QUANTIFICATION OF THE RELATIONSHIP BETWEEN THE SIZE OF A MUTUAL FUND AND ITS RISK. ASSESSMENT OF THE ASYMMETRY IN PROFITS AND LOSSES THROUGH THE SORTINO RATIO.

Autor: Gonzalo Hervás Múgica
Director: Sara Lumbreras Sancho
Director: Ignacio Cervera Conte

Madrid

Junio 2016

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESINAS O MEMORIAS DE BACHILLERATO

1°. Declaración de la autoría y acreditación de la misma.
El autor D. Gonzalo Hervás Múgica
DECLARA ser el titular de los derechos de propiedad intelectual de la obra:
Quantification of the relationship between the sie of a mutual fund and its risk. Assessment of the assymetry in profits and losses through the Sortino ratio,
que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

2°. Objeto y fines de la cesión.
Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

3°. Condiciones de la cesión y acceso
Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:
  a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar "marcas de agua" o cualquier otro sistema de seguridad o de protección.
  b) Reproducirla en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
  c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
  d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
  e) Asignar por defecto a estos trabajos una licencia Creative Commons.
  f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente).*

4°. Derechos del autor.
El autor, en tanto que titular de una obra tiene derecho a:
  a) Que la Universidad identifique claramente su nombre como autor de la misma
  b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
  c) Solicitar la retirada de la obra del repositorio por causa justificada.
  d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5°. Deberes del autor.
El autor se compromete a:
  a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
  b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
  c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que

pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.

d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6°. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

➢ La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.

➢ La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusive del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.

➢ La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.

➢ La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a ...16... de ....JUNIO............. de .2016

ACEPTA

Fdo...

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título:

"Cuantificación de la relación entre el tamaño de un fondo y su riesgo.

Valoración de la asimetría en pérdidas y ganancias a través del ratio Sortino"

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2º Máster en Ingeniería Industriales de mi autoría, original e

inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no

es plagio de otro, ni total ni parcialmente y la información que ha sido tomada

de otros documentos está debidamente referenciada.

Fdo.: Gonzalo Hervás Múgica   Fecha: 16./ 06./ 2016

Autorizada la entrega del proyecto

LOS DIRECTORES DEL PROYECTO

Fdo.: Sara Lumbreras Sancho   Fecha: 16./ 06./ 2016

Fdo.: Ignacio Cervera Conte   Fecha: 16./ 06./ 2016

Vº Bº del Coordinador de Proyectos

Fdo.: Jaime de Rábago Marín   Fecha: ....../ ....../ ......

QUANTIFICATION OF THE RELATIONSHIP BETWEEN THE SIZE OF A MUTUAL FUND AND ITS RISK. ASSESSMENT OF THE ASYMMETRY IN PROFITS AND LOSSES THROUGH THE SORTINO RATIO.

Autor: Gonzalo Hervás Múgica
Director: Sara Lumbreras Sancho
Director: Ignacio Cervera Conte

Madrid

Junio 2016

CUANTIFICACIÓN DE LA RELACIÓN ENTRE EL TAMAÑO DE UN FONDO Y SU RIESGO. VALORACIÓN DE LA ASIMETRÍA EN PÉRDIDAS Y GANANCIAS A TRAVÉS DE LA RATIO SORTINO.

Autor: Hervás Múgica, Gonzalo.

Director: Lumbreras Sancho, Sara.

Director: Cervera Conte, Ignacio.

Entidad Colaboradora: Universidad Pontificia Comillas – ICAI.

RESUMEN DEL PROYECTO

Introducción

El objetivo del proyecto es analizar si el rendimiento de un fondo de inversión se encuentra influenciado por su tamaño.

Los fondos de inversión tienen un registro histórico de rápido crecimiento cuando los inversores ven oportunidades ventajosas en ellos, causando a veces problemas para los administradores al crecer demasiado rápido o demasiado grande. Estos problemas podrían estar relacionados con el rendimiento. Precisamente este proyecto busca encontrar evidencia para determinar una relación entre el tamaño y el rendimiento de los fondos.

Los fundamentos del proyecto serán el análisis estadístico aplicado a diferentes medidas de rendimiento y riesgo de los fondos. La información utilizada para el estudio procede de la base de datos CRSP de la Universidad de Wharton, siendo una de las mayores bases de datos para la investigación de fondos de inversión y del mercado de valores.

Estado del Arte

A lo largo de los últimos años, se han realizado numerosos estudios que han tratado de encontrar una correlación entre el tamaño de un fondo y su rendimiento. Los resultados de estos estudios cubren todas las opciones posibles, dejando un gap en la literatura el cual este estudio pretende rellenar, con un resultado concluyente.

Los estudios académicos que han intentado encontrar una relación entre el tamaño de los activos y la rentabilidad de los fondos, aún no han sido capaces de dar un resultado común. Es cierto que gran parte de ellos coinciden en una relación negativa entre ambas variables, dando una primera impresión de que es el resultado hacia donde tiende este

estudio. Además, al haber evidencia estadística y matemática que apoya cada uno de estos estudios previos, este proyecto tendrá al resto en cuenta y tratará de dar la mejor conclusión posible, llenando el gap en la literatura y con el objetivo de descubrir si crecer en tamaño tiene o no algún efecto sobre las medidas de riesgo y rendimiento.

Descripción del modelo

La herramienta estadística a través del cual se analizarán los datos es STATA, que siendo una herramienta apropiada para trabajar con datos de panel, facilita el hallazgo de la relación, si la hay, entre el tamaño de un fondo y su rendimiento. El modelo explica la relación entre la variable dependiente, en este caso el rendimiento del fondo, y la variable independiente, que es el tamaño. La variable que representa el rendimiento es el Sortino, siendo esta la primera vez que se usa en un proyecto de investigación y una de las principales particularidades de este proyecto. Es una modificación del ratio Sharpe que tiene en cuenta la volatilidad más perjudicial del fondo, en lugar de toda ella. El ratio de Sharpe es una medida de rendimiento utilizada para representar el rendimiento de un portfolio determinado y tiene en cuenta el retorno extra por encima de la media del mercado y la desviación estándar de estos retornos. Por otro lado, la variable que representa el tamaño de los fondos es el valor liquidativo total, y se calcula como el total de activos de cada fondo, menos los pasivos totales a final de mes, dado en millones de dólares.

En cuanto al modelo de regresión utilizado para encontrar la correlación entre las dos variables estudiadas, se aplica la regresión de datos de panel. La regresión de datos de panel, al tratar al mismo tiempo la dimensión transversal como la temporal, permite tener en cuenta circunstancias que no se pueden observar o medir como factores culturales o diferentes estilos de inversión que se haya de forma implícita en los fondos. La regresión de los datos de panel también es útil para medir variables que cambian a lo largo del tiempo, pero no en todas las entidades, siendo esto muy relevante en un sector cíclico, como lo es el mercado de valores.

Análisis

Las primeras conclusiones extraídas a partir de los primeros resultados de la regresión indican que no existe una correlación lineal entre el tamaño y el rendimiento de los fondos. El análisis no ha mostrado evidencia suficiente para apoyar la relación entre ambas variables.

El siguiente gráfico muestra una representación del Sortino y del NAV, para así poder ver una representación real de los datos:
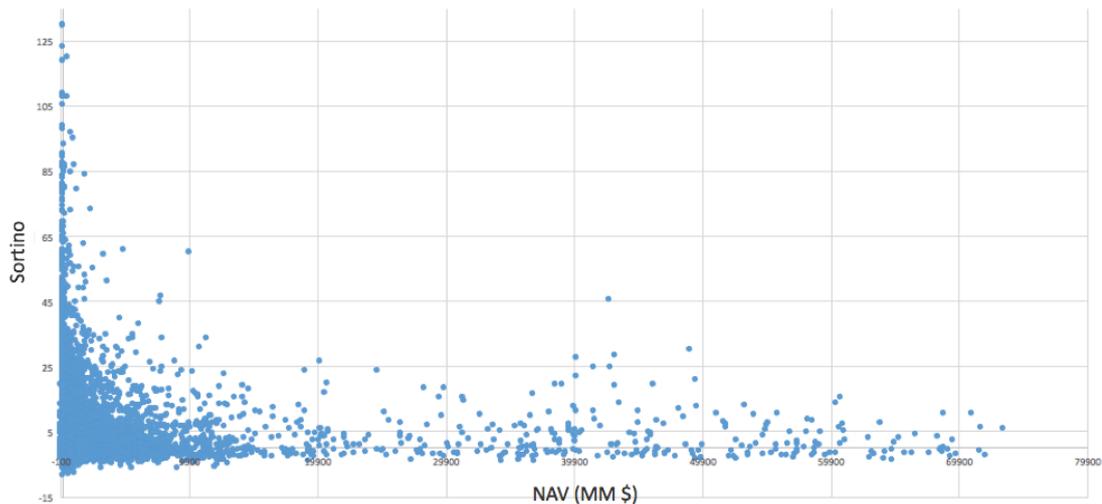

Figure 1: Sortino Vs. NAV

Análisis alternativo

Como los resultados de la regresión con datos de panel no muestran ninguna correlación lineal entre el tamaño y el rendimiento, se han explorado otros enfoques con el fin de evaluar esta relación.

Los distintos análisis son los siguientes:

1. Segmentación por tamaño

El objetivo de este análisis alternativo es el estudio de la muestra, dividiéndola en dos tamaños de fondos distintos, para ver si existe alguna correlación entre el rendimiento y el tamaño en cada una de las dos muestras por separado.

La base de datos se divide en fondos pequeños y grandes. Para la regresión de la muestra pequeña se llevan a cabo dos análisis diferentes. En primer lugar, el análisis con regresión lineal con datos de panel que se llevaba realizando hasta ahora, y en segundo lugar, una regresión cuadrática para ver si este nuevo modelo puede explicar mas datos que la anterior. Aunque no se encuentra correlación entre el tamaño y el rendimiento ya que los resultados no son estadísticamente significativos, el modelo se mejora ligeramente con el modelo cuadrático.

iii

2. Análisis individual

El siguiente análisis alternativo estudia cada fondo de manera individual. El objetivo de este análisis es ver cómo los fondos de la muestra se comportan como individuos, no como un grupo. Esto se realiza mediante la selección de siete fondos al azar de diferentes tamaños y se ejecuta un modelo de regresión lineal simple, sin datos de panel, a cada uno de ellos para estudiar así su comportamiento individual.

Tras analizar los datos, no se encuentra una tendencia clara en los resultados.

3. Análisis por periodos

El objetivo de este análisis consiste en dividir la muestra entre los periodos económicos más importantes, para ver si existe una correlación significativa entre el tamaño y el rendimiento. Por lo tanto, los datos se dividen en 5 muestras diferentes y se estudian de forma independiente con datos de panel.

Finalmente, los resultados no muestran una correlación entre el tamaño y el rendimiento en ningún período.

4. Análisis de dispersión

El último análisis alternativo busca explicar la tendencia de la dispersión del Sortino a medida que varía el tamaño de los fondos.

Los datos se agrupan en intervalos de tamaño similar (approx.1000 Sortinos en cada uno), contabilizando únicamente el valor máximo y mínimo de las observaciones en cada uno de los intervalos. De esta manera se tiene un valor máximo y un valor mínimo de cada intervalo, y analizando como estos valores se acercan o no a cero a medida que el tamaño de los fondos crece, el análisis serviría como evidencia para apoyar la hipótesis de que a medida que aumenta el tamaño del fondo, se reduce la dispersión del Sortino.

Tras analizar los datos, a medida que el tamaño del fondo crece, los valores mínimos del Sortino aumentan mientras que los valores máximos disminuyen, lo que determinaría que la dispersión del Sortino, efectivamente, disminuye a medida que el tamaño de los fondos (NAV) aumenta.

Conclusión

Como conclusión general del proyecto, después de analizar los datos, no se encuentra evidencia estadística para apoyar una relación lineal entre el tamaño de los fondos y su rendimiento, representado por el Sortino. Los resultados del análisis no han mostrado evidencia suficiente para justificar la correlación entre ambas variables.

Con respecto a esta medida de rendimiento, lo único que se puede afirmar, después de haber realizado numerosos análisis alternativos, es que la dispersión del Sortino para fondos pequeños es mayor que para los grandes fondos. Estos resultados apoyarían la idea de que los fondos pequeños se encuentran menos diversificados. Mientras los fondos más grandes tienen mecanismos de diversificación de riesgos y control más estándar, lo fondos pequeños son más diferentes entre sí. Estos resultados además tienen importantes usos prácticos para los gestores de fondos y los inversores. Para los inversores con aversión al riesgo podría ser mejor invertir en fondos pequeños, si conocen bien sus mecanismos de control de riesgo y estilo de inversión. Si no los conocen lo suficiente, el riesgo podría ser menor si se invierte en un fondo grande.

QUANTIFICATION OF THE RELATIONSHIP BETWEEN THE SIZE OF A MUTUAL FUND AND ITS RISK. ASSESSMENT OF THE ASYMMETRY IN PROFITS AND LOSSES THROUGH THE SORTINO RATIO.

PROJECT SUMMARY

Introduction

The objective of this project is analyze whether the performance of an investment fund is influenced by its size.

Mutual funds have a historical record of expanding rapidly when investors see advantageous opportunities in them, causing sometimes problems for managers when growing too fast or too big. The type of problems mutual funds might have would be related to performance and this project precisely seeks to find evidence of a relationship between the size and the performance of the funds.

The fundamentals of the project will be data analysis and statistics applied to different measurements of fund performance and risk in finance. The data used for the project is granted by the CRSP Wharton University Database, which holds some of the biggest historical databases for stock market research. Because of this, the evidence on which this study is based is wider than the vast majority of the literature research available.

State of the Art

There have been numerous studies that have tried to find a correlation between the size of a fund and its performance. Some of them agree in some aspects, but most of the results cover all possible options, leaving a gap for a conclusive result which is what this study aims.

Academic studies regarding a relationship between asset size and fund performance have yet not been able to give a conclusive result. It is true that a big number of studies agree on a negative relationship between both variables, giving a first impression that this might be where our study will be headed. It is true that there is statistical and mathematical evidence supporting each one of the studies, therefore this project will take all the evidence into account and will try to give the best conclusion possible, filling a literature gap with the objective of finding out if growing in size has any effect on risk measures.

Description of the model

The statistical tool through which the data will be analysed is STATA, allowing us to justify the relationship, if there is any, between the size of a fund and its performance. The model will find the relationship between the dependent variable, which would be the performance measure that is going to be explained by the model, and other independent variables or predictors, where the size of the fund will be included. The variable representing the performance is the Sortino ratio, being the first time used in a research and one of this project's main distinctions. It is a modification of the Sharpe that takes into account the downside volatility instead of all of it. The Sharpe ratio is a performance measurement used to represent the performance of a given portfolio and takes into account the average return in excess of the risk free rate and the standard deviation of the returns. On the other hand, the variable representing the size of the funds is the total Net Asset Value, and it is calculated as the total assets of each fund minus the total liabilities as of month-end, reported in millions of dollars.

Regarding the data regression model required to find the correlation between the two variables studied, this project applies Panel Data Regression Models. Panel Data will allow us to take into account circumstances that cannot be observed or measured like cultural factors or different investments styles across funds. Panel Data is also useful to measure variables that change over time but not across entities, being a very important aspect in a cyclical sector such as the stock market.

Analysis

The first conclusions that are drawn from the first regression results state that there is no correlation between the size and the performance of the funds. The analysis has not shown enough evidence to support the relationship between both variables as the indicators tell that the model doesn't explain the data. This would support the opinion of those authors that say that the performance of a fund is independent from its size.

The following graphic shows a representation of the Sortino ratio against the NAV, to see an actual representation of the data:
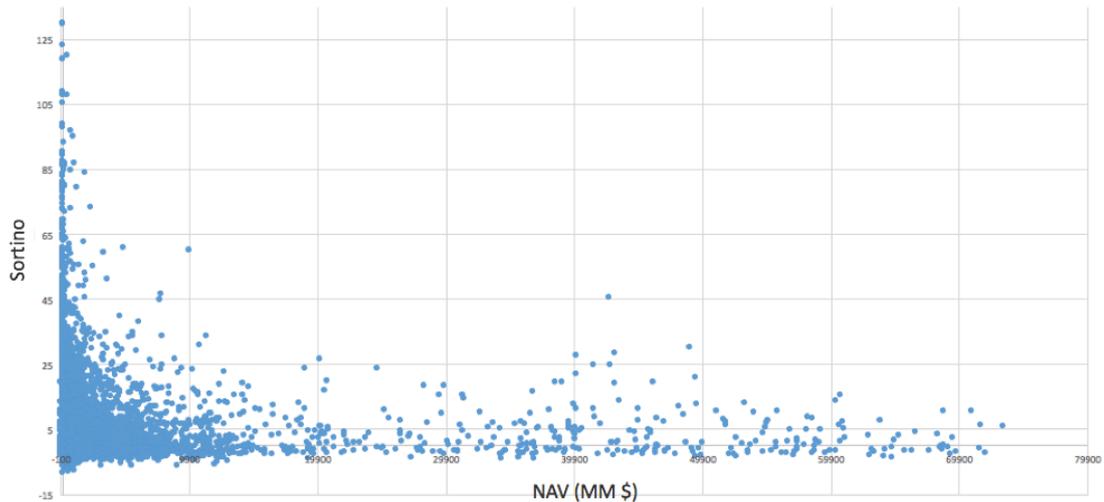
Figure 2: Sortino Vs. NAV

Alternative Analysis

As the results from the panel data regression show no linear correlation between size and performance, several other exploratory approaches were performed in order to assess that the relationship is indeed nonexistent.

The different analyses in addition to the one already done are the following:

1. Segmentation based on size

The objective of this alternative analysis is to study two different fund sizes independently, to see if any correlation between performance and size is found on each size, as separate samples.

The sample is divided into small and big funds. For the small funds regression, two different analyses are done. First, the regular linear regression analysis with panel data and random effects, and second, a quadratic regression to see if this new model can explain more than the linear one. Although no correlation was found between size and performance as the results are statistically insignificant, the model is slightly improved with the quadratic model.

ix

2. Individual Analysis

The next alternative analysis studies each fund individually. The objective of this analysis is to see how the funds of the sample behave as individuals, not as a group. This is done by selecting seven random funds of different sizes and running a simple linear regression model, with no panel data, to each one of them and study their individual behavior.

After analyzing the data, no clear trend was found in the results.

3. Period Analysis

The objective of this analysis if to see if by dividing the sample in the most important financial periods, there is a significant correlation between size and performance. Therefore, the data will be divided into 5 different samples and studied independently with 5 different regression analysis with panel data.

The results show no correlation between size and performance in any period.

4. Dispersion-Range Analysis

The last alternative analysis would like to explain the tendency of the Sortino's dispersion as the size of the funds changes.

The data is organized in intervals of similar size (approx.1000 Sortinos in each), the Maximum and Minimum value of the observations are obtained from each bin. This way we would have a maximum and a minimum value of each interval, and by analyzing how they come closer to zero as size grows, the analysis could work as evidence to support the idea that the dispersion is reduced as fund's size is increased.

As size grows, the minimum values of the Sortino increase while the maximum values decrease, which explains that the Sortino decreases as the size of the funds (NAV) increases.

Conclusions

As a general conclusion of the project, after analyzing the data, no statistical evidence was found to support a linear relationship between size and performance, represented by the Sortino ratio. The indicators of the analysis have not shown enough evidence to support the relationship between both variables.

Regarding this performance measure, the only think than can be stated after having done numerous alternative analyses, is that the dispersion of the Sortino for small funds is higher than for big funds. This evidence would support the idea that small funds are less diversified. While biggest funds have diversification and risk control mechanisms more standard, small funds might be more different among them. These results have important practical uses for portfolio managers and investors. For risk averse investors it might be better to invest in small funds, if they know well the investment style and risk control mechanisms of the fund. If they don't completely understand or now how the fund operates, the risk might be lower when investing in a big fund.

Index

Figure Index

Table Index

1. Introduction and Project approach

The objective of this project is analyze whether the performance of an investment fund is influenced by its size.

Mutual funds have a historical record of expanding rapidly when investors see advantageous opportunities in them, causing sometimes problems for managers when growing too fast or too big. The type of problems mutual funds might have would be related to performance, as when fund assets increase in a short period of time it might be more difficult for the fund to maintain its investment style. Transaction costs would increase and other problems, which will be deeply explained throughout the project, might appear as the asset size of the fund increases. The fundamentals of the project will be data analysis and statistics applied to different measurements of fund performance and risk in finance.

There have been numerous studies that have tried to find a correlation between the size of a fund and its performance. The conclusions of these studies are inconclusive as each one of the differ from the rest with well justified arguments, as well as statistical and mathematical support.

What makes this study different from the rest?

First of all, it has taken into account the knowledge brought by these studies into the financial sector, starting the project from a good knowledge status and not from scratch.

In addition to the performance measurements used in the past, such as the volatility or the Sharpe ratio, it is the first time the Sortino ratio is used being one of this project's main distinctions. The advantages of using this metric will be discussed in detail in this document.

Regarding the data regression required to find the correlation between the two variables studied, this project applies Panel Data Regression Models. Panel Data will allow us to take into account circumstances that cannot be observed or

measured like cultural factors or different investments styles across funds. Panel Data is also useful to measure variables that change over time but not across entities, being a very important aspect in a cyclical sector such as the stock market.

The data used for the project is granted by the CRSP Wharton University Database, which holds some of the biggest historical databases for stock market research. Because of this, the evidence on which this study is based is wider than the vast majority of the literature research available.

Finally, the statistical tool through which the data will be analysed is STATA, allowing us to justify the relationship, if there is any, between the size of a fund and its performance.

2. Literature Review (State of the Art)

Many economists have tried to find a correlation between the size of an investment fund and its performance. Throughout numerous studies there has been no single answer to this relationship. Some of them agree in some aspects, but most of the results cover all possible options, leaving a gap for a conclusive result which is what this study aims.

Academic studies regarding a relationship between asset size and fund performance have yet not been able to give a conclusive result. It is true that a big number of studies agree on a negative relationship between both variables, giving a first impression that this might be where our study will be headed, being this the most repeated result since 1990. It is true that there is statistical and mathematical evidence supporting each one of the studies, therefore this project will take all the evidence into account and will try to give the best conclusion possible, filling a literature gap with the objective of finding out if growing in size has any effect on risk measures.

The main results observed among the existing literature are the following:

- Linear-positive relationship: authors such as Amenc et al (2004) and Liang (1999) state that as the size of an investment fund increases, so does performance. The correlation between both two variables is linear-positive, therefore small funds are in competitive disadvantage in comparison to the biggest ones. To support their study, they give statistical and mathematical evidence, as well as descriptive explanations regarding economies of scale and possible operational advantages that the mutual funds with the biggest names in the sector might have.

- Quadratic - concave: authors such as Amman, Moerth (2005), Getmansky (2004), or Indro et al (1999) state that there is a specific size for which the performance of an investment fund is optimal. The relationship between the two variables is quadratic-concave, so the performance of a fund is expected

to grow as its size increases until an optimal point at which performance begins to decrease as the size continues to increase. According to these authors, funds must reach a determined size where their costs of trading and acquiring information is justified but once they surpass that optimal asset size, marginal benefits become negative.

- Linear-negative relationship: authors such as Agarwal et al (2004) and Cheng et al (2004), support that as the size of an investment fund increases, its performance decreases. Smaller funds will perform better due to management, organizational, and liquidity diseconomies of scale existing in bigger funds.

- No correlation: Gregoriou and Rouah (1994) and Clark (2003), state that there is no relationship between both variables due to lack of evidence.

A first conclusion from the given studies is that there is no unique answer to the fund size-performance relationship, leaving space for another study that might resolve this question in a more accurate way.

The reason why these studies have different results, could be explained by numerous factors: the number and type of funds used and their investment style, the different ratios used to explain the performance and the risk, and the different regression models used to find the correlation between the variables studied. These factors, among others, will be deeply explained in the following pages as understanding them is essential for the study and its interpretation.

2.1   What is an Investment Fund and what are their investment styles?

An investment fund is a source of capital that belongs to a big number of investors that use that capital to invest in securities at the same time investors monitor and keep the ownership of their shares. These funds give investors advantages such as a wide variety of opportunities, industry expertise and lower investment fees. There are numerous types of investment funds, with the most important being mutual funds, hedge funds, money-market funds and exchange-traded funds.

The job of a fund manager is to monitor the fund's investments. The manager has to decide what security he wants, how much capital he wants to invest and for how long. The job requires expertise and and extremely high market knowledge.

Investors choose to invest on a particular fund based on its past returns, risk, investment style, fees, goals and other personal factors they would like to take into account such as portfolio diversification.

The most common types of investment funds are the following:

- Hedge Funds: a hedge fund is characterized for having a lot of freedom in managing its assets. Unlike mutual funds, hedge funds often use aggressive investment strategies, unavailable to other types of funds, and liberally use financial techniques and products such as complex derivatives, short-selling and risk arbitrage.

- Money-Market Funds: a money-market fund is a type of investment fund that makes low risk investments. It is an investment option in which investors are guaranteed a safe place with easily accessible, liquid assets.

- Exchange-traded funds: these funds invest in securities created to behave the same way as a specific index or securities collection. Unlike mutual funds, these funds trade in a similar way as common stock on a stock exchange.

- Mutual Funds: these type of investment funds invest in all types of securities such as stocks, bonds, and other similar assets, giving expertise and market access market to small investors.

2.2   Mutual Funds

Mutual Funds, which indeed will be the ones studied and analyzed in this project, use the capital of different individual investors for investing in an asset portfolio.

Depending on the type and fund's investment style, the securities in which they mainly invest may be equities, which include common stocks, preferred stocks or other equities. This type of funds is called "stock" or Equity Funds. These funds offer investors a simple way to buy a portfolio of stocks, so they can easily diversify their investment and select the choice that best matches their risk, objectives, and exposure requirements.

The investment style is another important factor to take into account as there are two main investment strategies a fund manager can choose for generating returns: active or passive management.

- Active portfolio management funds buy stocks hoping they can outperform a specific stock index. This investment style requires paying special attention to political shifts, market trends, and other factors and irregularities that could affect a specific sector or company, so they can take advantage and program the purchase and sale of investments, maximizing returns. The fees assessed on active portfolios are generally higher than the ones for passive

ones as they require more experience from the manager to deal with more portfolio movements, implying also a higher number of commissions.

- Passive portfolio management funds try to create a specific asset portfolio that behaves very similar to a specific index, trying to generate a similar return to the index chosen with no need of outperforming it.

On the other hand, the other most common mutual funds would be the ones that invest in fixed-income investments, which are called Debt Funds. These funds invest in a wide range of bonds, going from corporate bonds, convertible bonds, government bonds, securitized products or money market instruments. Generally, returns are guaranteed in these type of funds, given the relatively low risk they are exposed to in comparison to equity funds.

Another important characteristic of a fund is its market capitalization or market "cap". It is used to measure a fund's size, which is of the two variables taken into account in this study. The NAV, or Net Asset Value, represents the value of the fund's underlying assets minus the liabilities, divided by the outstanding shares, in a similar way as in a regular public traded company. The fund's NAV is calculated once a day with the closing price of that same day and it basically represents a fund's market value per share. If you multiply the funds NAV by the total number of outstanding shares, you would get the total NAV of the fund which would represent its total size or market cap. Funds are often classified into different types of capitalization. Small Cap, Med Cap, and Big/Large Cap would be the main classifications depending on their total NAV, and this would let investors classify funds' risks and potential advantages in an easier way.

Now that we know how a mutual fund works and how its size is measured, it is very important to understand how to measure their performance and the risk they are exposed to.

Throughout the different studies that already exist in this topic's literature, we can find different performance measurements used to find the correlation between the two variables studied.

Performance and risk measurements previously used:

- Volatility: it statistically measures how dispersed the funds' returns are for a given portfolio. Higher volatility means higher risk, as the uncertainty for a security's value is higher, therefore its value can fluctuate faster and in a higher percentage.

- Alpha (α): it is a risk ratio used as a statistical measurement for calculating investor returns. It measures the difference between a fund's expected returns based on its beta (volatility regarding a specific index) and its actual returns. It can be interpreted as the return value added over a relevant reference risk. In other words, if a fund has a higher return than it is expected to have, given a specific beta, its alpha is positive. If it is lower, alpha would be negative. The higher the alpha, the better the portfolio.

- Sharpe: The Sharpe ratio is a performance measurement used quite often to represent the performance of a given portfolio. It measures risk-adjusted returns and the expression for its calculation is the following:

$$Sharpe = \frac{R_p - R_f}{\sigma_p} \qquad \text{(Equation 1)}$$

Where:

Rp: Return

Rf: Risk free rate

σ_p: Standard deviation

The Sharpe ratio is calculated dividing the average return earned by a portfolio in excess of the risk-free rate, by the standard deviation of the investment portfolio. In other words, it expresses the extra return obtained by the fund per unit of risk. The higher the Sharpe ratio, the more attractive the fund or portfolio is compared to another one, as it apparently would give an investor a higher return at the same level of risk. It is also a way of measuring how attractive a fund is, as the ratio may reflect if the fund's excess returns are due to risk exposure, or to brilliant diversification and investments.

- Treynor: The Treynor ratio is another performance ratio that expresses a very similar idea to the Sharpe ratio, which is that both are reward-to-volatility ratios. The main difference between both would be that the Treynor uses the beta of the fund's portfolio to measure its volatility regarding a specific market, instead of using the standard deviation.

2.3   Regression Models

Another important element that needs to be explained and clarified before describing the model used in the project is the regression analysis.

Regression models are basically equations used to express the relationship between two or more variables. A regression analysis generates an equation to describe the statistical and mathematical relationship between one or more variables and the explained or dependent variable, and eventually, once the model has been tested, it may be used to predict new observations.

There are numerous types of regression models that have been used in the literature when describing the correlation between the size and performance of the funds.

The linear-regression method generally used ordinary least-squares estimation, the equation which is obtained by minimizing the sum of squared residuals.

- Linear model: this method is the simplest regression model and it generally uses regular least squares estimation, the equation which is obtained by minimizing the sum of squared residuals. A linear relationship is a tendency in the data you want to explain that can be modeled by a straight line, with a positive or negative slope. We would use this model between two variables if it has previously been predicted that there is a clear linear relationship between the variables studied.

The simple linear model:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \text{(Equation 2)}$$
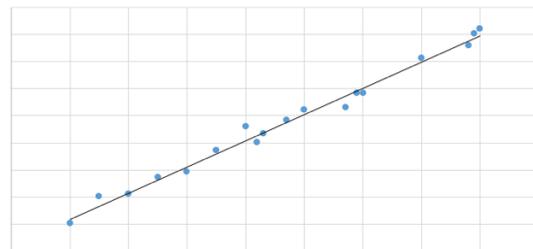
Figure 1 - Linear Relationship

- Quadratic regression: this model reflects the equation of a parabola that best fits a given set of data by minimizing the sum of squared residuals. It is normally used when you predict that there is an optimal point in the database and that the correlation between the two variables studied is not linear, but quadratic. It is the regression model used in the studies that argued that the asset size had an optimal point regarding performance.

The quadratic model:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \quad \text{(Equation 3)}$$
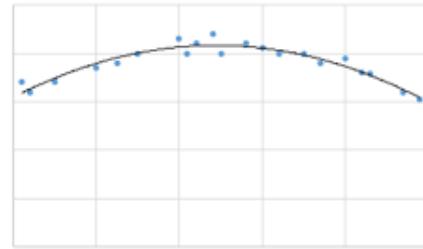
Figure 2 - Quadratic model

- <u>Multiple regression</u>: this regression uses more than one independent variable to explain the dependent variable of the model. Unlike the other two previously explained, with this model we can consider more than one factor, therefore it is more complex and it involves more data than the models with only one explanatory variable.

The multiple regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \dots + \beta_k X_k + \varepsilon \qquad \text{(Equation 4)}$$

Regression models will be tested in this research work to see if the model is good or if it is precise enough in predicting the dependent variable. A coefficient used for analyzing the model's accuracy would be the coefficient of determination (R^2), that ranges from 0 to 1. The closer to 1 this coefficient is, the better the model is in its predictions (the regression line perfectly fits the cloud of data), or in other words, the closer will the model be to the real data as R^2 as the dispersion would be the minimum.

2.4   Literature Analysis

In the previous points of the project, some of the most used performance and risk measurements in the literature have been explained. It is very important to understand very well what these measurements and ratios express, as well as a general approach of the type of funds, how they behave and how their size is measured, to better understand the conclusions that this study pretends to draw. It was also very important to explain the different types of regression models in order to justify which one will be chosen for this study.

As explained before, there are many different studies trying to find the correlation between size and performance.

Shawky and Li (2006) investigate the presence of an optimal asset size for only small-cap mutual funds. Their data source is Morningstar.com and they use a sample of 2921 mutual funds.

Their premise of study is the following (Shawky and Li 2006, page 79): *"If there is an optimal asset size for small-cap mutual funds, risk adjusted returns are expected to rise with increases in assets up to an optimal point, after which returns are expected to decline with further increases in assets"*.

The authors start their study by understanding the previous advantages that small funds might have against the big ones and vice versa.

Small funds (small asset base) advantages:

- They might have a higher performance by putting all their money into their best ideas or investment options.

Big Funds (big asset base) advantages:

- More resources.
- Higher ability to attract and retain talent.
- Better access to companies and clients.
- Higher efficiency and bargaining power with brokers.

These advantages are explained by the authors before starting the regression, to let us understand that both funds' size have positive aspects, therefore the quadratic regression model was going to be the most accurate for the study. Therefore, the initial model for the regression was the quadratic model, explained in point 1.3 of this project, with the total NAV as independent variable, and performance as dependent. The initial regression model was the following one:

$$P = \alpha + b_1 \, NAV + b_2 \, NAV^2 \quad \text{(Equation 5)}$$

In addition to the simple quadratic relationship, the authors considered adding more factors in order to take into account differences in the performance do to variations in market conditions of each year. Therefore, they implemented dummy (binary) variables into the model, each one representing a different year of the period studied, that went from 1997 to 2002. The final model was the following:

$$P = \alpha + b_1 \, NAV + b_2 \, NAV^2 + b_3 \, D_{97} + b_4 \, D_{98} + b_5 \, D_{99} + b_6 \, D_{00} + b_7 \, D_{01} + b_8 D_{02}$$

$$\text{(Equation 6)}$$

Where:

NAV: Net Asset Value representing the size.

D_{ii}: dummy variables representing a different year within the period.

The coefficients and parameters were statistically significant after the regression, supporting the initial idea and existence of an optimal small-cap asset size. After having statistically quantified the evidence, according to Hany A. Shawky and Liuling Li, a possible explanation for the presence of an optimal fund size could be the following. First, as previously conjectured before having the results of the regression, as a fund grows in size, it starts investing in other securities rather than

in the best options, as it has more money than the one it could optimize its investments with. Therefore, there might be a limited supply of optimal small cap stocks in the equity market, leaving worse or less attractive choices for the biggest funds. Second, diseconomies of scale regarding the production of the necessary information on the buy side of the market. Finally, regarding costs, monitoring and trading costs increase as small-cap funds increase in size.

These ideas are very useful as they at least support the concept of diseconomies of scale that have been numerously mentioned in other studies apart from this one, showing that investment funds' size is indeed correlated with their performance, and that in most cases, the correlation is negative.

It is very interesting that according to Amenc et al (2004), Liang (1999), and Ding et al (2009), the size-performance relationship is linear-positive. All of them worked with hedge funds instead of mutual funds giving to the literature a possible idea that maybe the positive gain in performance related to size is related to completely different investment styles, securities and strategies, such as the ones that hedge funds use.

To get more in depth into why fund size matters, some points will be clarified in order to support the idea of a correlation between both variables:

Initially, when a mutual fund is relatively small in size and starts growing, size growth provides the fund with higher net returns, therefore the fund starts to gain economies of scale (Indro, Jiang, Hu, & Lee, 1999). This is due to the fact that as transaction volume starts increasing, brokerage commissions in trade executions begin to decrease for big funds. Growing funds gain other cost advantages related to the access to data, research services, support, and other expenses, that are spread throughout a bigger platform, therefore the marginal cost is smaller than in smaller funds, giving growing funds a positive competitive advantage.

As much as big funds seem to have lower marginal costs related to the brokerage commissions explained before, these commissions are just a small part of the total

transactions costs that a fund has to deal with (Wagner and Edwards, 1999). This fact, in addition to the disadvantages that bigger funds obtain when growing fast without control, will give growing funds the following disadvantages:

- Big and well known funds attract attention as they are seen as good market representatives, therefore their stock selection might seem a good initial option. Not only the stock selection but the timing of the investment, which is sometimes as important as the selection, is analyzed by smaller funds giving bigger funds the disadvantage of being noticed, or at least of not being able to work without letting their competitors know what their next step is going to be.

- Growing in size means growing in capital and as more human capital is required, managing a higher number of investments becomes more difficult. Coordination between different fund managers and personnel as the fund grows in size becomes more difficult and intense, therefore there is a loss in control in comparison to smaller funds with smaller information networks.

- Regarding the transaction costs and according to Loeb (1983), unlike with smaller investments, the sale of large blocks of stock come with a higher bid-ask spread, becoming for difficult for the completion of the purchase. While small block trades can be almost instantly and anonymously done, large investments are usually done through intermediaries, incurring a higher cost for the fund.

- Finally, as explained before with a very basic concept, smaller funds might have a higher performance by putting all their money into their best ideas or investment options. There are limited optimal investment opportunities in the market, therefore when the fund has more capital than the one required for that specific amount of stock, fund managers choose assets that they

would normally not invest in according to their initial investment strategy, giving the impression that big funds invest in assets they would not invest in if they were smaller.

Again these points suggest that diseconomies of scale are present when mutual funds grow in size, but at the same time big funds have other advantages, leaving again uncertainty regarding the relationship between size and performance. That is why Indro, Jiang, Hu, and Lee (1999) thought that another characteristic should be taken into account when studying the relationship between both variables. That is why the authors took into account the fund's investment orientation in addition to the funds investment management style.

There are two main fund categories regarding the fund manager's investment orientation, which basically takes into account the approach they take when investing, the way they choose the stocks, and the type of markets that best matchl their requirements:

- Growth funds: the objective of these type of funds is to invest in companies fund managers believe will grow faster than the average market, regarding revenues, cashflows or earnings (high Price to Earnings ratios). As the risk is higher compared to value funds, when investing in these type of companies, growth oriented funds usually need a higher tolerance for risk.

- Value Funds: the goal of these funds is to invest in undervalued companies, or in other words, companies which stock prices do not represent their real value (low price-to-tangible-book-value ratios), hoping this will be recognized by the market eventually.

- Blend Funds: these funds do not have a particular investment orientation. They invest in both value and growth stocks.

Indro, Jiang, Hu, and Lee (1999) analyzed mutual funds under the investment orientation and management style scope and draw interesting conclusions supported by statistical evidence. The authors found that actively managed mutual funds must reach a minimum fund size before they attain enough returns to cover their acquiring and trading on information costs. They also support the idea that there are diminishing marginal returns to information activities and that there are negative marginal returns when a funds grows above its optimal size. Summing the information up, Indro, Jiang, Hu, and Lee (1999) found that from a sample of almost 700 equity mutual funds, studied from 1993 to 1995, fund size does affect the performance of the fund. They must reach an optimal size to justify their acquiring and trading on information costs.

By taking all these studies into account and as it was said at the beggining of the literature review, studies regarding a relationship between asset size and fund performance have yet not been able to give a conclusive result. A big number of studies agree on a negative relationship between both variables, but there are studies supporting a positive relationship, quadratic, or even none, having each one of them statistical and mathematical evidence supporting each one of their conclusions.

As there is no unique answer to the fund size-performance relationship, this has been considered as a gap in the literature for another study that might resolve this question in a more accurate way, as all the previous studies' information and conclusions will be taken into account.

3. Description of the model

This project will try to explain the possible relationship between the size of a mutual fund and its risk-adjusted performance, with the objective of finding out if growing in size has any effect on risk measures.

3.1   Objectives and specifications

The model will find the relationship between the dependent variable, which would be the performance measure that is going to be explained by the model, and other independent variables or predictors, where the size of the fund will be included.

Linear Model

To fully understand the final model that is going to be used to find the relationship between the two variables studied, the basic model from which the most complex ones are elaborated is the linear model.

As explained before, this model looks for a trend in the data that can be modeled as a straight line. It is very limited, as only basic relationships can be explained by a straight line, but to start off it is the easiest model to understand and eventually build up a more complex model from it.

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \qquad \text{(Equation 7)}$$

Where:

Y: the dependent variable in this study is the the Sortino ratio, based on the presumption that its value is influenced by the size of the fund.

X_1: the independent or explanatory variable is the total Net Asset Value, that represent the size of the fund.

β₀: the intercept would be the term that would explain the value of the dependent variable when the explanatory is 0.

β₁: regression coefficient that represent the slope of the linear model or relationship between the both variables. If this term is positive, the performance of the fund would increase as the fund's size increases.

ε: the error or disturbance term includes all the factors that the explanatory variable is not capable of explaining.

This model is very simple but as terms are added in the following steps, it will gain a good enough complexity for explaining the relationship between size and performance.

Panel Data

It is one of the main characteristics of the project and it differentiates it from the rest of studies, as it adds more complexity into the regression model.

Panel Data receives its name from the format in with the database is displayed. It is basically a dataset in which the behavior of entities is analyzed through time.

The format in which the variables and data must be displayed in order to apply panel data is the following:

Entity	Date	Y	X
1	2010	1,2	5,1
1	2011	1,3	5,3
1	2012	1,1	5,4
2	2010	0,8	8
2	2011	0,75	8,5
2	2012	0,9	8,3

Table 1: Exampe of Panel Data Structure

The entities studied must be located in the first column, followed by the time period of each entity, which are followed by the data of each entity and date. Eventually, the behavior of entities, which in the case of this project are the funds, can be analyzed through time.

The main advantages of panel data are the following:

- With the data displayed in the format described in Table 1, panel data allows the analyst to control those factors that cannot be measured or controlled like cultural factors or difference in business strategies across companies (funds).
- It could also control variables that vary throughout time but not across entities. It takes into account individual heterogeneity.

There are two main ways of analyzing panel data:

1. Fixed Effects Model:

Fixed-Effects is used to analyze the impact of variables that change through time.

- Tries to explain the relationship within dependent and independent variables within an entity (fund).
- Each fund has its own features that could influence its behavior or business practices
- Time-invariant characteristics are unique to the fund and not correlated with other fund features.

The expression of the Entity Effect Model is the following:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mathbf{F_i} + \varepsilon_{it} \quad \text{(Equation 8)}$$

Where:

Y_{it}: Dependent variable (performance) where i=entity/fund and t=time.

X_{it}: Explanatory variable (size).

β_0: Intercept.

β_1: Regression coefficient among variables.

ε: Error term.

F_i: Entity fixed effect variable (i=first fund identifier… last fund identifier).

This model takes into account the differences among individuals, as explained before. The variable F_i (i =fund identifier) allows to explain the differences among the funds regarding business characteristics.

In addition to the the entity effects, you can also add time effects variables (T_t) into the model in order to explain differences within time periods and to increase complexity, becoming a time and entity fixed effects regression model.

The expression of the *time and entity fixed effect model* is the following:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + F_i + \mathbf{T_t} + \varepsilon_{it} \quad \text{(Equation 9)}$$

Where:

Y_{it}: Dependent variable (performance) where i=entity/fund and t=time.

X_{it}: Explanatory variable (size).

β_0: Intercept.

β_1: Regression coefficient among variables.

ε: Error term.

F_i: Entity fixed effect variable (i=first fund identifier… last fund identifier).

T_t: Time period fixed effect variable (t= different periods within the sample).

2. <u>Random Effects:</u>

The other type of panel data model application would be the random effects model. While in fixed effects we assume that each entity has its own individual characteristics that influence the predictor variable, in random effects model the variation across entities is assumed to be random.

If you think differences between entities affect the dependent variable, this model should be used.

The expression of the *Random Effects Model* is the following:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + U_{it} + \varepsilon_{it} \quad \text{(Equation 10)}$$

Where:

Y_{it}: Dependent variable (performance) where i=entity/fund and t=time.

X_{it}: Explanatory variable (size).

β_0: Intercept.

β_1: Regression coefficient among variables.

U_{it}: Between-entity error.

ε_{it}: Within-entity error.

This model assumes that the entity's error term is not correlated with the explanatory variables of the dependent variable, allowing time-invariant variables to be taken into account as predictors.

The panel data models will try to find the relationship between the variables studied, and to explain the correlation among them as best as possible, if there is any.

3.2   Database

Once it has been determined what regression model will be the one chosen to explain the relationship between fund's size and performance, the following step is determining the data sample that is going to be studied.

Through the Wharton Research Data services (WRDS), from Wharton, University of Pennsylvania, access was granted in order to find a trustful source to support the veracity in the conclusions of the project. WRDS is an award-winning platform and business tool used in more than 30 countries, that include data related to Accounting, Banking, Economics, ESG, Finance, and related topics, which are constantly reviewed and validated by their analytics professionals. Through this tool the CRSP mutual fund database could be accessed, being the financial source chosen for the correlation study.

3.2.1   CRSP Mutual Fund Database

The database used for study is the CRSP Mutual Fund Database and it provides all the information needed for the project.

The CRSP Mutual Fund Database was designed for research and educational purposes. The Center for Research in Security Prices (CRSP) has supplied scholar researchers with high quality data since 1960. It covers finance, economics and similar topics, widening the literature with every research. CRSP databases are used throughout the whole world in almost 500 institutions, in more than 35 countries, to support both academic lessons and research.

The database used for this project included a history of each mutual fund's name, fee structure, holdings, asset allocation that let us classify the fund, asset class codes, and even management company contact information. All of these data is provided from 1962 to 2015, depending on the fund and on the type of data which can be daily, monthly, or annual.

Initially, the original CRSP Mutual Fund Database was generated by Mark M. Carhart and included mutual fund data from 1961 to 1995 to widen the finance literature regarding mutual funds, as he considered there was a lack of data available. Progressively, more and more data has been added to the original database, becoming one of the best sources of funds and finance information. CRSP workers are constantly updating and correcting the database when errors are identified, therefore data quality is exceptional and constantly being improved.

3.2.2   Database used for the study

The CRSP Mutual Fund database provides a tool that lets you filter the entire dataset, depending on your requirements and your research needs and criteria. Once the equity funds for the study were manually selected following a criterion later explained, the initial filter is applied, choosing the variables and information required of each fund as well as the time period wanted to study.

<u>Fund Selection</u>

The funds studied in the project are Equity Funds. As explained before they are basically those mutual funds that invest mainly in stocks. The objective of this stage is selecting the optimal funds for the final sample in which the regression analysis will be done.

During this initial fund identification stage, which is very important as it will determine the characteristics of the funds studied, the CRSP database had a filter that provided variables that showed the proportion of the fund invested in each type of stock, bond or security, allowing the user to know the asset composition of each fund and therefore letting him know which funds invested principally in stocks.

This filter tells you what percentage of the fund is invested in common stock, preferred stock, other equities, convertible bonds, corporate bonds, municipal bonds, government bonds, other bonds, asset-backed securities, cash and others.

The variables that indicated the fund's asset composition that corresponded to an equity fund, were the following:

- Amount of fund invested in Common Stocks.
- Amount of fund invested in Preferred Stocks.
- Amount of fund invested in Other Equities.

For a proper equity fund identification, the funds were manually selected following the following criteria:

- Funds are selected if the sum between the amount of fund invested in common stocks, amount of fund invested in preferred stocks and amount of fund invested in other equities is in average higher than 90%.
- Those funds with a clear focus on bond investments are discarded.
- With these criteria, we guarantee that the proportion of the fund invested in of stock is higher than 90%, giving the fund an equity asset composition.

A 100% would mean a pure equity fund which would be ideal, but it is very difficult to find as almost no fund invests 100% of its assets.

Here there are a couple of examples of funds that have been considered equity funds.

Fund Identifier	Date	fund invested in Common Stocks	fund invested in Preferred Stocks	fund invested in Convertible	fund invested in Corporate	fund invested in Municipal	fund invested in Government	fund invested in Other	of fund invested in Cash	fund invested in all Bonds	invested in Asset-Backed Securities	fund invested in	fund invested in other	Fund Name
2830	09/30/2002													AFBA 5Star:Small Ca
2830	12/31/2002													AFBA 5Star Fund, In
2830	03/31/2003	98,54	0	0	0	0	0	0	1,46	0	0	0	0	AFBA 5Star Fund, In
2830	06/30/2003	97,57	0	0	0	0	0	0	2,43	0	0	0	0	AFBA 5Star Fund, In
2830	09/30/2003	91,08	0	0	0	0	0	0	8,92	0	0	0	0	AFBA 5Star Fund, In
2830	12/31/2003	81,33	0	0	0	0	0	0	18,67	0	0	0	0	AFBA 5Star Fund, In
2830	03/31/2004	84,68	0	0	0	0	0	0	15,32	0	0	0	0	AFBA 5Star Fund, In
2830	06/30/2004	91,02	0	0	0	0	0	0	8,98	0	0	0	0	AFBA 5Star Fund, In
2830	09/30/2004	100	0	0	0	0	0	0	0	0	0	0	0	AFBA 5Star Fund, In
2830	12/31/2004	92,41	0	0	0	0	0	0	7,59	0	0	0	0	AFBA 5Star Fund, In
2830	03/31/2005	91,91	0	0	0	0	0	0	8,09	0	0	0	0	AFBA 5Star Fund, In
2830	06/30/2005	90,57	0	0	0	0	0	0	9,43	0	0	0	0	AFBA 5Star Fund, In
2830	09/30/2005	94,37	0	0	0	0	0	0	5,63	0	0	0	0	AFBA 5Star Fund, In
2830	12/30/2005	94,37	0	0	0	0	0	0	5,63	0	0	0	0	AFBA 5Star Fund, In
2830	03/31/2006	94,8	0	0	0	0	0	0	5,2	0	0	0	0	AFBA 5Star Fund, In
2830	06/30/2006	94,77	0	0	0	0	0	0	5,23	0	0	0	0	AFBA 5Star Fund, In
2830	09/29/2006	93,02	0	0	0	0	0	1,79	5,18	0	0	0	0	AFBA 5Star Fund, In
2830	12/29/2006	90,66	0	0	0	0	0	0	9,34	0	0	0	0	AFBA 5Star Fund, In
2830	03/30/2007	93,04	0	0	0	0	0	0	6,96	0	0	0	0	AFBA 5Star Fund, In
2830	06/29/2007	96,02	0	0	0	0	0	0	3,98	0	0	0	0	AFBA 5Star Fund, In
2830	09/28/2007	48,88	0	0	0	0	0	0,79	50,34	0	0	0	0	AFBA 5Star Fund, In
2830	12/31/2007	50,31	0	0	0	0	0	0,75	50,22	0	0	0	0	AFBA 5Star Fund, In
2830	03/31/2008	94,43	0	0	0	0	0	1,9	3,67	0	0	0	0	AFBA 5Star Fund, In
2830	06/30/2008	96,6	0	0	0	0	0	3,28	0,12	0	0	0	0	AFBA 5Star Fund, In
2830	09/30/2008	96,06	0	0	0	0	0	2,25	1,69	0	0	0	0	AFBA 5Star Fund, In

Figure 3 – Example1 of Equity fund selected.

As seen in Figure 3, throughout most of the time the fund selected invested most of its capital in common stocks (average 91% of their capital is invested in common stocks). In addition to this, the fund does not make any investments at all in the bond market.

Figure 4 – Example 2 of Equity fund selected.

Figure 4 shows another example of an equity fund selected for the sample. The minimum percentage of total capital invested in equities is 97%, therefore its clearly an equity fund. This example shows how even though in one year the fund invested 2% of its capital in bonds, the other 98% of capital was invested in common stocks, following the fund's asset composition strategy throughout the whole period of time studied.

Figure 5 – Example of fund NOT selected.

Figure 5 is an example of a fund that clearly invests in the bond market, as all of its capital is invested in Municipal bonds. This is an example of a fund profile that will not be selected for the final sample.

After a detailed analysis, 300 equity funds are identified and included in the final sample.

Time period

This project wanted to include as much data as possible in order to have a better and more complete result. As the study uses daily observations and the database can provide daily information from 1998 to 2015, the study includes daily observations of 17 years.

Another point regarding the time period and panel data characteristics is that datasets are usually brought in two types of distribution:

- Balanced panel: each entity is observed for the same period of time.
- Unbalanced panel: entities are observed for different periods.

Although the sample used for the study is observed from 1998 to 2015, there are some funds that only have data from smaller periods within that bigger period of time (they either disappeared halfway or started within the period), therefore as there is not the same amount of data for every fund, the panel data studied is unbalanced.

Information for each fund

Once the time period is selected, the next step is selecting the information needed from each fund for the study:

- Fund identifier (CRSP_Fundno): This variable is used to identify the fund. It lets you know what fund information you are looking at, and it is an essential variable in the fund selection process for the fund identification.

- Date (date): as the project considers daily observations, the date variable shows daily information (day-month-year), and it is essential regarding the panel data approach.

- Total NAV (NAV): The total net asset value is equal to the total assets minus the total liabilities as of month-end and it is reported in millions of dollars. It is added later to the sample, once the data is organized in months for the panel data approach.

- Daily Net Asset Value (NAV per share): it represents the daily value of the mutual fund's underlying assets, including cash, minus the liabilities, divided by the total number of shares outstanding. It is important for the results of the project to use daily observations, in order to take into account as much data as possible, as monthly observations would not consider the daily volatility. The daily data available started in September 2, 1998, therefore this date is the begging of the studied window.

It is important to clarify that although the data is later organized and presented monthly for the panel data approach, the study considered daily observations to be as accurate as possible.

Number of funds

Taking into account that the study analyzes daily observations from a time period of 17 years, 300 funds is considered as the optimal sample size. 300 funds would give the study almost 1 million of total observations, which is considered to be an

excellent size for an accurate statistical study as its also the typical size of this type of publications.

Database Summary:

- **Size:** 300 Mutual Funds.
- **Period:** 1998-2015.
- **Type of funds:** Equity Funds.

3.3   Calculation of performance and risk measurements

Once the sample is identified, the next step is calculating the performance and risk measurements, which are the dependent variables that the regression model would try to explain.

These performance and risk measurements are performed for every month and are calculated with daily data, taking into account as much information as possible within the study.

The first thing that needs to be done in order to obtain the ratios, is calculating the returns of the fund, starting from the daily NAV per share.

The daily returns were calculated following the next expression:

$$Return \; _t = \text{Ln} \; (\frac{NAV_t}{NAV_{t-1}}) \qquad \text{(Equation 11)}$$

Where:

> **Return**: rate of return over a single day, assuming continuously compounded return.
>
> **NAV**: daily (t) net asset value *per share*.
>
> **t:** day to which the rate of return refers to (t-1would correspond to the day before).

To compare returns over time periods of different lengths on an equal basis, it is very useful to convert each return into an annual equivalent rate of return, or annualised return.

Once the daily returns for each fund are obtained, the measurements are calculated:

Volatility

As both the Sharpe and the Sortino ratio are calculated taking into account the average return in excess of the risk free rate (R-Rf), it makes sense to study the volatility of that precise daily value.

Volatility, as explained in the literature review, is a statistical measurement of the return dispersion. It is measured with the standard deviation which analyzes the spread of the values about the mean of the sample studied.

One of the main characteristics and advantages of this project is the huge amount of data taken into account and that is justified in this precise step. As the performance measurements are monthly ratios, the data had to be observed daily in order to consider all the variations the NAV per share does along one month. That way all the dispersion is taken into account within the study.

Here is an example of how the volatility is calculated regarding one month (February) of a specific fund:
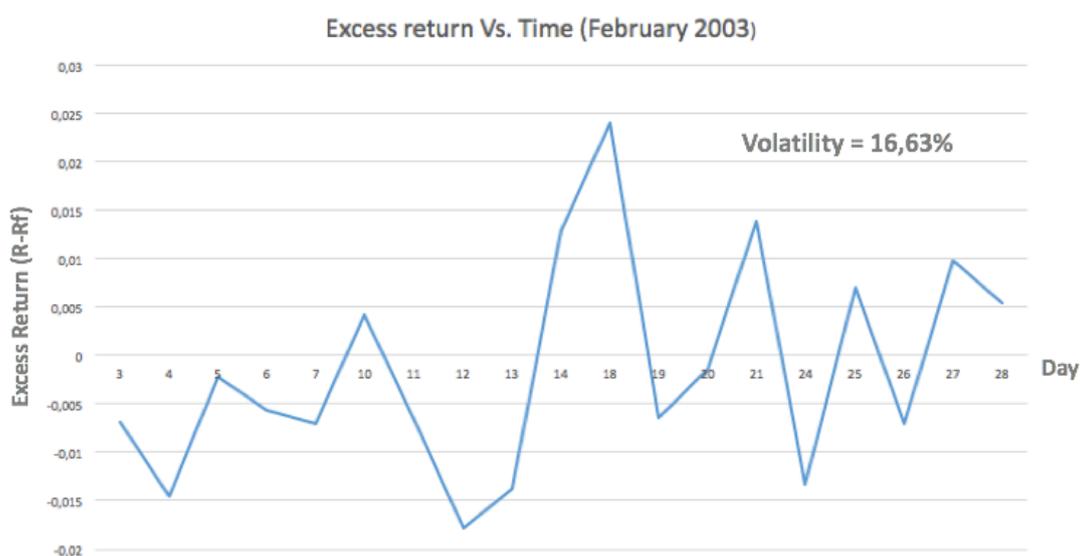
Figure 6 - Example of volatility calculation of one month of one fund

Figure 6 shows how the process of calculating the excess return volatility during one month of a single fund. The standard deviation of the excess return (R-Rf) was calculated for all the time horizon (1998-2015) and for the 300 mutual funds included in the study. It is an important step in order to help in the process of calculating the Sortino, which will be the main performance risk of the study, in addition to drawing additional conclusions about the fund's behavior.

Sharpe ratio

Just as a reminder, the Sharpe ratio is a performance measurement used quite often to represent the performance of a given portfolio and takes into account the average return in excess of the risk free rate and the standard deviation of the returns.

It was also calculated to add information to the project in case the ratio showed something in particular, but this performance measurement is not the core element of the project.

Sortino ratio

Introducing the Sortino ratio in the analysis is one of the main contributions of this project. It is a modification of the Sharpe that takes into account the the downside volatility instead of all of it.

The Sharpe ratio assumes that returns follow a normal or symmetrical distribution, and uses the mean return as a target return. The Sortino's main advantage is that it uses a relative target rate of return instead of simply using the average return and that it quantifies downside volatility without penalizing upside volatility.

The performance measurement is calculated with the next expression:

$$Sortino = \frac{R - R_f}{\sigma_d} \qquad \text{(Equation 12)}$$

Where:

R: Return

Rf: Risk free rate

σ_d: downside deviation or standard deviation of negative asset returns.

The standard deviation of the negative asset returns takes into account the volatility of the "bad" returns. Therefore, a high Sortino of a specific period of time would be a positive aspect as it would mean a low probability of a high loss as the bad risk to which the assets are exposed is relatively low.

To calculate the Sortino ratio, you must subtract the risk-free rate to the return in the same way as in the Sharpe ratio, calculating the return in excess of the risk-free rate. The difference between the Sortino and the Sharpe comes when dividing the excess return by the standard deviation of the negative asset returns.

The way in which the downside risk is calculated is the following: for a single month, the positive excess returns are accounted for as zeros. The process would be then basically changing the positive excess returns into zeros and leaving the negative ones as negatives and calculating the standard deviation of all the excess returns.

There is another form of Sortino in which instead of transforming the positive excess returns into zeros, it simply takes into account the negative excess returns when calculating the downside risk, leaving totally aside the positive excess returns in the calculations. This project uses the first form of Sortino, as it uses every excess return in the Sortino calculation, regardless of wether is positive or negative, therefore this method takes into account all the data available, unlike the second method explained.

Here is an example of how the downside risk is calculated regarding one month of a specific fund:

crsp_fundno	Date	Return (daily)	Rf (daily)	Excess Return (R-Rf)	Negative Excess Return (daily)	SORTINO
8750	01/12/1998	-0,02101	0,00018861	-0,02119865	-0,0211987	**1,089**
8750	02/12/1998	0,00166389	0,00018861	0,001475282	0	
8750	03/12/1998	-0,0171876	0,00018861	-0,01737621	-0,0173762	Downside risk (annualized)
8750	04/12/1998	0,01801851	0,00018861	0,017829894	0	0,150954434
8750	07/12/1998	0,00497101	0,00018861	0,004782401	0	Avg Excess Return (annualized)
8750	08/12/1998	-0,0074658	0,00018861	-0,00765443	-0,0076544	0,164441546
8750	09/12/1998	-0,002501	0,00018861	-0,00268966	-0,0026897	
8750	10/12/1998	-0,0176847	0,00018861	-0,01787328	-0,0178733	
8750	11/12/1998	-0,0046838	0,00018861	-0,00487246	-0,0048725	
8750	14/12/1998	-0,0167854	0,00018861	-0,01697403	-0,016974	
8750	15/12/1998	0,01678542	0,00018861	0,016596805	0	
8750	16/12/1998	0,00255755	0,00018861	0,002368934	0	
8750	17/12/1998	0,02189561	0,00018861	0,021707	0	
8750	18/12/1998	0,00498547	0,00018861	0,004796858	0	
8750	21/12/1998	0,00866521	0,00018861	0,008476596	0	
8750	22/12/1998	-0,0004109	0,00018861	-0,00059954	-0,0005995	
8750	23/12/1998	0,01832672	0,00018861	0,018138108	0	
8750	24/12/1998	0,00362538	0,00018861	0,00343677	0	
8750	28/12/1998	-0,0044328	0,00018861	-0,00462142	-0,0046214	
8750	29/12/1998	0,0092463	0,00018861	0,009057685	0	
8750	30/12/1998	0,00239808	0,00018861	0,002209471	0	
8750	31/12/1998	-0,0015981	0,00018861	-0,00178669	-0,0017867	
8750	04/01/1999	0,0019972	0,00017441	0,001822794	0	

Tabla 2: Example of Sortino and downside risk calculation

Table 1 shows the calculation process of the Sortino ratio for one month of one specific fund. As it can be seen, the negative excess return column shows how positive excess returns are transformed into zeros to calculate after the downside risk for the Sortino. This process was replicated throughout the whole database, in order to calculate al the Sortino ratios of every month in the time-period studied, and for every fund.

To finish the calculations step, although the study will use the Sortino ratio as the performance-dependent variable, the other performance and risk measurements were calculated as well, due to the fact that they give a lot of extra information about the behavior of the funds, and they could also be useful for the final conclusions.

3.4   Data Compilation

Once the performance and risk measurements have been calculated, the next step is compiling the data and organizing it in order to create a database that fits the regression software requirements. That would be, the panel data structure.

In order to make the regression analysis, the data is organized in the following way:

- CRSP_Fundno - 1st Column: in this column, the variable *CRSP_Fundno* is located. This variable, which represents the fund identifier mentioned before, is positioned first in order to have all the data pre-organized by funds, as we previously know that the weight of the fund will be important in the regression, regarding the panel data.

- Date - 2nd Column: as each fund had daily returns in order to take into account more data to try to make the study more accurate, once the performance and risk measurements were calculated, the *Date* variable was needed to reorganize the performance measurements as they represented monthly summaries. That is why the format of the data variable was converted into *month-year*, from original data that reflected daily information.

- Total NAV - 3rd Column: after the date, the independent variable of the regression that represents the size of the fund is positioned.

- Performance and risk measurements - Final Columns: the dependent variables that reflect the performance of the funds are placed in the final columns. Each variable summarizes monthly information, therefore there

are twelve measurements per year, per fund and per performance measurement.

An example of how the data was compiled and organized can be seen in the following Table:

CRSP_Fundno	Date	Total NAV	Sharpe	Sortino	Volatility
7873	dic-02	3,4	-2,90	-3,19	0,15
7873	ene-03	3,5	-0,10	-0,12	0,18
7873	feb-03	3,4	-1,50	-2,02	0,17
7873	mar-03	3,4	0,21	0,33	0,23
7873	abr-03	3,6	4,55	6,43	0,14
…	…	…	…	…	…

Tabla 3: Sample organized ready for regression analysis

Almost one million lines of data had to be converted and adapted to the format shown in the previous table in order to create a well-adapted database that fits the panel data requirements.

To finish up and summarize the data compilation step, all the daily data that was used for the calculation, which included variables not shown in the final sample such as net asset value per share, daily returns, risk-free rates, downside risks and other binary control variables used for the performance ratio calculations, were compiled into the variables useful for the study and converted from daily observations into monthly measurements.

4. Analysis

4.1   Statistical Analysis

<u>Hausman Test</u>

The first step is choosing what panel data model should be applied to the data sample and this is done through the Hausman test. This test determines if we should use a fixed-effects model or a random effects model by using a null hypothesis that supports that the preferred model is random effects against the fixed effects. This way the model selected will be the optimal one.

The results for the Hausman test of the data sample are the following:

```
. hausman fixed random

                     ─── Coefficients ───
                     (b)          (B)            (b-B)        sqrt(diag(V_b-V_B))
                    fixed        random        Difference          S.E.

     total_nav     -.00002    -.0000383         .0000182          .0001518

                        b = consistent under Ho and Ha; obtained from xtreg
          B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

            chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                    =        0.01
            Prob>chi2 =      0.9043
```

Figure 7 - Hausman test results

According to Hausman, if Prob>chi2 is lower than 0,05 you should use a fixed effects model, and if not, a random effects model. The results show a Prob>chi2 = 0,9043, therefore the panel data regression will be done with the Random Effects Model.

Once the data is tested and the random effects model is chosen for the regression, the next step would be preparing the data and finally running the model to see if the variables studied are correlated.

Data preparation

```
. xtset crsp_fundno date2, monthly
       panel variable:  crsp_fundno (unbalanced)
        time variable:  date2, October 1998 to September 2015, but with gaps
                delta:  1 month
```

Figure 8 - Previous step: Variable preparation

In this step, the fund identification is selected as the panel variable to let the model know how to weight the difference between funds, and the variable "Date", that includes the monthly dates that summarized the performance of the fund, is selected as the time variable. We also indicate that the time variable is monthly.

Regression

The final step is performing the regression analysis, showing the results in the following picture:

```
. xtreg sortino total_nav, re

Random-effects GLS regression              Number of obs      =       45256
Group variable: crsp_fundno                Number of groups   =         298

R-sq:  within  = 0.0000                     Obs per group: min =          47
       between = 0.0001                                    avg =       151.9
       overall = 0.0000                                    max =         204

                                           Wald chi2(1)       =        0.01
corr(u_i, X)   = 0 (assumed)               Prob > chi2        =      0.9080
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	-.0000383	.0003309	-0.12	0.908	-.0006868	.0006103
_cons	5.116964	2.275128	2.25	0.025	.6577948	9.576134
sigma_u	36.571882					
sigma_e	164.68044					
rho	.04700058	(fraction of variance due to u_i)				

Figure 9 - Regression Results

The main indicator that tell if the correlation exists whatsoever is the R squared, that with a value of 0.0001, indicates that the model doesn't explain the data.

At first sight, the first conclusions that are drawn from these results are that there is no correlation between the size and the performance of the funds. The analysis has not shown enough evidence to support the relationship between both variables as the indicators tell that the model doesn't explain the data. This would support the opinion of those authors that say that the performance of a fund is independent from its size.

There are many reasons that could explain why the results show no correlation:

- The period of time taken into account might be too wide. It could make no sense to compare fund behaviors of 1998 funds with 2015 funds. Their nature or investment style could be completely different due to a lot of factors such as technology, markets' behavior, etc.

- It might have been very ambitious to try to create a model with such a big amount of data, as it wanted to explain daily observations.

- Simply, there might not be a relationship between size and performance, therefore the project would support that result, which was a possible option since the start of the study.

4.2   Alternative Analysis

As the results from the panel data regression show no linear correlation between size and performance, several other exploratory approaches were performed in order to assess that the relationship is indeed nonexistent.

Before proceeding to these alternatives, it is important to graphically represent the Sortino ratio against the size of the funds to see if there is any evident graphical relationship, to give us possible ideas for the alternative analysis.

In the following picture, the Sortino is represented in the "y" axis, with the total NAV in the "x" axis:
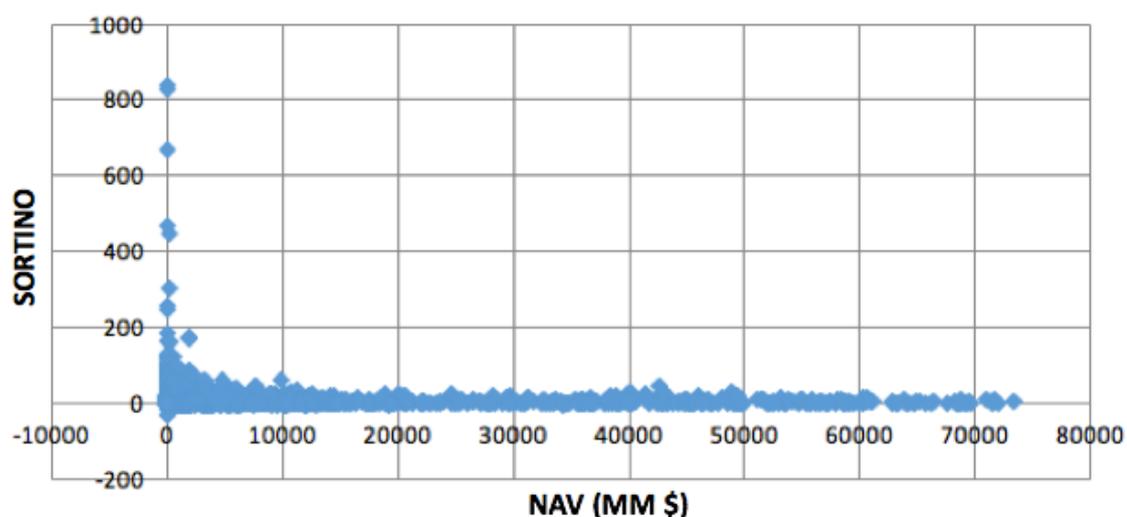
Figure 10 - Sortino (Performance) VS. NAV (size) I

The graphic is rescaled, getting rid of some outliers, to see better the tendency of the points. The rescaled graphic is shown next:
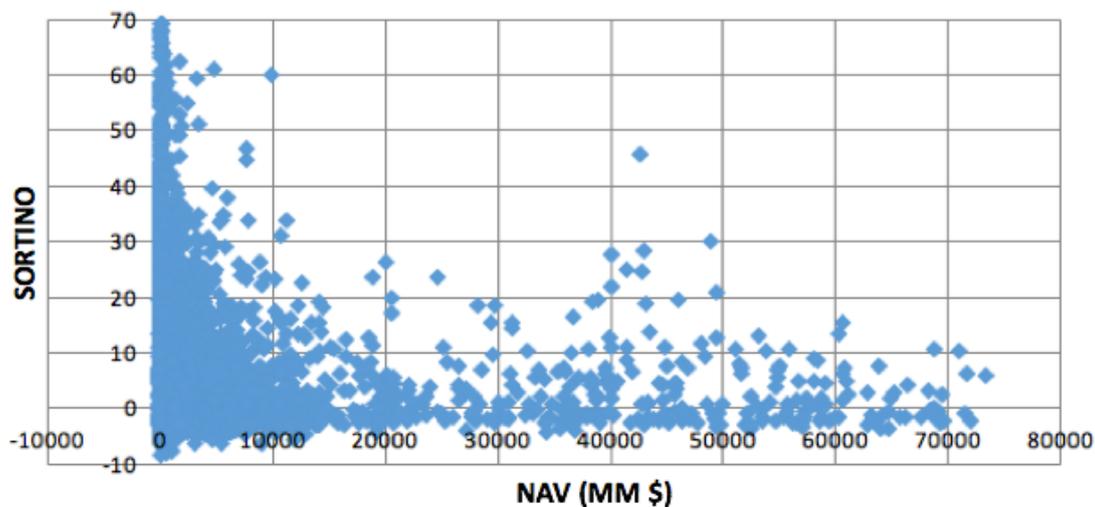
Figure 11 - Sortino VS. NAV II

The different analyses in addition to the one already done are the following:

1. Size division

2. Individual Analysis

3. Period Analysis

4. Dispersion-Range Analysis

4.2.1   Size division

The objective of this alternative analysis is to study two different fund sizes independently, to see if any correlation between performance and size is found on each size, as separate samples.

To choose the NAV value for the division, 5 different size values will be proposed, selecting the final division value after a small analysis. The options are 6000, 7000, 8000, 10000 and 15000 MM $, which have been visually proposed after looking at the graphic. With each option, the cloud for the big size will be analyzed, seeing what linear regression better models the correlation, justifying that value for the division.

The size division is the following one:
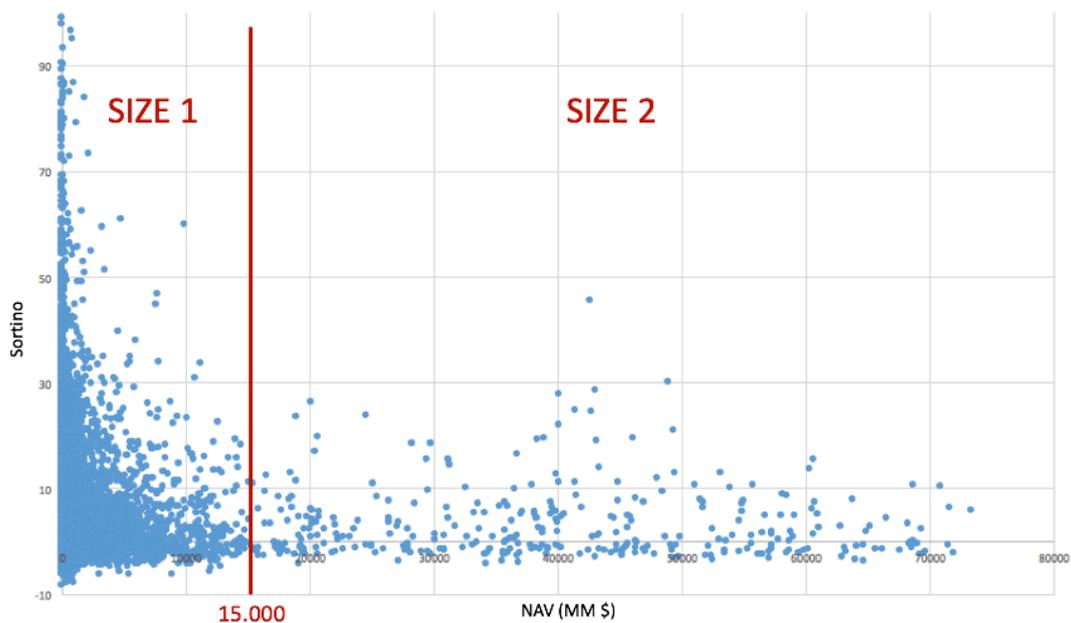
Figure 12 - Size division

The 5 different sizes proposed are analyzed to see which one gives a better result for the division.

The procedure is the following: if for example 6000 MM $ is the one to be tested, only the dots that correspond to sizes higher than 6000 MM$ (NAV), are taken into account in the regression analysis. At the end, the regression model with the best result will determine the value used for the size cut.

- NAV > 6000


```
. xtreg sortino total_nav, re

Random-effects GLS regression              Number of obs     =       822
Group variable: crsp_fundno                Number of groups  =        13

R-sq:  within  = 0.0004                     Obs per group: min =         1
       between = 0.0362                                     avg =      63.2
       overall = 0.0000                                     max =       203

                                            Wald chi2(1)      =      0.02
corr(u_i, X)   = 0 (assumed)                Prob > chi2       =    0.9011
```

| sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
total_nav	−1.76e−06	.0000141	−0.12	0.901	−.0000295 .000026
_cons	3.168794	.3952062	8.02	0.000	2.394204 3.943384

Figure 13 – NAV: 6000 regression analysis

- NAV > 7000


```
. xtreg sortino total_nav, re

Random-effects GLS regression              Number of obs     =       711
Group variable: crsp_fundno                Number of groups  =        12

R-sq:  within  = 0.0011                     Obs per group: min =         1
       between = 0.0219                                     avg =      59.2
       overall = 0.0002                                     max =       197

                                            Wald chi2(1)      =      0.14
corr(u_i, X)   = 0 (assumed)                Prob > chi2       =    0.7037
```

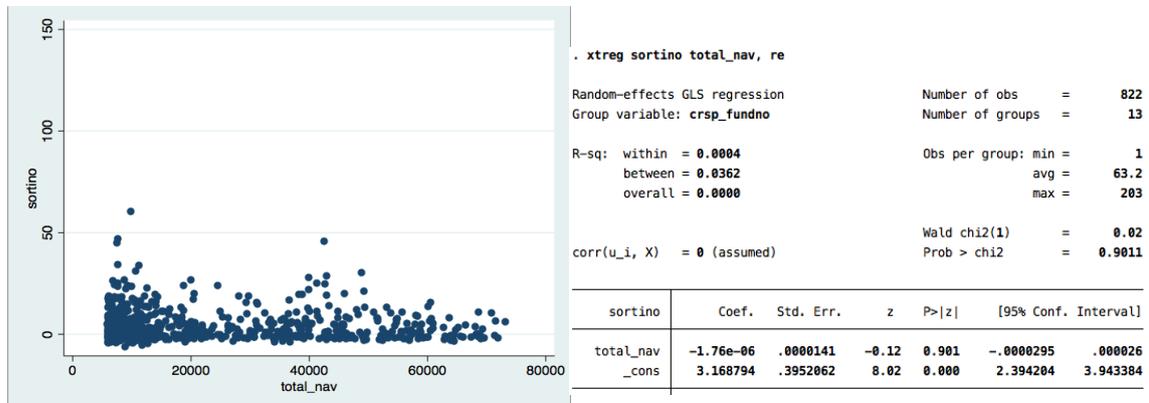| sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
total_nav	−5.84e−06	.0000154	−0.38	0.704	−.000036 .0000243
_cons	3.344284	.4597487	7.27	0.000	2.443193 4.245375

Figure 14 – NAV: 7000 regression analysis

- NAV > 8000
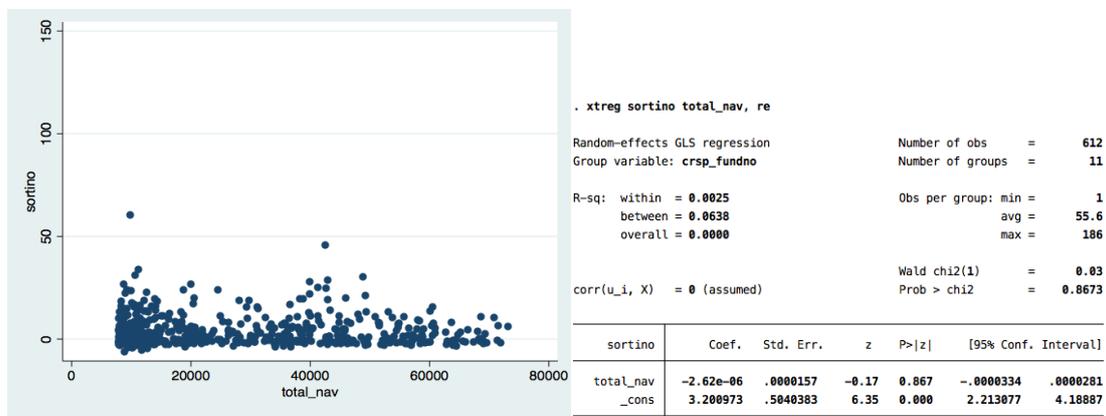
Figure 15 – NAV: 8000 regression analysis

- NAV > 10.000
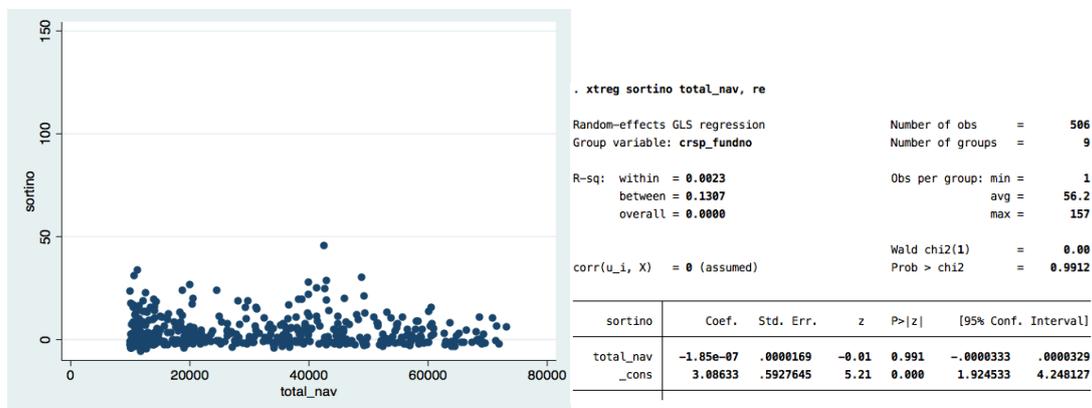
Figure 16 – NAV: 10.000 regression analysis

- NAV > 15,000

Figure 17 – NAV: 15.000 regression analysis

Once all 5 options have been tested, the one with the best results is selected. In the following Table, the main statistical indicators can be seen, showing a comparison between all the options.

Fund Size - NAV (MM$)	N of groups	R within	R between	R overall	P >|z|
> 6000	13	0,0004	0,0362	0	0,901
> 7000	12	0,0011	0,0219	0,0002	0,704
> 8000	11	0,0025	0,0638	0	0,867
> 10000	9	0,0023	0,1307	0	0,991
> 15000	5	**0,0132**	**0,3819**	**0,0022**	**0,387**

Tabla 4: Regression comparison for size división

As seen in Table 5, the regression analysis for funds bigger than 15.000 MM$ had the best results, being this size the one selected for the size division.

- Big funds (>15.000 MM$):

The regression analysis for this size is already done, as its results are the one used to justify the size cut to divide the funds into two sizes, small and big.
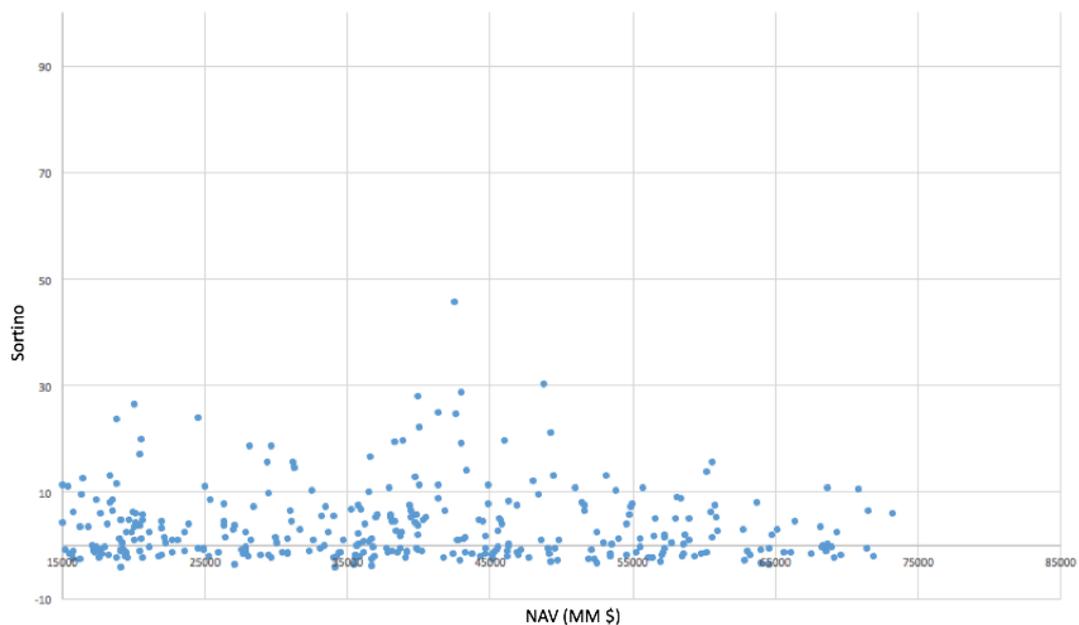
Figure 18 - Sortino VS. NAV (NAV >15.000 MM$)

Although the regression results are relatively good compared to the ones previously obtained, the most important size is size 1 as most of the observations are included in it.

- Small funds (<15.000 MM$):
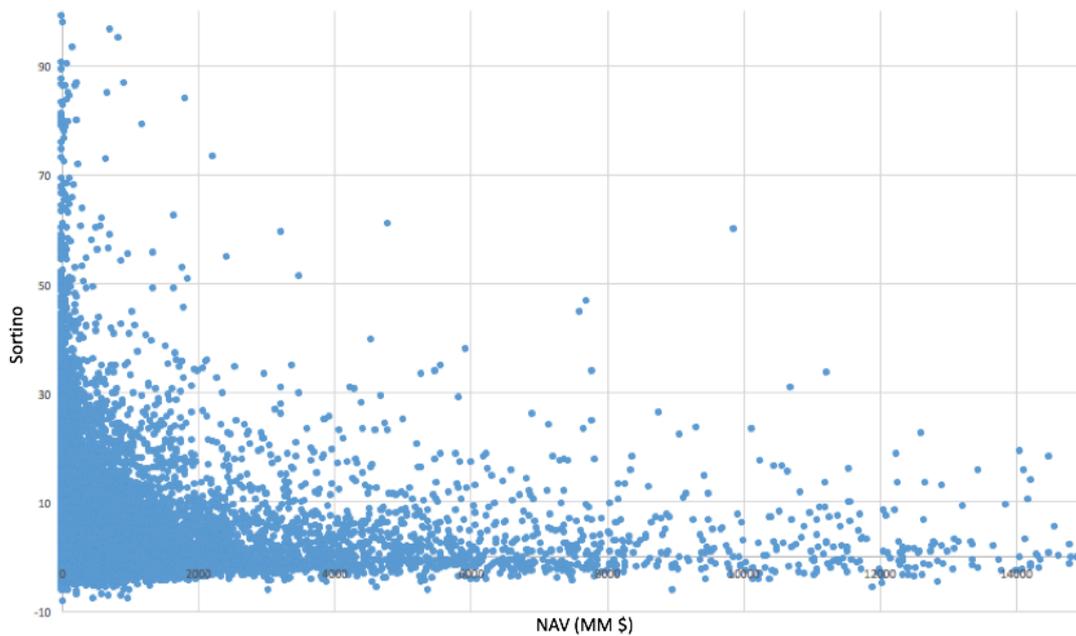
Figure 19 - Sortino VS. NAV (NAV < 15.000 MM$)

For the small funds regression, two different analyses are done. First, the regular linear regression analysis with panel data and random effects, which is the only one used until now, and second, a quadratic regression to see if this new model can explain more than the linear one.

o Linear

```
. xtreg sortino total_nav, re

Random-effects GLS regression              Number of obs     =      44921
Group variable: crsp_fundno                Number of groups  =        298

R-sq:  within  = 0.0000                    Obs per group: min =         47
       between = 0.0004                                   avg =      150.7
       overall = 0.0000                                   max =        204

                                           Wald chi2(1)      =       0.03
corr(u_i, X)   = 0 (assumed)               Prob > chi2       =     0.8556
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	-.0001944	.0010683	-0.18	0.856	-.0022882	.0018994
_cons	5.165129	2.299588	2.25	0.025	.6580189	9.672239

Figure 20 – Linear regression results for small funds.

After analyzing the results in Figure 19, no correlation was found again. R squared improved from 0.0001 to 0.0004, but as expected, it is not significant and the model doesn't explain the data.

o Quadratic relationship

By introducing NAV^2 as the independent variable instead of using simply the NAV, the model can be quadratic instead of a linear, which makes more sense given the previous results and the graphical distribution of points.

For this model, the dependent variable is again the Sortino, but the independent variable of the model is modified. Instead of using the NAV, this values are transformed into NAV^2. In this way, the model is transformed into a quadratic relationship, and hopefully the model is improved.

```
. xtreg sortino nav2, re

Random-effects GLS regression              Number of obs      =      30387
Group variable: crsp_fundno                Number of groups   =        283

R-sq:  within  = 0.0000                    Obs per group: min =          1
       between = 0.0022                                   avg =      107.4
       overall = 0.0000                                   max =        204

                                           Wald chi2(1)       =       0.25
corr(u_i, X)   = 0 (assumed)               Prob > chi2        =     0.6189

-------------------------------------------------------------------------------
     sortino |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
        nav2 |  -.0001792   .0003603    -0.50   0.619    -.0008853    .0005269
       _cons |   4.740185   1.326829     3.57   0.000     2.139647    7.340723
-------------+-----------------------------------------------------------------
     sigma_u |          0
     sigma_e |   201.1518
-------------------------------------------------------------------------------
```

Figure 21 – Quadratic regression results for small funds.

After analyzing the results in Figure 19, no correlation was found again. R squared improved from 0.0001 to 0.0004, but as expected, it is not significant and the model doesn't explain the data.

Although **no correlation** was found between size and performance as the results are statistically insignificant, the model is slightly improved from an R squared between of 0,0004 in the linear model to 0.0022 in the quadratic one, showing that the data is better explained with this model.

4.2.2   Individual Analysis

The next alternative analysis studies each fund individually. The objective of this analysis is to see how the funds of the sample behave as individuals, not as a group. This is done by selecting seven random funds of different sizes and running a simple linear regression model, with no panel data, to each one of them and study their individual behavior.

The analysis will try to see if there is any general tendency in the individual results, regarding indicators such as the regression coefficients sign for example.

- Fund 7922

Figure 22 – Linear regression result (Fund 7922)

- Fund 8060

Figure 23 – Linear regression result (Fund 8060)

- Fund 8309

Figure 24 – Linear regression result (Fund 8309)
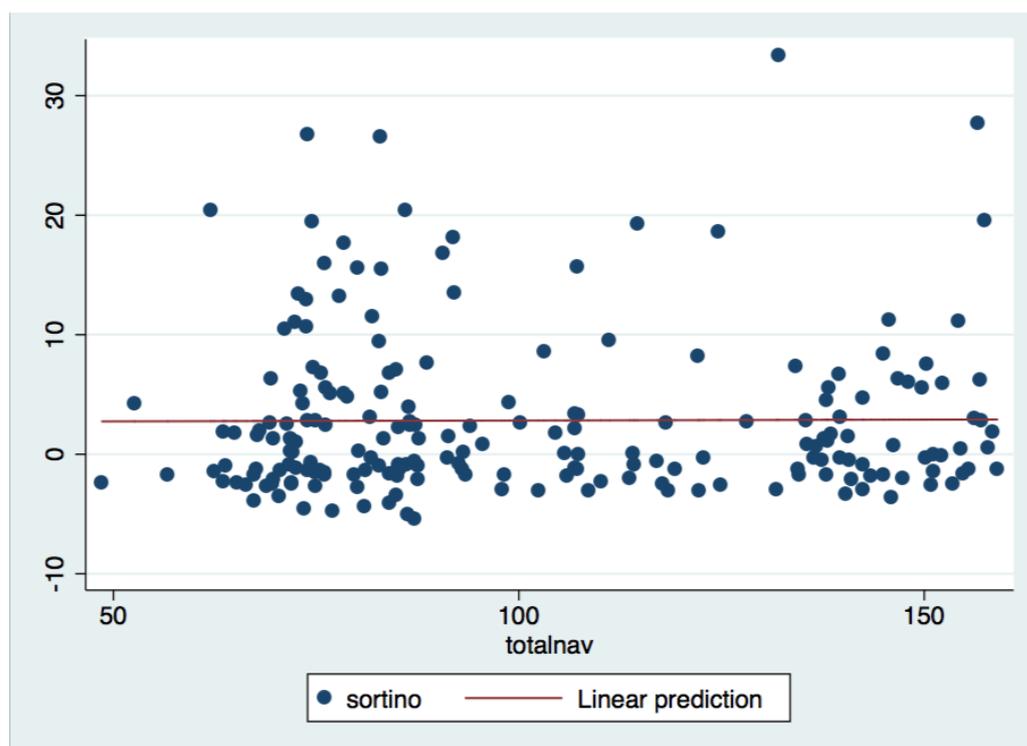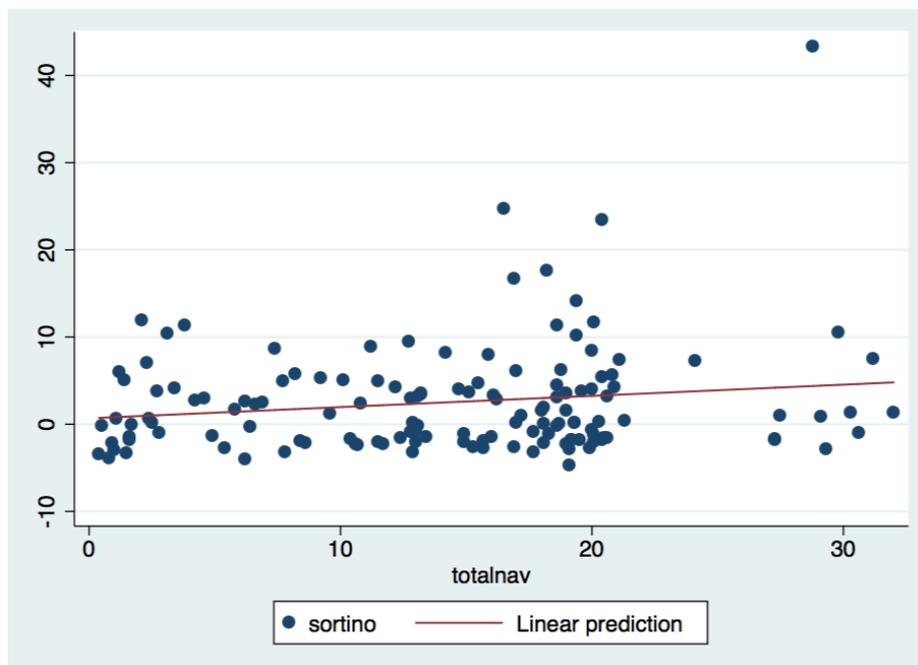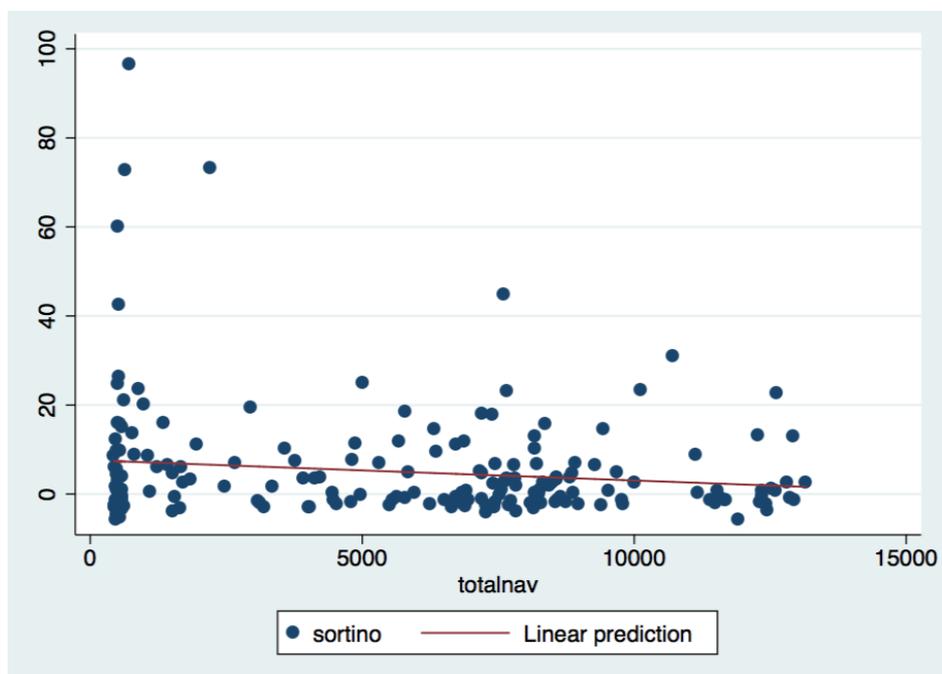
- Fund 7884
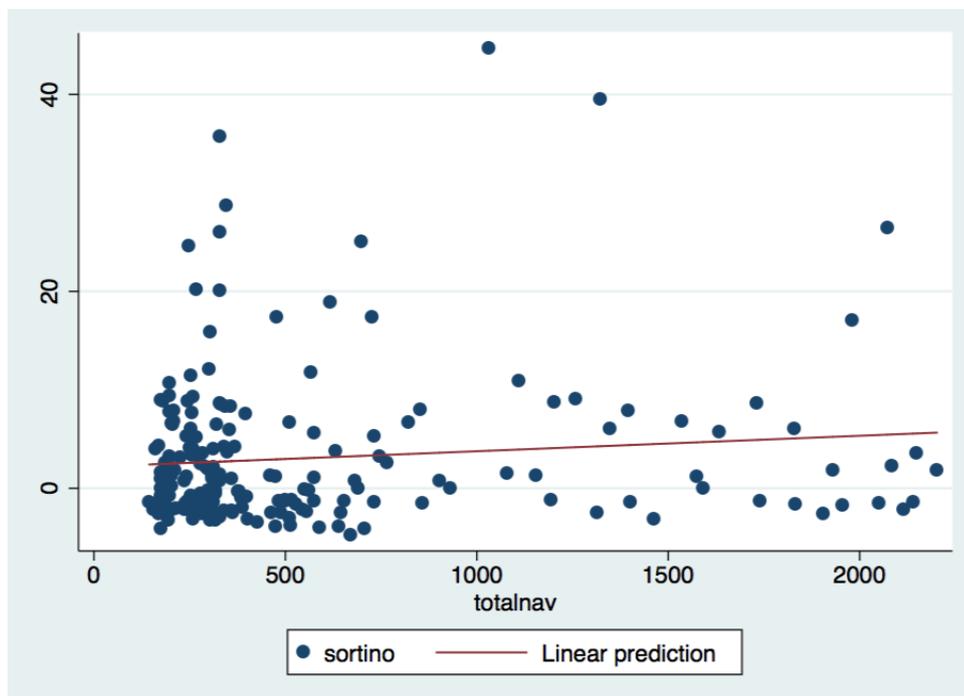
Figure 25 – Linear regression result (Fund 7884)

- Fund 8808
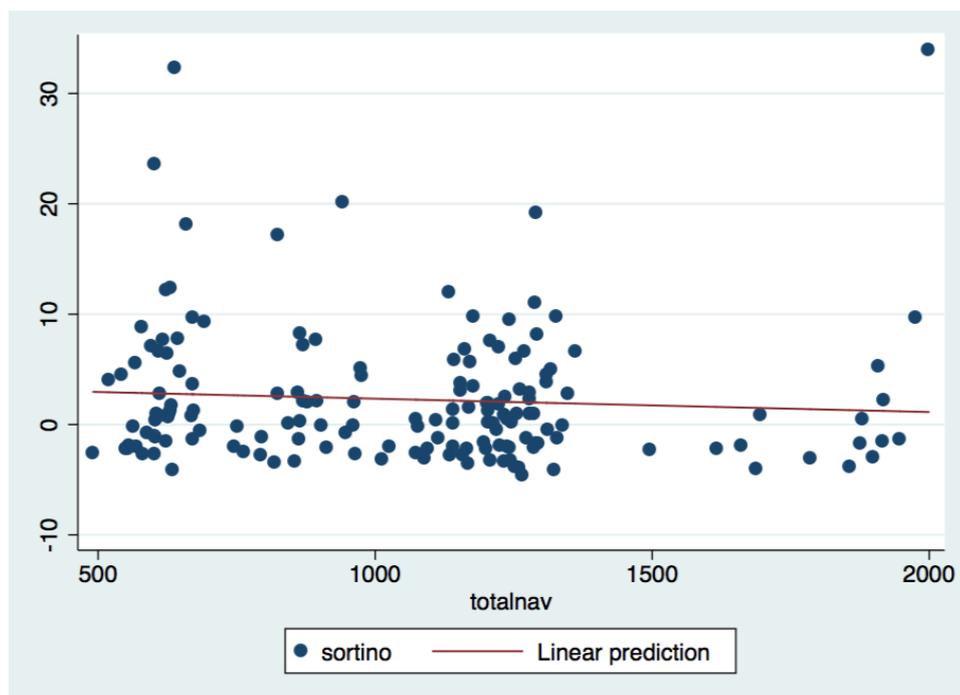
Figure 26 – Linear regression result (Fund 8808)

- Fund 9165
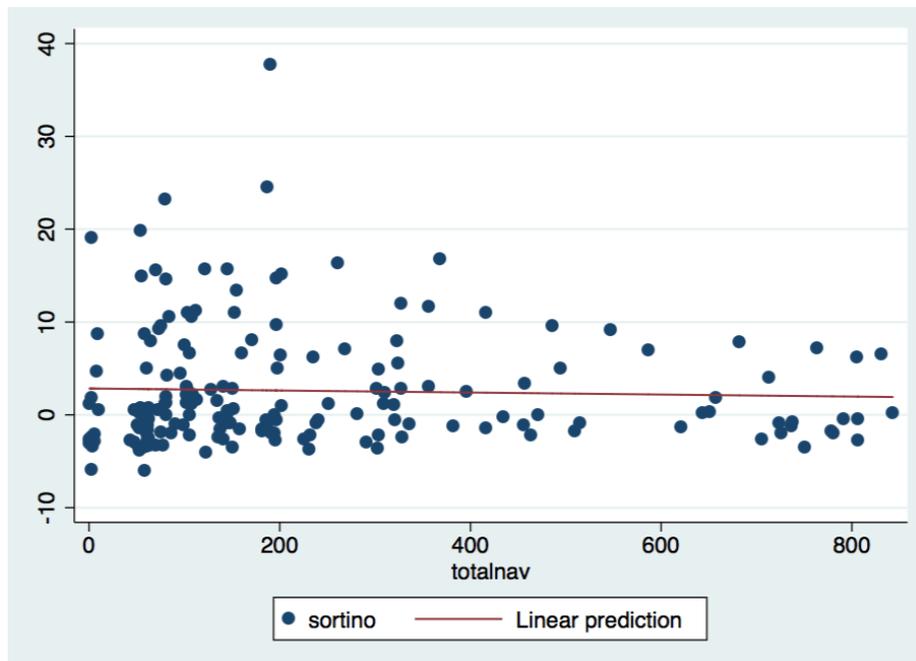
Figure 27 – Linear regression result (Fund 9165)

- Fund 8764
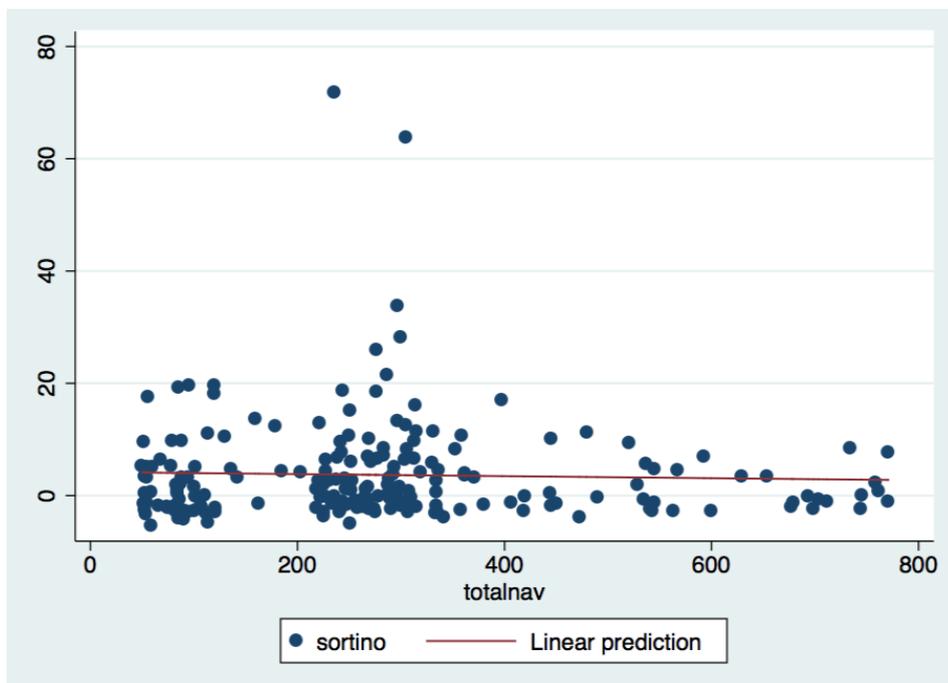
Figure 28 – Linear regression result (Fund 8764)

After applying a simple linear regression to the random selected funds, the results are analyzed and each indicator compared between all the funds. The following Table is precisely a comparison between the results of each fund's analysis:

Fund id	Min Size	Max Size	R squared	Correlation	Coeff.	P >|t|
7922	50	150	0	Positive	0,0014	0,927
8060	0	30	0,0255	Positive	0,128	0,061
7884	200	2000	0,0111	Positive	0,00157	0,134
8309	500	14000	0,0194	Negative	-0,00045	0,0049
8808	500	2000	0,0052	Negative	-0,0012	0,36
9165	50	800	0,0014	Negative	-0,0011	0,616
8764	100	800	0,0013	Negative	-0,00183	0,614

Table 5: Regression comparison for size división

As it can be seen, there is no clear trend in the results. Regarding the signs of the coefficients, there are funds that have have been m

odelled with a positive correlation and others with a negative one, showing no tendency in the correlation sign. In addition to this, there are some regressions that are statistically significant, with R-squared of 2,6% and almost 2% and P>ItI close to 0, which is relatively good regarding modelling funds' performance, while there are others that have results that are not statistically significant at all. Summing up, the results of this alternative analysis show no tendencies within the funds.

4.2.3 Period Analysis

The objective of this analysis if to see if by dividing the sample in the most important financial periods, there is a significant correlation between size and performance. Therefore, the data will be divided into 5 different samples and studied independently with 5 different regression analysis with panel data.

Procedure: Linear regression with panel data by periods:

- Period 1: Dotcom bubble (1998-2000)
- Period 2: Crash (2001-2003)
- Period 3: Boom (2004-2007)
- Period 4: Financial crisis (2008-2011)
- Period 5: Sovereign debt crisis (2012-2015)

Once the main economic periods are delimited, the data is analyzed. The results are shown in the following images:

- **Dotcom Bubble (1998-2000)**

```
Random-effects GLS regression              Number of obs     =        4153
Group variable: crsp_fundno                Number of groups  =         168

R-sq:  within  = 0.0004                     Obs per group: min =          3
       between = 0.0076                                    avg =       24.7
       overall = 0.0004                                    max =         27

                                           Wald chi2(1)      =        0.95
corr(u_i, X)   = 0 (assumed)               Prob > chi2       =      0.3303
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
total_nav	-.0004772	.0004902	-0.97	0.330	-.0014378 .0004835
_cons	5.475888	.5611646	9.76	0.000	4.376026 6.575751
sigma_u	3.9321554				
sigma_e	27.406431				
rho	.02017006	(fraction of variance due to u_i)			

Figure 29 - Period 1 Results

- **Crash (2001-2003)**

```
Random-effects GLS regression              Number of obs      =      7410
Group variable: crsp_fundno                Number of groups   =       253

R-sq:  within  = 0.0006                     Obs per group: min =         1
       between = 0.0080                                    avg =      29.3
       overall = 0.0000                                    max =        36

                                            Wald chi2(1)       =      0.13
corr(u_i, X)   = 0 (assumed)                Prob > chi2        =    0.7136
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	.000036	.0000981	0.37	0.714	-.0001563	.0002284
_cons	2.78436	.1425038	19.54	0.000	2.505058	3.063663
sigma_u	1.719285					
sigma_e	7.088675					

Figure 30 - Period 2 Results

- **Boom (2004-2007)**

```
Random-effects GLS regression              Number of obs      =     13121
Group variable: crsp_fundno                Number of groups   =       298

R-sq:  within  = 0.0004                     Obs per group: min =         2
       between = 0.0003                                    avg =      44.0
       overall = 0.0000                                    max =        48

                                            Wald chi2(1)       =      0.52
corr(u_i, X)   = 0 (assumed)                Prob > chi2        =    0.4689
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	-.0000135	.0000186	-0.72	0.469	-.0000499	.000023
_cons	2.606496	.0764081	34.11	0.000	2.456739	2.756253
sigma_u	.85224198					
sigma_e	6.4577167					

Figure 31 - Period 3 Results

- **Financial Crisis (2008-2011)**

```
. xtreg sortino total_nav, re

Random-effects GLS regression              Number of obs     =       11931
Group variable: crsp_fundno                Number of groups  =         283

R-sq:  within  = 0.0000                     Obs per group: min =          1
       between = 0.0001                                    avg =       42.2
       overall = 0.0000                                    max =         48

                                            Wald chi2(1)      =        0.01
corr(u_i, X)   = 0 (assumed)                Prob > chi2       =      0.9099
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	-.0001752	.0015477	-0.11	0.910	-.0032086	.0028583
_cons	9.44977	7.508591	1.26	0.208	-5.266798	24.16634
sigma_u	113.69708					
sigma_e	315.73261					

Figure 32 - Period 4 Results

- **Sovereign debt crisis (2012-2015)**

```
Random-effects GLS regression              Number of obs     =        8641
Group variable: crsp_fundno                Number of groups  =         208

R-sq:  within  = 0.0002                     Obs per group: min =          3
       between = 0.0004                                    avg =       41.5
       overall = 0.0000                                    max =         45

                                            Wald chi2(1)      =        0.09
corr(u_i, X)   = 0 (assumed)                Prob > chi2       =      0.7606
```

sortino	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
total_nav	-5.12e-06	.0000168	-0.30	0.761	-.0000381	.0000278
_cons	4.073264	.0931678	43.72	0.000	3.890658	4.255869
sigma_u	0					
sigma_e	8.4938731					

Figure 33 - Period 5 Results

The results show no correlation between size and performance in any period. The highest R-squared is not higher than 0,008, meaning the models don't explain the variation of the data, therefore, there is no relationship existent, according to the panel data linear regression models.

4.2.4   Segmentation based on size

The last alternative analysis would like to explain the tendency of the Sortino dispersion as the size of the funds changes.

As it can be seen in the following representation of the Sortino ratio against the NAV, visually, dispersion decreases as the size of the funds (NAV) increases.
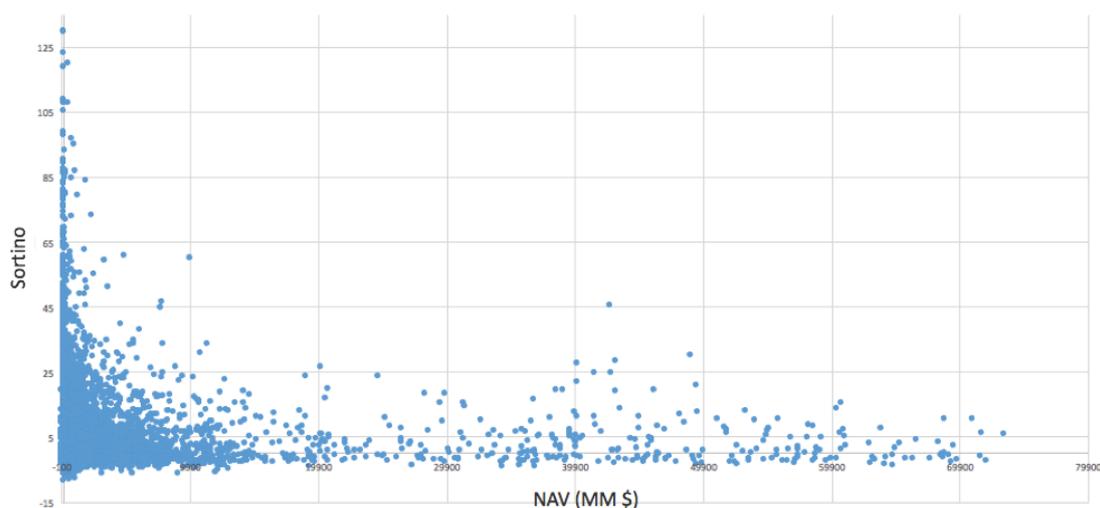
Figure 34 - Sortino Vs. NAV

To explain this, we propose the following procedure:

1st Reorganization of data.

- Data: 45.345 Sortino observations that include one Sortino per month, per fund, for the whole period studied.
- The data is separated into 43 intervals with respect to the funds' NAV, of approximately 1000 observations each.
- It is important for all the intervals to be of similar size in order to have intervals as representative as possible.

2nd Maximum and Minimum of each interval

Once the data is organized in intervals of similar size (approx.1000 Sortinos in each), the next step would be obtaining the Maximum and Minimum value of the observations included in each interval.

This way we would have a maximum and a minimum value of each interval, and by analyzing how the maximums and minimums of the intervals come closer to zero as size grows, the analysis could work as evidence to support the idea that the dispersion is reduced as fund's size is increased.

3rd Analysis

In the following chart, the number of the interval, the number of observations in each one, the maximum and minimum Sortino values in each one, and the representative NAV of the interval (calculated as the average NAV value of that bin), are shown:

Interval	Min	Max	Observ.	Rep. NAV
1	-6,2670138	142,640964	939	0,025
2	-5,5712238	826,223192	1489	0,085
3	-5,4336559	465,842157	1113	0,285
4	-4,7680379	118,890691	965	0,675
5	-5,6415587	72,9800247	1186	1,2
6	-5,6382921	45,8603306	1152	1,75
7	-5,2412614	87,3664352	1074	2,25
8	-5,9371078	35181,7079	1095	2,9
9	-5,6055643	83,1111871	977	3,7
10	-5,583846	81,0825958	964	4,7
11	-18,559964	161,441307	1063	6
12	-32,004877	64,0860913	1045	7,55
13	-5,1678637	129,599479	979	9,25
14	-8,4168935	97,687502	1009	11,25
15	-5,5715082	835,562345	1075	13,9
16	-4,8887182	129,692046	1019	16,95
17	-5,3313386	50,2401592	1078	20,25
18	-5,4338385	105,450724	1052	24,25
19	-5,1122425	43,6893911	1028	28,85
20	-6,3094097	185,184951	1028	33,6
21	-30,24297	77,7366604	1097	38,5
22	-5,4223393	79,643103	1007	44
23	-5,6610997	107,927654	1051	50,75
24	-6,0170842	86,1295737	1017	59
25	-5,6161276	90,2728466	1044	68,25
26	-5,3625387	123,11589	1093	79
27	-6,2119329	255,767635	998	91
28	-4,9721597	107,686845	1032	104
29	-5,1123252	303,25764	1024	120,5
30	-4,8952398	93,1708602	1025	142
31	-5,2119006	159,212155	1016	169,5
32	-7,1791078	86,0415181	1036	202,5
33	-6,5971281	448,211305	1006	242,5
34	-5,1497939	63,7412838	996	292,5
35	-6,3088392	50,1399506	991	332,5
36	-5,498509	107,813445	1073	387,5
37	-7,9618803	119,855527	1019	475
38	-7,0333697	72,8261344	1021	590
39	-7,2392156	96,5660881	994	760
40	-7,8573229	86,6800021	1056	1030
41	-5,5401479	62,3461818	1000	1475
42	-6,2746988	175,509724	1018	2575
43	-6,2990811	60,8985645	1360	4200

Table 6: Data considered of each bin

MAXIMUM

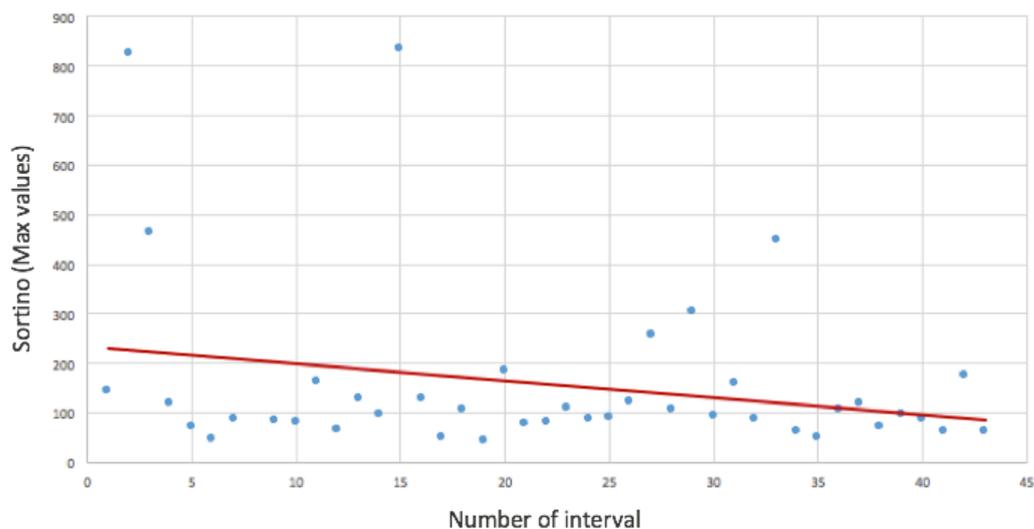In the following graphics, the maximum values of each interval are represented:

Figure 35 – Maximum Sortino values of each interval

Graphic 34 shows the maximum Sortino value of each interval considered. The trend line shows a reduction in the maximum Sortino values as the number of the interval increases.

However, for the linear regression of the values, the independent variable used in the analysis is the interval's representative point which takes the value of the average NAV of all the values included in each interval. For the dependent variable, the maximum Sortino values are taken, as it can be seen in the next graphic:
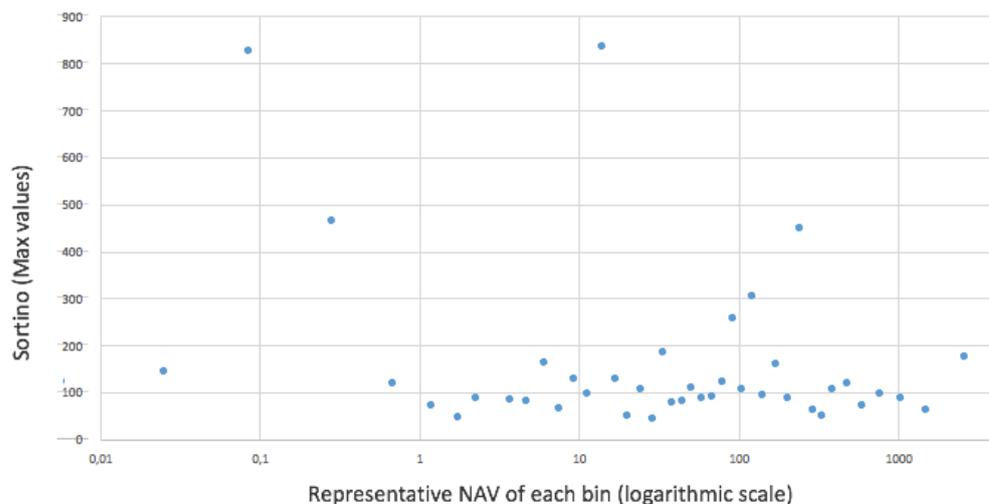
Figure 36 – Maximum Sortino values for each interval's representative NAV

The results of the linear regression analysis between the maximum Sortinos and the interval's middle point are shown in the following chart:


```
. regress max middlepoint

      Source |       SS       df       MS              Number of obs =      43
-------------+------------------------------           F(  1,    41) =    0.19
       Model |  5607897.62      1  5607897.62           Prob > F      =  0.6631
    Residual |  1.1938e+09     41  29117313.7           R-squared     =  0.0047
-------------+------------------------------           Adj R-squared = -0.0196
       Total |  1.1994e+09     42  28557565.7           Root MSE      =    5396

-------------------------------------------------------------------------------
         max |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
  middlepoint |  -.4751944   1.082797    -0.44   0.663    -2.661947    1.711558
       _cons |   1124.228   892.4579     1.26   0.215    -678.1272    2926.583
-------------------------------------------------------------------------------
```

Figure 37 - Linear regression between maximum Sortinos and NAV of bins

Even though the value of the R-squared is 0,0047, meaning the relationship between the variables is not very representative, the linear regression coefficient is -0,475, therefore, there is a negative relationship between the maximum Sortino values and the size of the funds.

As size grows, the maximum values of the Sortino are reduced with a rate of 0,475 per NAV unit, which explains what expected in the first place by visually analyzing the data, which is that the dispersion decreases as the size of the funds (NAV) increases.

MINIMUM

As in the previous study with the maximum values, the minimum Sortino values of each interval are represented in the following graphic:
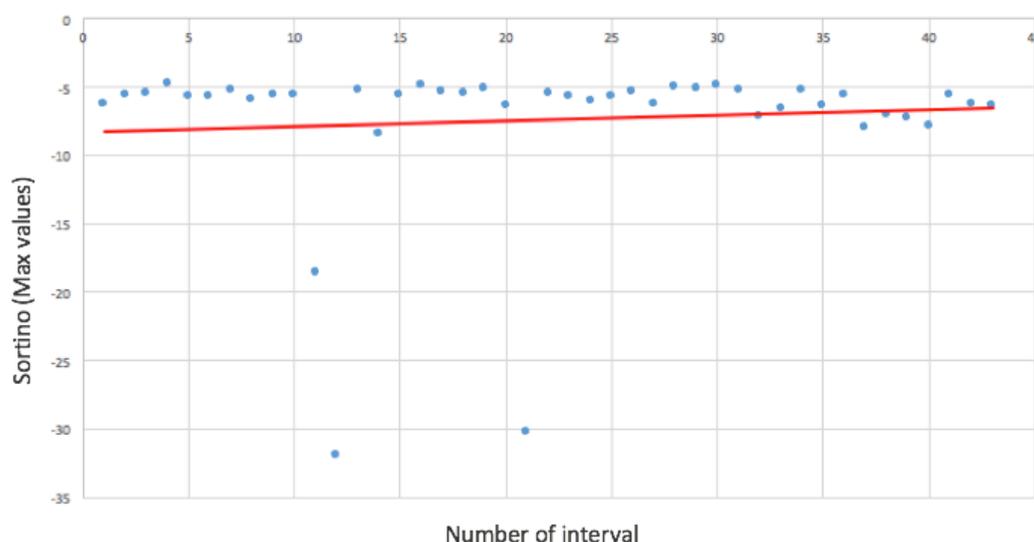

Figure 38 – Minimum Sortino values of each interval

This graphic shows the minimum Sortino value of each interval considered. The trend line shows an increase in the minimum Sortino values, meaning they come closer to zero as NAV increases.

As in the other analysis, for the linear regression of the values, the independent variable used in the analysis is the interval's average NAV, while for the dependent variable, the minimum Sortino values are taken, as it can be seen in the next graphic:
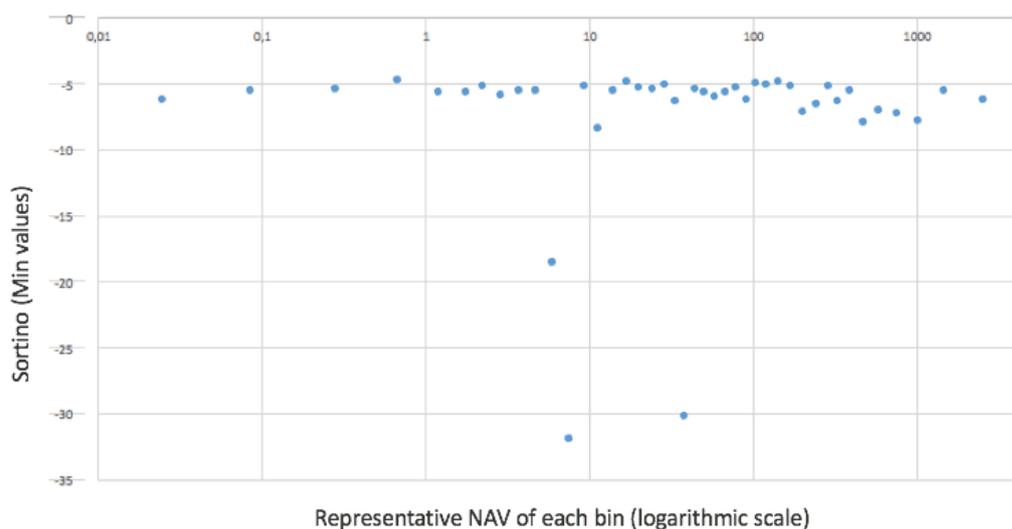

Figure 39 – Minimum Sortino values for each interval's representative NAV

```
. regress min middlepoint
```

Source	SS	df	MS		Number of obs =	43
					F(1, 41) =	0.18
Model	6.0720076	1	6.0720076		Prob > F =	0.6719
Residual	1367.7146	41	33.3588927		R-squared =	0.0044
					Adj R-squared =	-0.0199
Total	1373.78661	42	32.709205		Root MSE =	5.7757

min	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
middlepoint	.0004945	.001159	0.43	0.672	-.0018461	.0028351
_cons	-7.510054	.9552519	-7.86	0.000	-9.439225	-5.580884

Figure 40 - Linear regression between maximum Sortinos and NAV of bins

As in the analysis of the maximum values, the value of the R-squared in this analysis is 0,0044, meaning the relationship between the variables is also not very representative. However, the linear regression coefficient is 0,0005, which means there is a positive relationship between the minimum Sortino values and the size of the funds.

As size grows, the minimum values of the Sortino are increased with a rate of 0,0005 per NAV unit, which explains what expected by visually analyzing the data, which is that the dispersion, taking into account both, the maximum and minimum values, decreases as the size of the funds (NAV) increases.

This evidence would support the idea that small funds are less diversified. While biggest funds have diversification and risk control mechanisms more standard, small funds might be more different among them.

5. Conclusions

This project analyzes whether the performance of an investment fund is influenced by its size. The database used for study is the CRSP Mutual Fund Database and it provides all the information needed for the project. The CRSP Mutual Fund Database was designed for research and educational purposes and has supplied scholar researchers with high quality data since 1960. The sample includes daily observations from 1998 to 2015, taking into account 300 funds, giving the study almost 1 million of total observations, considered an excellent size for an accurate statistical study as its also the typical size of researches of this topic.

There have been numerous studies that have tried to find a correlation between the size of a fund and its performance. There are studies that support that small funds might have a higher performance by putting all their money into their best ideas or investment options. Others state that the bigger in size the better, as they have more resources available, higher ability to attract and retain talent, better access to companies and clients, and a higher efficiency and bargaining power with brokers.

After analyzing the data, no statistical evidence was found to support a linear relationship between size and performance, represented by the Sortino ratio. The indicators of the analysis have not shown enough evidence to support the relationship between both variables. This would support the opinion of those authors that say that the performance of a fund is independent from its size.

Regarding this performance measure, the only think than can be stated after having done numerous alternative analyses, is that the dispersion of the Sortino for small funds is higher than for big funds. This evidence would support the idea that small funds are less diversified. While biggest funds have diversification and risk control mechanisms more standard, small funds might be more different among them.

These results have important practical uses for portfolio managers and investors. For risk averse investors it might be better to invest in small funds, if they know well the investment style and risk control mechanisms of the fund. If they don't completely understand or now how the fund operates, the risk might be lower when investing in a big fund. Cost considerations are secondary.

6. References

Cheong Sing, T. (n.d.). Effects of Expenditures and Size on Mutual Fund Performance. *Singapore Management Review*, 31-45.

Indro, D., Jiang, C., Hu, M., & Lee, W. (1999, June). Mutual Fund Performance: Does Fund Size Matter? *Association for Investment Management and Research*, 74-87.

Laurent Bodson, L. C. (2011). Does size affect mutual fund performance? A general approach. *Journal of Asset Management*, 163-171.

Loeb, T. (1983). Trading Cost: The Critical Link between Investment Information and Results. *Financial Analysts Journal, 39*, 39-43.

Philpot, J., Hearth, D., Rimbey, J., & Schulman, C. (2001). Active management, fund size, and bond mutual fund returns. *The Financial Review*, 115-126.

Shawky, H., & Liuling Li. (2006). Optimal Asset Size for U.S. Small-Cap Equity Mutual Funds. *The Journal of Investing*, 79-86.

Torres-Reyna, O. (2007, Decembar). Panel Data Analysis Fixed and Random Effects using Stata (v 4.2). *Data & Statistical Services - Princeton University*, 1-39.

Zera, S., & Madura, J. (2001). The empirical relationship between mutual fund size and operational efficiency. *Applied Financial Economics*, 243-251.

Investopedia, (2012). Investment Fund Definition | Investopedia. [online] Available at: http://www.investopedia.com/terms/i/investment-fund.asp [Accessed 22 Feb. 2016].

Fidelity Investments (1998-2016). Growth versus Value Investing - Fidelity. Availableasshttps://www.fidelity.com/learning-center/investment-products/mutual-funds/growth-vs-value-investingn.d.   Web.[Accesed   07 Mar. 2016].

CRSP (1960)"CRSP Survivor-Bias-Free US Mutual Fund Guide for CRSPSift." CRSP
    Survivor-Bias-Free US Mutual Fund Guide for CRSPSift. Web.[Accesed 09
    Mar. 2016]

Investopedia, (2012). Mutual Fund Definition | Investopedia. [online] Available at:
    http:// http://www.investopedia.com/terms/m/mutualfund.asp [Accessed 24
    Feb. 2016].