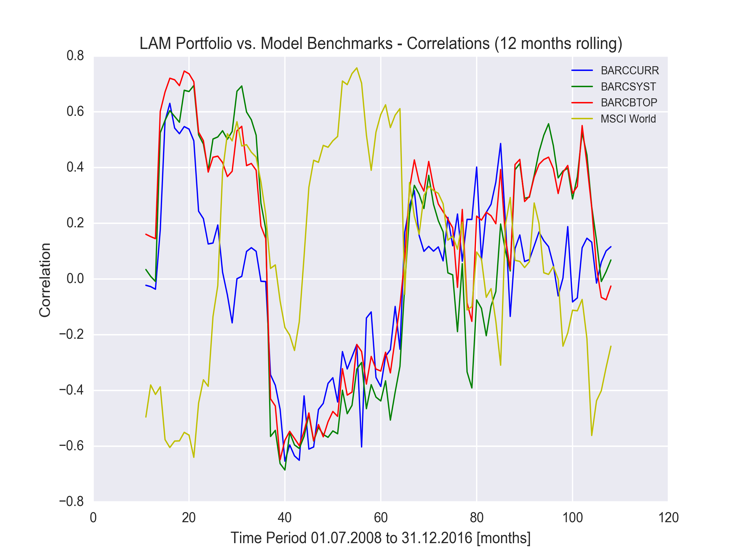
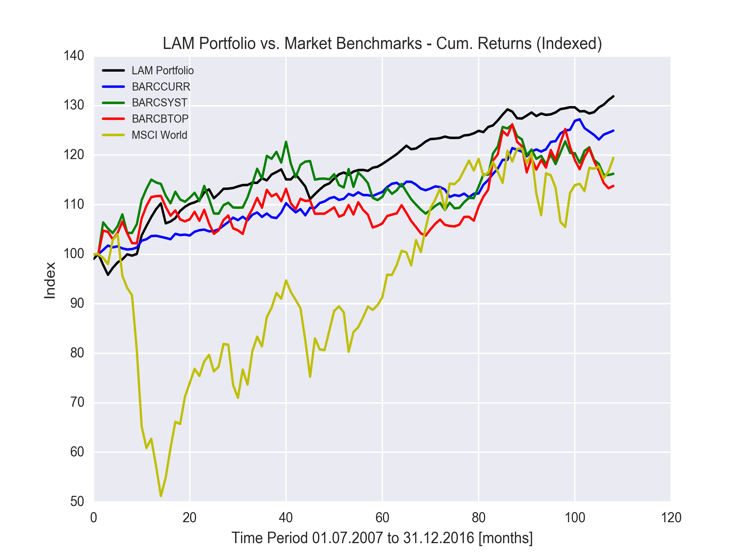
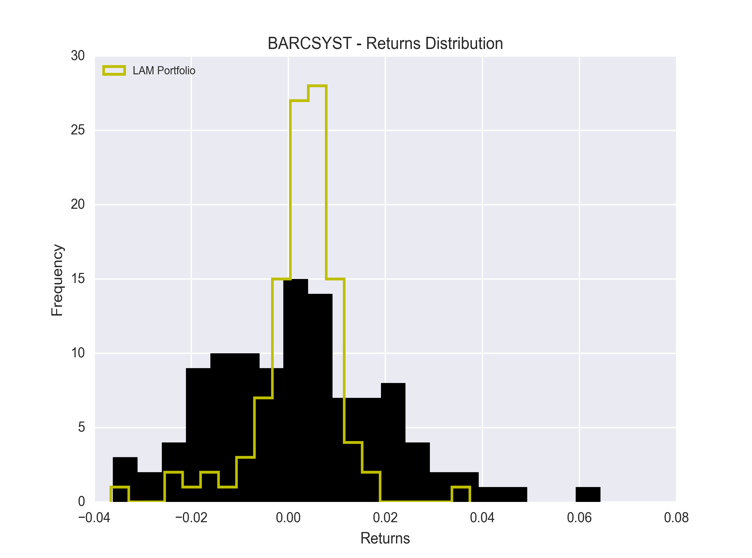
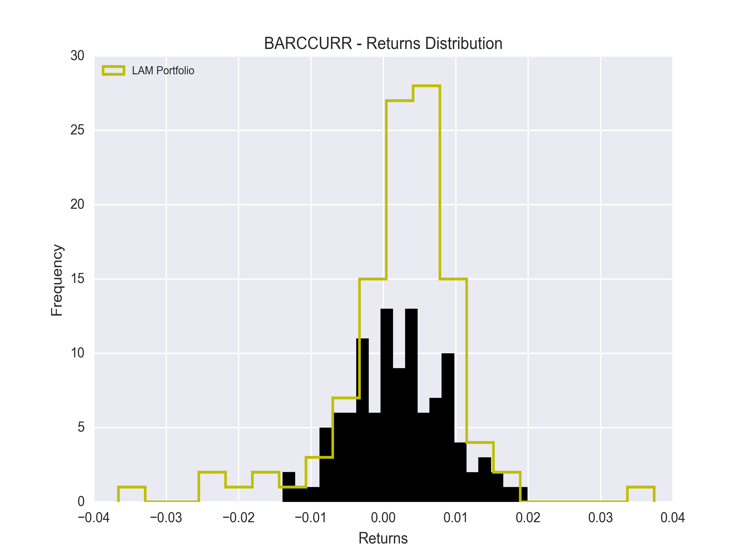


Figure ‑ – LAM Portfolio vs. Market Benchmarks – Cum. Returns / Correlations



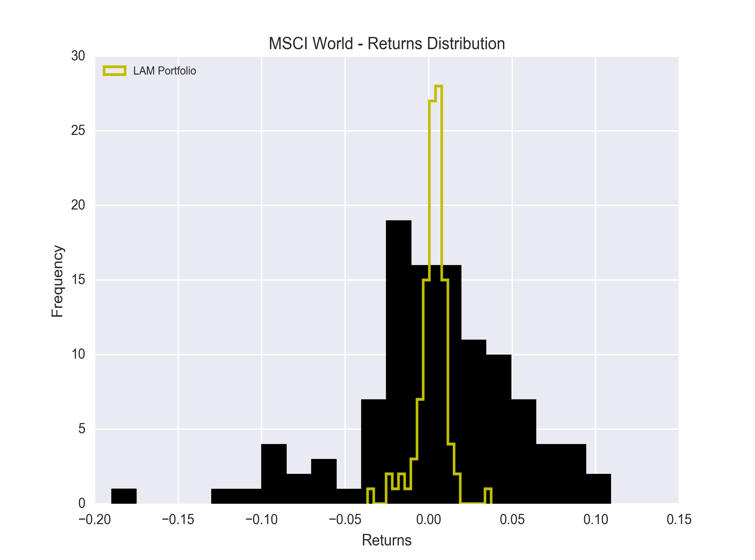
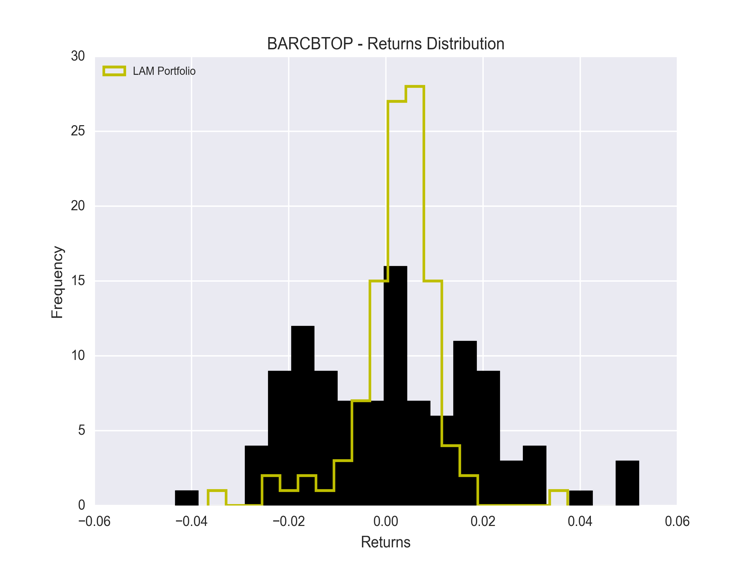


Figure ‑ – Market Benchmarks – Returns Distributions

# Conclusions

The purpose of this paper was to investigate the potential of an option-based portfolio derived from the positioning of market participants. In more detail, the goal was to assess the predictive power of the CoT and TFF Reports released by the CME every week, develop econometric models that improve upon a simplistic trend-following strategy, and implement an asset selection process that yields an investable portfolio.

Two econometric models were conceived (AR, GARCH) and trading strategies were constructed around them through appropriate optimization processes. For the AR Model, an optimization was conducted to yield an allocation between the forecasted positioning and the latest positioning information. Additionally, a filter was designed to exclude trades whereby the market positioning is close to even ( filter). For the GARCH Model on the other hand, a volatility filter was conceived, whereby trades are not taken if the forecasted variance is higher than historical averages. On an individual basis, significant improvements in performance metrics were observed for both models against the raw model (BASE Model) in both in-sample and out-of-sample environment. While the AR predictions broadly yield more accurate and less volatile returns, the GARCH model appears to perform better in certain circumstances. Given the number of combinations of assets, market participants and econometric models was in excess of 100, an appropriate selection model was implemented; the Limited Asset Markowitz model analyses the investment universe available, and selects five assets along with the respective allocations that yield the best Sharpe Ratio for the in-sample data. These parameters are then applied in out-of-sample period, re-balancing taking place at the end of every year. The results obtained yield the *LAM Portfolio*, whose Cumulative Return stands at +31.8% over a period of 10 years. This is shown to outperform a LAM portfolio compiled only with raw signals (*Benchmark 1*), and to be comparable to a, usually hard to beat, equally weighted LAM portfolio (*Benchmark 2*). Finally, a comparison of the *LAM Portfolio* against broadly employed market benchmarks evidenced the portfolio consistently outperforms the latter.

The analysis conducted throughout this paper evidenced not only the overall predictive power of the CFTC positioning data, but also allowed the identification of shortcomings and improvements of the methodologies employed. Given the results and procedures presented in this paper, the following extensions are proposed:

* The specification of the options analysed in this paper was arbitrary, and was derived from a real-life implementation of the BASE model. An optimization process can therefore be devised that yields the optimum duration and strike of the options traded for each asset and market participant.
* The option data collected through Bloomberg do not include any commissions payable for their stipulation. Financial counterparties usually include such costs in the premia paid or received by the client for the options, a feature not present in the Bloomberg data. Given the low trading frequency of the *LAM Portfolio* (maximum of 5 trades per week), commissions are not expected to have a significant impact on the overall performance, but would come in handy for the sake of completeness.
* Currencies and precious metals options discussed in this paper are OTC products; their pricing can therefore vary according to the market parameters (e.g. implied volatility) employed by a counterparty. Bloomberg appears to address this issue by monitoring these market parameters from a range of liquidity providers. Nonetheless, to eliminate any potential bias or erroneous data, options could be priced using CME Options. While the product offering of the CME is vast, it should be noted the it does not feature options for all the products included in the CFTC reports. An alternative would be to analyse and aggregate data from different venues (e.g. exchanges, banks, brokers etc.).
* The AR Model presented in Section 4.3 features an optimisation process yielding an optimal allocation between the forecasted return and the latest return. Since this process is carried out only once in the in-sample data, and not at the end of every year, it leaves a lot of room for improvement. By conducting the optimisation every year, the optimal allocation will be allowed to shift, therefore providing the AR Model with an additional degree of freedom. Furthermore, such improvement would align well with the yearly LAM Model rebalancing process.
* The implementation of the LAM Model features a LASSO regression. This is a penalized-regression model introduced by (Tibshirani, 1996) that simultaneously identifies and calculates the most important coefficients by employing a shorter sample period by betting on sparsity. While this brings many advantages to the particular selection process, it also features some shortcomings. For instance, if two predictors are highly correlated, LASSO can end up dropping one of them arbitrarily. Within the context of this paper, placebo tests could be conducted to ensure the process adds to the forecasting power of the LAM Model.
* Additional work can also be conducted on fine-tuning the LAM Model. A modification of the LAM Model could allow for a varying number of assets in the portfolio in each period. Furthermore, rebalancing could take place at shorter / longer intervals, or even at irregular intervals using, for instance, band rebalancing. Finally, the LAM Model could be allowed to appoint also negative allocations to its selections, effectively taking the opposite side of the trade than what the underlying strategy is suggesting.
* While efforts were made to collect as many data as possible, the time-intensive process required to do so through Bloomberg made obtaining option pricing information for all the assets not-feasible. To obtain a fully cross-asset portfolio, the current paper needs to expand into indices and treasuries; complete positioning data for these are available[[43]](#footnote-44).

Since implementing the above suggestions would require substantial effort, it is left for future research. While there are numerous extensions to the presented methodologies, the basic premise of this paper was to prove the forecasting power of the CFTC reports. This first step has been completed, leaving a level field for additional work to be conducted on this.

# Bibliography

Alexander, C. (2008). *Market Risk Analysis, Practical Financial Econometrics.* Chichester, West Sussex, England: John Wiley & Sons.

Babcock, B., & Williams, L. (1989). *The Dow Jones Irwin Guide to Trading Systems.* NY: Dow-Jones Irvin.

Barro, R. (1989). *Black Monday and the future of financial markets.* NY, NY, USA: Richard D Irwin.

Berger, D., & Turtle, H. (2012). Cross-sectional performance and investor sentiment in a multiple risk factor model. *Journal of Banking & Finance*(36), 1107-1121.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.

Brown, G., & Cliff, M. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*(11), 1-27.

Brown, S., & Goetzmann, W. (1995). Performance Persistence. *Journal of Finance, 50*(2), 679-698.

Brunnermeier, M., & Nagel, S. (2004). Hedgefunds and the technology bubble. *The Journal of Finance, 59*(5), 2013-2040.

Buchananan, W., Hodges, P., & Theis, J. (2001). Which way the natural gas price: an attempt to predict the direction of natural gas spot price movements using trader positions. *Energy Economics, 23*, 279-293.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance, 52*(1), 57-82.

CFTC. (2006). *Comprehensive Review of the Commitments of Traders Reporting Program.* NY: Commodity Futures Trading Commission.

Chan, W. (2003). Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics*(70), 223-260.

Chang, E. (1985). Returns to Speculators and the Theory of Normal Backwardation. *The Journal of Finance, 40*(1), 193-208.

Cook, J. (2012). *CFTC - Commitments of Traders Report Update.* CME Group, Research & Product Development. CME Group.

Cromwell, J., Hannan, M., Labys, W., & Terraza, M. (1994). *Multivariate Tests for Time Series Models.* Thousand Oaks: Sage Publications.

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). *Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?* London: Oxford University Press.

Fama, E. (1965). The Behaviour of Stock market prices. *Journal of Business, 38*, 34-105.

Friedman, M. (1953). The Case for Flexible Exchange Rates. *Essays in positive economics*, 157-203.

Hartzmark, M. (1987). Returns to Individual Traders of Futures: Aggregate Results. *The Journal of Political Economy, 95*(6), 1292-1306.

Hartzmark, M. (1991). *The Distribution of Large Trader Returns in Futures Markets: Theory and Evidence.* Chicago: University of Chicago.

Henriksson, R., & Merton, R. (1981). On Market Timing and Investment Performance. II. Statistical Procedures for Evauating Forecasting Skills. *The Journal of Business, 54*(4), 513-533.

Holderness, C. (2016). Problems Using Aggregate Data to Infer Individual Behavior: Evidence from Law, Finance, and Ownership Concentration. *Critical Finance Review*, 1-40.

Houthakker, H. (1957). Can Speculators Forecast Prices? *The Review of Economics and Statistics, 39*(2), 143-151.

Hsiao, C. (2005). Why panel data? *Singapore Economic Review, 50*(2), 1-12.

Leuthold, R. L., Garcia, P., & Lu, R. (1994). The Returns and Forecasting Ability of Large Traders in the Frozen Pork Bellies Futures Market. *The Journal of Business, 67*(3), 459-473.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91.

Miller, F., Vandome, A., J., & McBrewester, J. (2009). *Efficient-Market Hypothesis.* Alphascript Publishing.

Mirabile, K. (2016). *Hedge Fund Investing.* New Jersey, USA: John Wiley & Sons.

Notter, L. (2013). *Sentiment Strategies for Currency Trading.* St. Gallen: University of St. Gallen.

Sanders, D., Irwin, S., & Merrin, R. (2007). Smart Money? The Forecasting Ability of CFTC Large Traders. In D. Sanders, S. Irwin, & R. Merrin (Ed.), *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management* (p. 20). Chicago: Sanders, D.R.; Irwin, S.H.; Merrin, R.

Savona, P., Kirton, J., & Oldani, C. (2011). *Global Financial Crisis: Global Impact and Solutions.* Surrey, England, UK: Ashgate Publishing Limited.

Siegel, J. (1992). Equity Risk Premia, Corporate Profit Forecasts, and Investor Sentiment around the Stock Crash of October 1987. *The Journal of Business, 65*(4), 557-770.

Stewart, B. (1949). *An Analysis of Speculative Trading in Grain Futures.* United States Department of Agriculture. Commodity Exchange Authority.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society.*, 267-288.

Tu, J., & Zhou, G. (2008). *Being Naive about Naive Diversification: Can Investment Theory Be Consistently Useful?* Washington University, Olin School of Business. Washington: Washington University.

Upperman, F. (2006). *Commitments of Traders: Strategies for Tracking the Market and Trading Profitably.* New Jersey: John Wiley & Sons, Inc.

Wang. (2001). Investor Sentiment and Return Predictablity in Agricultural Futures Markets. *The Journal of Futures Markets, 21*(10), 929-952.

Wang, C. (2003). Investor sentiment, market timing, and futures returns. *Applied Financial Economics*(13), 929-952.

Yan, X. (2009). *Linear Regression Analysis: Theory and Computing.* Singapore: World Scientific.

Zhang, W., & Skiena, S. (2010). Trading Strategies to Exploit Blog and News Sentiment. *Fourth International AAAI Conference on Weblogs and Social Media* (pp. 375-378). NY: Fourth International AAAI Conference on Weblogs and Social Media.

# Appendix

## CFTC Current Reporting Levels



## Performance Metrics

The Sharpe Ratio, an industry standard for portfolio performance measurement, is a number that yields the risk-adjusted return of an asset or a portfolio. Developed by Nobel laureate William F. Sharpe, the Sharpe Ratio is the average return earned in excess of a risk-free rate per unit of total risk[[44]](#footnote-45):

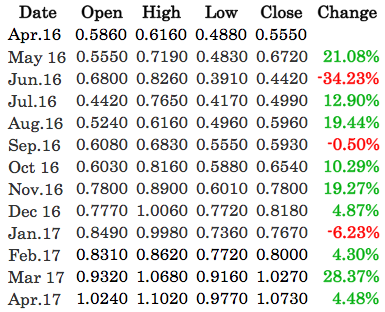
,where and are the expected portfolio and risk-free asset returns, while is the portfolio standard deviation.

Maximum DrawDown is defined as the maximum peak-to-trough decline of an investment; it therefore measures how sustained a portfolio’s losses can be:

,where is the portfolio’s peak value before the largest drop and is the lowest value before a new high.

## US 1-Year Bond Yield Data / Calculation

The Risk-Free rate was defined as the 12-month average of the US 1-Year Bond Yield. Below the historical monthly data employed for the calculation of the Risk-Free rate used in all portfolio calculations:



## Market Benchmarks

Find below the descriptions of the benchmarks employed[[45]](#footnote-46):

* The Barclay CTA Index (BBGID: BARCCURR) is an industry benchmark of representative performance of currency and commodity trading advisors.
* The Barclay BTOP50 Index (BBGID: BARCBTOP) attempts to replicate the composition of the managed bond industry with regard to overall market exposure and trading approach.
* The Barclay SYST (Systematic Traders) Index (BBGID: BARCSYST) is an equal weighted composite of managed programs whose approach is at least 95% systematic.
* The MSCI World Index is a broad global equity benchmark featuring large and mid-cap equity performances across 23 developed countries.

1. See (Carhart, 1997) and (Brown & Goetzmann, 1995). [↑](#footnote-ref-1)
2. Mutual funds are often compelled to publish their trades and performances aiming to attract new investors. [↑](#footnote-ref-2)
3. Monday, October 19, 1987, also referred as a „Black Monday” in finance, was one of the largest one-day percentage declines for stock exchanges around the world (Barro, 1989). [↑](#footnote-ref-3)
4. Updated research points to 75%-90% negative performances in the futures market (Babcock & Williams, 1989). [↑](#footnote-ref-4)
5. This conclusion is in line with the more recent research of (Leuthold, Garcia, & Lu, 1994) who reported significant profits for all large traders analysed. [↑](#footnote-ref-5)
6. (Henriksson & Merton, 1981) [↑](#footnote-ref-6)
7. In his paper (Chang, 1985) distinguished between rewards for forecasting abilities and for taking risk (risk premia). (Hartzmark M. , 1991) [↑](#footnote-ref-7)
8. Of interest, (Hartzmark M. , 1987) results indicated that 0.1% of the large hedgers earned 58% of the total profits in the period analysed. [↑](#footnote-ref-8)
9. In (Hartzmark M. , 1991) study, consistent forecasting abilities were defined as consistent profits. It must be noted that most of the results were not statistically significant. [↑](#footnote-ref-9)
10. The results of (Wang, 2001) were based on six agricultural futures markets. [↑](#footnote-ref-10)
11. As discussed in the Section 2.1, the CoT reports include information spanning across asset classes (commodities, currencies, indices, fixed income instruments etc.). [↑](#footnote-ref-11)
12. The data are produced every Tuesday, and released to the public every Friday. [↑](#footnote-ref-12)
13. The TFF report is also delivered in the same formats as the CoT; "futures only" and "futures and options". [↑](#footnote-ref-13)
14. Markets covered include commodities, indices, currencies, treasuries, bonds etc., across national and local exchanges (e.g. NYSE, NYMEX, CBOE, KCBT etc.). [↑](#footnote-ref-14)
15. The investment process is built on trading options priced on Mondays at GMT 14:00 and expiring the following Monday at GMT 14:00. [↑](#footnote-ref-15)
16. Public holidays are replaced by the next available trading day (usually Tuesday). [↑](#footnote-ref-16)
17. The ADF test employs an autoregressive model and optimizes an information criterion across multiple different lag values. The null hypothesis (H0) states the time series can be represented by a unit root that is not stationary (i.e. time series has a time-dependent structure), while the alternate hypothesis (H1), is that it is stationary (i.e. time series has no time-dependent structure) (Cromwell, Hannan, Labys, & Terraza, 1994). [↑](#footnote-ref-17)
18. ADF test was applied on 134 complete asset’s time series; the null hypothesis (H0) was rejected at the 1% significance level for 82.1% of them. [↑](#footnote-ref-18)
19. Rejecting the null hypothesis at the 1% significance level implies that the probability the ADF test results are a statistical fluke are low. [↑](#footnote-ref-19)
20. The (P)ACF gives the (partial) correlation of a time series with its own lagged values. [↑](#footnote-ref-20)
21. The Delta of an option is a value that estimates the rate of change in the price of an option given a 1 point move in the underlying asset. Using a higher delta when pricing an option increases the distance of the strike from the spot, thus making potential execution of the option less likely; subsequently the likely premium received is also reduced. For this paper a 20-delta strike was arbitrary chosen, in line with a real life implementation of this strategy. [↑](#footnote-ref-21)
22. MATLAB utilized for model development and optimization, while Python employed for cross-validation of results, portfolio construction and plotting. [↑](#footnote-ref-22)
23. As an example, if in Dealers have a net long (short) position in an asset at any point in time, the model will opt to sell the put (call) options. [↑](#footnote-ref-23)
24. Times series exhibit strong 1-period autocorrelations. [↑](#footnote-ref-24)
25. In the 110 equity lines analysed, the 10 largest losing trades are generated with a probability of 73.4% from positioning data in the area. [↑](#footnote-ref-26)
26. The data (*N* data points) are split in half and the optimization is run on the in-sample part of the data (*0* to *N/2*). Subsequently the resulting allocation is employed in the out-of-sample data (*N/2* to *N*). [↑](#footnote-ref-27)
27. The CoT report covers Commercials and Non-Commercials, while the TFF data reports Leverage Money, Dealers and Asset Managers (Section 2.1). [↑](#footnote-ref-28)
28. The first consists of the same portfolio but with equal, instead of optimized, allocations to the constituents, while the second is a LAM-derived portfolio constructed only with the BASE signals (Section 4.5.3). [↑](#footnote-ref-29)
29. Each equity line is based on trading a USD 1m notional on the resulting signals, as per Section 3.3 [↑](#footnote-ref-30)
30. Derived from the 12-month average of the US 1-Year Bond Yield between April 2016 and March 2017) (Appendix 7.3). [↑](#footnote-ref-31)
31. A randomized selection algorithm was designed in Python that picked the following combinations of assets/market participants: GBPUSD Commercial, USDCAD Asset Manager, USDJPY Dealer, AUDUSD NonCommercial, EURUSD Leverage Money (the algorithm features appropriate constraints to yield different assets and participants). [↑](#footnote-ref-32)
32. The Hit Ratio is defined as the number of positive trades to the overall number of trades. [↑](#footnote-ref-33)
33. The Profit Factor is defined as the total profit divided by the total loss generated from the trades undertaken. [↑](#footnote-ref-34)
34. As outlined in Section 4.3, the AR Model, as opposed to the BASE Model, is not always in the market, since the filter yields potential neutral periods. For the presented equities these neutral periods amounted to, on average, of the investment horizon. [↑](#footnote-ref-35)
35. Comparing the GARCH Model against either the BASE or GARCH Model, a reduction in the volatility of the Japanese Yen and British Pound equities is observed. [↑](#footnote-ref-36)
36. Performance metrics based on a leverage of 1, as described in the introduction of Section 4.5 [↑](#footnote-ref-37)
37. Note that since the portfolio is composed of multiple investments, the Hit Ratio in this context is defined as the number of positive weeks to the total number of traded weeks. [↑](#footnote-ref-38)
38. Of the 10 years examined, March is the worst performing month in 70% of the cases. [↑](#footnote-ref-39)
39. (DeMiguel, Garlappi, & Uppal, 2007) examined 14 models of optimal allocation relative to a “naïve” allocation benchmark portfolio; in their results, no single model consistently delivered a Sharpe Ratio or CEQ return higher than the benchmark portfolio. (Tu & Zhou, 2008) performed similar tests on a wider range of theory-based portfolio strategies, obtaining extremely similar results. [↑](#footnote-ref-40)
40. The universe of strategies derived from the BASE Model is provided to the LAM Model, which selects assets and allocates weights with the same criteria as in the *LAM Portfolio*. [↑](#footnote-ref-41)
41. Note that all the indices selected (except for the MSCI World) are produced on a monthly basis; the performance metrics in this section are therefore derived from the monthly returns. [↑](#footnote-ref-42)
42. Also known as the global financial crisis, economists consider this to be the worst financial crisis since the Great Depression in the 1930s. It originated as a crisis in the subprime mortgage market, was magnified by the excessive exposure of the banks, and resulted in extended bail-outs of financial institutions and numerous monetary and fiscal policies to prevent a collapse of the world’s financial system (Savona, Kirton, & Oldani, 2011). [↑](#footnote-ref-43)
43. A total of 122 complete positiong datasets are available accross teh CoT and TFF reports. In this paper 22 of these were examined. [↑](#footnote-ref-44)
44. In the Sharpe Ratio calculation, risk is commonly measure by means of volatility. [↑](#footnote-ref-45)
45. Data collected from Bloomberg and Barclays Bank PLC., through BarclayHedge. [↑](#footnote-ref-46)