

A Work Project, presented as part of the requirements for the Award of a Master Degree  
in Finance from the NOVA – School of Business and Economics

**FINANCIAL STATEMENT INFORMATION FOR VOLATILITY  
ESTIMATION**

JOÃO ALBUQUERQUE DA CRUZ ALVES | 860

A Project carried out under the supervision of Professor Paulo M. M. Rodrigues

May, 2016

## FINANCIAL STATEMENT INFORMATION FOR VOLATILITY ESTIMATION

### **Abstract**

This work extends Sridharan's (2015) results, who found a significant relationship between financial variables and realized volatility. In particular, the introduction of *Size*, *Research and Development Expenditures*, *Sales Growth*, *Cash Flow Volatility*, *Earnings Opacity*, *Leverage*, *Return on Assets* and *Equity Book-to-Market Ratio* in a model based on the volatility implied in option market prices presented improved results. Applying a similar methodology to a different set of data, it is found that only three of those variables affect realized volatility in my sample. *Leverage* and *Equity Book-to-Market Ratio* have a negative impact and *Return on Assets* a positive impact. It is hypothesized that financial variables should also be useful in the construction of a traditional ARCH model. This is confirmed empirically and it is shown that the better volatility forecasts, provided by the introduction of these financial variables, can be used to construct a successful investment strategy that outperforms the market.

### **Keywords:**

Volatility    Modeling    Financial Variables    ARCH

## **I. Introduction**

Eugene Fama (1970), in his influential article “Efficient Capital Markets”, termed as efficient a market in which prices always ‘fully reflect’ available information and this became known as EMH (Efficient Market Hypothesis). A semi-strong form of this EMH implies that all publicly available information is already incorporated in prices and, consequently, it is not possible to detect mispriced securities and design a profitable investment strategy using information such as financial statement figures. Despite the widely acceptance and ‘intellectual dominance’ of the efficient-market revolution (Malkiel, 2003), a counterrevolution is going on, conducted by fundamental analysts, who believe it is possible that, in the short run, the market misprices securities, and by econometricians, who argue that stock returns are, to a considerable extent, predictable.

The discussion, arguments and results about the efficiency of the market are not very relevant for this dissertation. What is important to highlight is that the current literature presents a wide range of studies incorporating financial statement information in the prediction of returns, such as Caneghem, *et al*, (2002), Alexakis, *et al*, (2010) and Goslim, *et al*, (2012). By contrast, however, regarding volatility estimation, the existing models are much more closed and leave out most of the information about the firms which is usually considered in returns modeling. I believe that there are, at least, two reasons for this. First, the specific characteristics of financial data, namely the evidence of volatility clustering, low decay correlations, volatility persistence and ‘leverage’ effects have resulted in a competitive demand for a time varying conditional variance model, able to incorporate all the specific characteristics of a returns series. Second, the dissemination of stock options and the Black and Scholes formula for its prices, resulted in an efficient market approach to volatility modeling, similar to that of equation (1) (see for instance, Claessen and Mittnik, 2002), that is,

$$\sigma_{i,t+\tau}^{RV} = \beta_0 + \beta_1 \sigma_{i,t,\tau}^{IV} + \beta_2 Year + \varepsilon_{i,t} \quad (1)$$

where  $\sigma_{i,t+\tau}^{RV}$  stands for realized volatility of firm  $i$  between  $t$  and  $t + \tau$ ;  $\sigma_{i,t,\tau}^{IV}$  stands for the implied volatility of firm  $i$  between  $t$  and  $t + \tau$  and  $Year$  is a vector of year fixed effects, with some authors showing that implied volatility for an at-the-money option is an unbiased estimator for the average volatility over the remaining life of the option (Christensen and Strunk, 2002).

Among all these models and contrasting to what is found in the returns literature, financial statement information was, surprisingly, never hypothesized to be relevant for the explanation of volatility. An original study conducted by Sridharan (2015) tested the usefulness of adding financial information into a volatility model, similar to that in equation (1), but controlling for past volatility,  $\sigma_{i,t-1}^{RV}$ , and liquidity,  $Spread$ , resulting in

$$\sigma_{i,t+\tau}^{RV} = \beta_0 + \beta_1 \sigma_{i,t,\tau}^{IV} + \beta_2 V\_FSV_{i,t}^j + \beta_3 Spread_{it} + \beta_4 \sigma_{i,t-1}^{RV} + \beta_5 Year + \varepsilon_{i,t} \quad (2)$$

where  $V\_FSV_{i,t}^j$ , represents a set of financial statement variables that were separately included in the equation.

Considering the logical assumption that a fully efficient option market would imply  $\beta_1$  not statistically different from one and  $\beta_2$  not statistically different from zero, Sridharan (2015) verifies in her empirical analysis that some financial variables are able to provide additional information to the market expectations about future volatility reflected in options prices. The obtained results were important in the sense they attested implied volatility as a biased estimator of future realized volatility, but especially because they revealed the importance of fundamental variables for volatility estimation.

However, a limitation of these results is that they are only applicable to implied volatility models. It is obvious to note that it is not always possible to estimate a model

of this nature, because options on a certain stock are not always available. Analyzing worldwide stocks, one realizes about the existence of many companies which, despite having equity shares publicly traded on a stock exchange, do not have marketed stock options. The main reason for this, is the fact that some companies do not have enough dimension to assure liquidity in derivative instruments. Also, it is unusual to find a highly developed derivatives industry, such as in America and Asia. Still, one may find it necessary to estimate returns volatility on a stock without options. Some examples may be highlighted: risk management purposes, evaluation of portfolio performance, determination of option prices in a first stock option issue.... In these cases, the use of a traditional volatility model is required<sup>1</sup>.

My goal is to extend Sridharan's (2015) results, by analyzing whether accounting information can also be used in the construction of a traditional ARCH family volatility model. I use a different set of data (medium European companies, instead of big American companies), which also contributes to extend the results to another geographical area and to companies with different fundamental characteristics. In addition, I also intend to evaluate an investment strategy based on the results obtained. My research is divided in five parts. The next section discusses relevant literature for my analysis, section III describes the sample used, section IV presents the methodological design of the study, section V shows the results obtained and finally, in section VI, some conclusions are drawn.

## **II. Literature Review**

As previously mentioned, Sridharan (2015) was the first study considering the hypothesis of using financial statement information for volatility estimation. The study

---

<sup>1</sup> I use the word *traditional* to describe models that do not use implied volatility such as Autoregressive Conditional Heteroscedasticity (ARCH) models.

was initially published as a working paper, in 2012, and obtained good academic acceptance. For example, in a similar study, Goodman, Neamtiu and Zhang (2013), using data from different companies and testing a different set of financial variables, confirmed that information about a firm's fundamental volatility is not fully priced in option contracts. Moreover, in the second edition of the book "Volatility Trading", Sinclair (2013, Chapter 4.5), presents a summary of Sridharan's results to teach the importance of fundamental information in volatility forecasting.

In the first part of my research I will follow the methodology used by this author. Therefore, it will be useful to describe in more detail some aspects of the referred paper. A sample comprising 1,126 firms with quarterly observations during 14 years of data was used to test two hypothesis. The first was that financial variables are related with future equity returns volatility and the second was that options markets are not fully efficient. As explained in the previous section, I do not intend to explore the discussion about the options market efficiency and, so, I will only focus on the results obtained by testing the first hypothesis. The following model was considered,

$$\sigma_{i,t+\tau}^{RV} = \beta_0 + \beta_1 V\_FSV_{i,t}^j + \beta_2 Year + \varepsilon_{i,t} \quad (3)$$

where V\_FSV represents eight fundamental variables considered to be potentially relevant to affect realized volatility: *Size*; *Research and Development Expenditures*; *Sales Growth*; *Cash Flow Volatility*; *Earnings Opacity*; *Leverage*; *Returns on Assets*; and *Equity Book-to-Market Ratio*. Each variable showed evidence to be related with realized volatility, as the coefficients were all significant at a 1 percent significance level. *Size* and *Return on Assets* were negatively related with the dependent variable, while the remaining variables were positively related with it. The results showed that the variables tested are important for the estimation of volatility and that is why I will use a database with

variables similar to those. In the next subsection, I review the literature that justifies why these specific eight variables are apparently related with volatility.

## **II.1 Financial Statements and Volatility**

Although the referred study was original directly relating fundamentals with volatility estimation, prior literature had already documented significant relations between financial information and measures of risk or uncertainty, such as default risk, the incidence of extreme returns and growth opportunities. In its most basic financial definition, volatility is also a measure of risk or uncertainty and, therefore, we can suppose that a factor related with default risk, for instance, is also likely to affect volatility. Following this reasoning:

*Size* – With mixed results there is some literature relating default risk with firm size. For example, Bonfim (2007) notes that firms in default tend to be slightly bigger firms, but Eklund *et al* (2001) estimated a negative relationship between default probability and firm size, as measured by total assets. However, larger firms are less likely to report earnings surprises (Barton and Simko, 2002) and have a lower probability of becoming a target of a merger or acquisition (Palepu, 1986). Size is, therefore, expected to be negatively related with volatility.

*Personal Expenses to Assets* – I use this variable to proxy for Research and Development (R&D) expenses. R&D expenses are frequently used as a measure of growing opportunities for a firm (see for instance, Grullon, Lyandres and Zhdanov, 2012), since they are incurred in anticipation of future products and revenues. I expect that variables that measure investment or growth opportunities should be positively correlated with returns volatility, since they indicate, to some extent, uncertainty about a firm's future performance. In fact, Chan, Lakonishok and Sougiannis (2001) verify a positive association between R&D intensity and return volatility. They note that R&D intensive firms tend to pay little or no dividends and that low dividend stocks tend to have higher

volatility. Although the authors had verified this relationship, they did not develop a model for volatility estimation, neither went further investigating the relationship between other financial variables and returns volatility. The reason why I use personal expenses to assets instead of R&D expenses is that most of the firms in the sample do not report R&D in their statements. Nevertheless, it is plausible to assume that personal expenses and R&D expenses are positively related, since the variation of both might represent an investment or disinvestment pursuit by a firm.

*Sales Growth* – Sales Growth can also be considered as an indicator of growth opportunities and, therefore, the same reasoning used for R&D expenditures can be applied. Moreover, Baneish (1999) shows that higher sales growth may be indicative of a higher probability of earnings manipulation and Sun (2009) notes that the suspicion of manipulation and anticipation of restatements increases uncertainty and hence volatility. I expect sales growth to be positively related with returns volatility.

*Cash Flow Volatility* – Corporate risk management theory indicates that cash flows volatility is negatively valued by investors. One reason for this is revealed by Minton and Schrand (1999), who empirically showed that higher cash flow volatility not only increases the likelihood that a firm will need to access capital markets, but also increases the costs of doing so. Consistent with that, Allayannis and Weston (2005) estimate that one standard deviation increase in cash flow volatility should result in a decrease in the firm value between 0 and 14 percent. Considering that portfolios composed by firms with lower market capitalization tend to have higher volatility than portfolios of firms with higher market capitalization (Fama and French, 2008), I expect cash flows volatility to be positively related with equity volatility.

*Earnings Opacity* – Hutton, Marcus and Tehranian (2009) developed a measure of earnings opacity to proxy for managed earnings. They conclude that higher earnings

opacity indicates less disclosure of firm-specific information, which suggests higher probability of managed earnings. Moreover, they show that more opaque firms are more likely to experience stock price crashes. I expect a positive relationship between earnings opacity and returns volatility.

*Leverage* – Some literature established a link between leverage levels and probability of default. Hui, Lo and Huang (2007), Beaver (1966) and Kaplan and Gabriel (1979) all show a positive relationship between those two variables. Considering that default probability increases uncertainty about future returns, it should increase volatility as well. I expect returns volatility to increase with a firm's leverage.

*Returns on Assets* – Along with leverage, another financial ratio commonly used in predictive default models is the return on assets (ROA), a profitability measure. Maricia and Vintila (2012) and Beaver (1966), for instance, present a negative correlation between ROA and default probability. A potential reason is that capital markets are concerned about a firm's ability to repay its debt and profitability is an important indicator of that (Beaver, McNichols and Rhie 2005). In opposition to leverage, I expect ROA to be negatively related with returns volatility.

*Equity Book-to-Market Ratio* – Some studies, such as Fama-French (1994), suggest that stocks with higher Book-to-Market (BTM) ratios earn higher returns than low BTM stocks and give as possible explanation that high BTM equity firms are assigned with a higher risk premium, due to their higher risk of distress. Consistent with that, Chen and Zhang (1998) found that high BTM are usually firms under distress, with high financial leverage and substantial future earnings uncertainty. I expect, therefore, a positive relationship between BTM and returns volatility.

## II.2 Volatility Estimation

When modeling financial market volatility, one must consider several salient features that financial time series usually present, known as stylized facts. These include fat tail distributions of risky asset returns, asymmetry and mean reversion, co-movements of volatilities across assets and financial markets (Poon and Granger, 2003) and the fact that price changes are not independent over time, firstly documented by Mandelbrot (1963, p.418), who noted that “large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes”, a phenomenon known as ‘volatility clustering’. Some statistical methods were proposed to capture the dynamic behavior of volatility. The simplest is the Random Walk model, where  $\sigma_{t-\tau}$  is used as a forecast for  $\sigma_t$ . Other methods based on past standard deviations were introduced, from which the Exponentially Weighted Moving Average (EWMA) presented by Roberts (1959) stands out. However, since Engle (1982), a new class of stochastic volatility processes called Autoregressive Conditional Heteroskedasticity (ARCH) has been explored by researchers and have been trying to capture all the specific features of a financial time series. Engle (1982) presents the ARCH model in its basic form, recognizing immediately its potential applications to financial data and maximum likelihood estimators as the most efficient. The method allows the conditional variance to change over time as a function of past errors, leaving the unconditional variance constant. Bollerslev (1986) introduces the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, an ARCH generalization, to allow for past conditional variances in the current conditional variance equation. Several adaptations such as E-GARCH, CGARCH, TGARCH, PARARCH, AGARCH and GARCH-M have been introduced since then and it is a common methodology to apply all these different

specifications to a financial time-series, choosing the best fitting model (see, for instance, Gökbulutand and Pekkaya, 2014).

### III. Data

My sample is composed by 92 quarterly observations of 173 firms from March 2004 to June 2015, resulting in a panel with 15,916 observations. Firms were divided into 10 industries according to the Global Industry Classification Standard, as shown in Table 1.

Sectors				
Sector	Obs	Sector	Obs	
Energy	12	Health Care	6	
Materials	19	Financials	36	
Industrials	31	Information Technology	5	
Consumer Discretionary	21	Telecommunication Services	7	
Consumer Staples	22	Utilities	14	

**Table 1:** Firms by Sector

Bloomberg provided all the data collected. The criteria for selection consisted in picking all the European firms that, in September 2015, were publicly traded but did not have stock options issued and trading on an exchange. Most of the companies are from countries where financial markets are significantly developed, while the derivatives industry is still recent or poorly developed (see Table 2).

Country					
Country	Obs	Country	Obs	Country	Obs
Austria	19	Hungary	4	Luxembourg	1
Bulgaria	14	Iceland	4	Poland	25
Czech Republic	9	Ireland	2	Portugal	16
Estonia	11	Lithuania	19	Slovakia	1
Greece	17	Latvia	26	Slovenia	5

**Table 2:** Firms by country

Because companies usually report their statements with two months delay, I make sure that all information is available for market participants at the quarter in which the information is used. For instance, at quarter  $t$ , investors will have access to the daily stock price, but will only have access to the balance sheet of the quarter  $t-1$ . By lagging the

variables I guarantee that results can be used to forecast without falling into a forward looking bias.

The construction of the variables purposely follows a similar scheme used by Sridharan (2015). *Size* is the value of total assets as reported in the quarter t-1. *PEA* is the ratio between personnel expenses and total assets as reported at quarter t-1, where personnel expenses include wages and salaries, social security, pension, profit-sharing expenses and other benefits related to personnel. *SGI* is the quarterly growth of the estimate comparable sales figure between t-2 and t-1.  $\sigma_{i,t}^{CF}$  is the standard deviation of operating cash flows over total assets from t-11 to t-1, where operating cash flows is generally calculated as

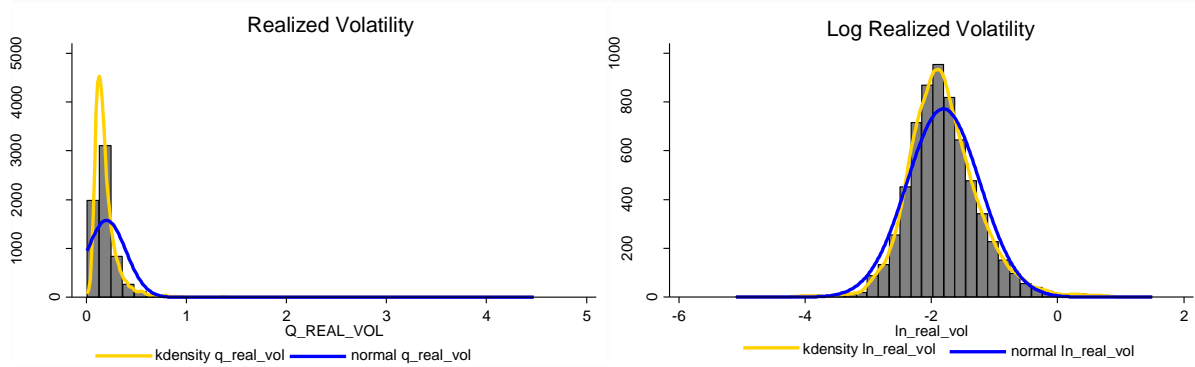
$$OCF = Net\ Income + Depreciation\ \&\ Amortization + Other\ Noncash\ Adjustments \\ + Changes\ in\ Non-cash\ Working\ Capital$$

*Opacity* is the sum of discretionary accruals ( $D\ Acc_{i,t}$ ) over t-1, t-2 and t-3 where  $D\ Acc$  is the residual of the cross sectional estimation

$$\frac{Net\ Income_{i,t} - Operating\ CF_{it}}{Assets_{it}} = \beta_0 + \beta_1 \frac{1}{Assets_{it}} + \beta_2 \frac{\Delta Sales_{it}}{Assets_{it}} + \beta_3 \frac{PPE_{i,t}}{Assets_{it}} + \varepsilon_{i,t} \quad (4)$$

*Lvg* is the ratio between total liabilities and total assets in quarter t-1. *ROA* is the average from t-5 to t-1 of the ratio between 12 month net income and total assets. *BTM* is the ratio between book value of equity and market value of equity, where the former is calculated as the difference between total assets and total liabilities at the quarter t-1 and the latter is market capitalization of the company at the end of the day before the end of quarter t. *Q\_REAL\_VOL* is the quarterly realized volatility, measured as the standard deviation of the daily returns, between the last days of quarters t and t+1, multiplied by  $\sqrt{60}$ .

Figure 1 shows that realized volatility is negatively skewed and leptokurtic compared to a normal distribution. Figure 2 indicates that log realized volatility roughly fits a log-normal distribution. For this reason I use log realized volatility in my analysis.



**Figure 1:** Density plot – Realized Volatility

**Figure 2:** Density plot – Log Realized Volatility

#### IV. Hypothesis Development and Research Design

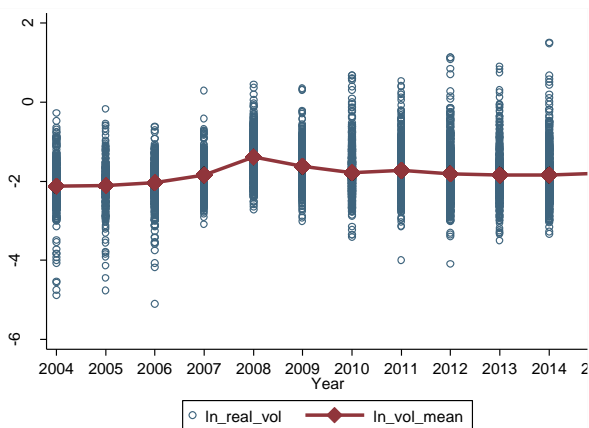
I divide my empirical analysis into four stages. First I check if the fundamental variables above described are related with future volatility. Second, based on the literature discussed in Section II.2 I find the traditional volatility model that best fits my data. Third, I test the significance of each financial variable incremental to the model estimated. Finally I design an investment strategy to try to ‘arbitrage’ on volatility. In this chapter, I describe the methodology and assumptions followed in each stage of my analysis. According to what was already described, I can formulate my first hypothesis as follows:

*Hypothesis 1: Financial information of a firm is related with the future volatility of that firm’s stock returns. Volatility increases with leverage, cash flow volatility, personnel expenses, book-to-market ratio and earnings opacity and decreases with size and return on assets.*

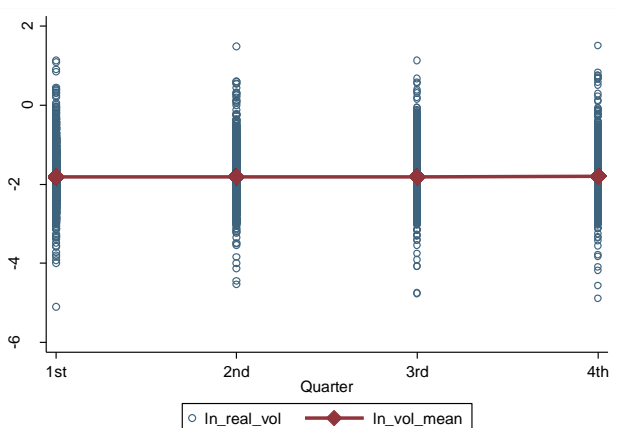
I test this hypothesis by estimating equation (5) systematically for each of the eight variables that are being considered and testing the significance of  $\beta_1$

$$\sigma_{i,t+\tau}^{RV} = \beta_0 + \beta_1 V\_FSV_{i,t}^j + \beta_2 Year + \varepsilon_{i,t} \quad (5)$$

$V\_FSV_{i,t}^j$  ( $j=1, \dots, 8$ ), is an indicator variable that equals one if the level of the accounting based variable  $j$  for firm  $i$  is above its industry median and equals zero otherwise. For instance,  $V\_SIZE$  is one if firm  $i$  in quarter  $t$  was larger than the median of its respective industry according to the computations made in section III. The estimation of equation (5) uses fixed effects and slightly differs from the estimation of equation (3) presented in section II. While Sridharan (2015) follows a two-way industry and quarter clustered standard errors approach I follow Petersen (2009), who argues that, in panels with more firms than years, the best approach is to absorb time effect by the inclusion of a dummy variable for years and then cluster by firm. I found no evidence of being necessary to control in two time dimensions, both by year and by quarter. From Figures 3 and 4 one realizes that the dependent variable presents some heterogeneity by year, but almost none by quarter. Moreover, to include a control for industry would be redundant, since the construction of the dependent variables already considers the relationship between a firm and its respective industry. Nevertheless, non-tabulated results show that there are no qualitative differences if the estimation is made clustering in two dimensions by industry and quarter, industry and country, or country and quarter.



**Figure 3:** Year Effects: Heterogeneity across years



**Figure 4:** Quarter Effects: Heterogeneity across quarters

To find the volatility model that best fits my data I first have to convert the panel dataset I am using into a time series. Since Cermeno and Grier (2001) there have been

some experiences trying to adapt ARCH models to a panel data context. However, I chose not to follow this approach for two reasons. First, Stata does not have an installed routine for this adaptation and an accurate replication of the process introduced by Cermeno and Grier (2001) requires substantial Stata programming skills. Second, this extension relies on the assumption that the disturbances in the model are cross-sectionally independent. This would not probably hold in my sample, since many firms belong to the same industries and countries. An alternative way to address this problem would then be to use the multivariate GARCH model, introduced by Bollerslev, Engle & Wooldridge (1988) in which the conditional covariance matrix H at time t (for GARCH(1,1) specification) is given as

$$vech(H_t) = C + A vech(\varepsilon_{t-1} \varepsilon_{t-1}') + B vech(H_{t-1})$$

However, being heavily parameterized, multivariate GARCH is tractable only for a small number of series. For example, a sample composed by only three firms requires an estimation of 78 coefficients. Even with the simplification assumptions of Bollerslev, Engle & Wooldridge, who assume the matrices A and B are diagonal, the number of coefficients to be estimated is still very large (18). A sample as the one I am using would require a not reasonable number of coefficients to be estimated. Moreover, the main applications of multivariate GARCH are conditional CAPM, futures hedging and volatility spillovers modeling, which are not the main focus of my research. Therefore, I aggregate the firms of my sample into an index, weighted by the market capitalization of each firm in the previous day of the earnings announcement.

To choose the mean equation for the volatility models of the index, I considered the Autoregressive Moving Average model (ARMA)

$$Y_t = c + \sum_{i=1}^m \varphi_i Y_{t-i} + \sum_{i=1}^n \varepsilon_{t-i} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (6)$$

and estimated it for different values of  $m$  and  $n$  with and without the constant term  $c$ . I then compare the Bayesian Information Criterion of each estimation to decide which ARMA( $m,n$ ) to use on the volatility models.<sup>2</sup>

Some of the most complex ARCH family models, such as CGARCH, GARCH-M, or AGARCH, require a considerable amount of data to be estimated. Because my sample is composed by quarterly observations and the time series is relatively small, only some of the most basic specifications were considered.

For (6) I considered the specifications ARCH( $q$ ) where  $\sigma_t^2 = \omega + \alpha_i \sum_{i=1}^q u_{t-i}^2$ , GARCH( $q,p$ ), where  $\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \beta_j \sum_{j=1}^p \sigma_{t-j}^2$ , EGARCH(1) where

$$\ln(\sigma_t^2) = \omega + \alpha \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right] + \gamma \frac{u_{t-1}^2}{\sqrt{\sigma_{t-1}^2}} + \beta \ln(\sigma_{t-1}^2)$$

with  $\gamma$  being a component for leverage

effects and PGARCH(1,1,1) where  $\sigma_t = \omega + \sum_{i=1}^q \alpha(|u_{t-1}| + \gamma u_{t-1}) + \sum_{i=1}^q \beta \sigma_{t-j}$ . I then compared, the Bayesian Information Criterion of each estimation to decide on the volatility model most suitable for my data. After that, I tested a second hypothesis, formulated as follows:

*Hypothesis 2: For stocks without options, the estimation of realized volatility using an ARCH family model can be improved with the inclusion of variables based on accounting information.*

I did so by adding the financial variables that showed significance when testing *Hypothesis 1* and analyzing the new model. All the referred estimations were done considering only the first 30 observations of the index and using robust estimators. The remaining were left for out-of-sample tests.

---

<sup>2</sup> This is a common methodology to model returns. See, for instance, Gökbulutand and Pekkaya (2014).

## V. Results

According to the methodology described above, this section presents and analyzes the main results of my research. All estimations were obtained using Stata13.

### V.1 Hypothesis 1

The estimation results of equation 5 that tests *Hypothesis 1* is summarized in table 3.

	I	II	III	IV	V	VI	VII	VIII
<b>V_TA</b>	-0.032 (-0.896)							
<b>V_PEA</b>		-0.012 (-0.355)						
<b>V_SGI</b>			0.003 (0.246)					
<b>V_σCF</b>				0.042 (-1.649)				
<b>V_Opacity</b>					-0.021 (-1.197)			
<b>V_LVG</b>						0.067** (2.124)		
<b>V_ROA</b>							-0.145*** (-6.208)	
<b>V_BTM</b>								0.122*** (4.241)
<b>Obs</b>	5260	2555	3075	5076	3799	5260	5528	4604
<b>Adj. R-square</b>	0.120	0.118	0.204	0.062	0.139	0.106	0.109	0.142

**Table 3:** Regression of Log Realized Volatility on accounting-based fundamental variables – Summary of results. t-statistics in parenthesis below coefficients. \*\*\*, \*\* and \* indicate statistical significance at 1, 5 and 10 percent, respectively.

From the table, *Size*, *Personal Expenses to Assets*, *Sales Growth*, *Cash Flow Volatility* and *Earnings Opacity* show no significant statistical correlation with the dependent variable. However, *Leverage* and *Book-to-Market ratio* are positively related to volatility and *Return on Assets* is negatively related to it. These three variables present coefficients significantly different from zero – *Leverage* at 5% significance level and the others at 1% significance level – and relate to volatility in the way predicted in Section II.1. The interpretation of the coefficients must consider the fact that the dependent variable is a logarithm. Therefore, the coefficient 0.067 on *Leverage*, for instance, indicates that firms with a level of leverage above their industry’s median, have on average 6.7% higher

returns volatility than firms below the industry median. Comparing these results with those of Sridharan (2015), a substantial difference is easily noted. Despite still positive to support that fundamental variables are related with realized volatility (*Hypothesis 1*), my results are not so good, considering either the number of relevant variables or the adjusted R-Square of the estimations. It is a fact that a much lower number of observations were used. However, that number should be enough to produce efficient estimates. The methodology used was virtually the same and the variables were constructed in a similar way. I am, therefore, lead to conclude that differences in the data are responsible for differences in results. It is important to consider that I used a more recent dataset than Sridharan, with firms of a different geographic location and with different specific features.

## V.2 Mean Equation of the Volatility Models

The Bayesian Information Criterion resulting from the estimation of ARMA(q,p) models with  $q=\{0,1,2\}$  and  $p=\{0,1,2\}$  is summarized in tables 4 and 5.

ar	ma	0	1	2
0		17,4473	15,8050	<b>14,0601</b>
1		17,7996	16,5616	16,5167
2		19,9178	19,8821	18,6232

**Table 4:** Bayesian Information Criterion on ARMA(q,p) estimations. With constant term.

ar	ma	0	1	2
0		14,3062	<b>13,4287</b>	16,0669
1		14,8613	15,8828	19,2402
2		17,1796	19,2393	18,4651

**Table 5:** Bayesian Information Criterion on ARMA(q,p) estimations. Without constant term.

From the tables above it is possible to observe that removing the constant term usually improves the results of the estimation. Moreover, the best fitted model is a ARMA(0,1) with no constant. Performing a Portmanteau test for white noise on the chosen specification and using 12 lags (value chosen by default), a Q statistic of 6.6983 was obtained. Comparing it to a  $\chi^2$  distribution with 12 degrees of freedom, one does not reject the null hypothesis of no autocorrelation in the residuals, indicating that the residuals apparently behave as white noise, as is desired.

### V.3 Volatility Models

Using the ARMA(0,1) specification for the mean equation, different volatility models were estimated. The results are summarized in Table 6.

	<b>ARCH (1)</b>	ARCH (2)	GARCH (1, 1)	GARCH (1, 2)	EGARCH (1, 1)	PGARCH (1, 1, 1)
$\theta$ (ARMA)	<b>0.0818</b> <b>(0.2130)</b>	0.1176 (0.4120)	0.1154 (0.3910)	0.0429 (0.8100)	0.0426 (0.814)	0.1591 (0.3140)
$\omega$ (constant)	<b>0.0095</b> <b>(0.0030)</b>	0.0099 (0.0050)	0.0100 (0.0060)	0.0098 (0.0070)	-7.1050 (0.0000)	0.0000 (0.8990)
$\alpha_1$ (ARCH)	<b>1.6109</b> <b>(0.1450)</b>	1.5763 (0.1290)	1.5789 (0.1310)	1.7735 (0.1320)	0.3522 (0.1450)	8.7166 (0.6710)
$\alpha_2$ (ARCH)		-0.0128 (0.7760)				
$\beta_1$ (GARCH)			-0.0077 (0.7570)	0.0062 (0.6070)	-0.8911 (0.0000)	-0.0000 (0.6960)
$\beta_2$ (GARCH)				-0.0118 (0.0000)		
$\Upsilon$ (Leverage)					1.1194 (0.0030)	8.0706 (0.0150)
AIC	<b>-13.5715</b>	-11.7027	-11.6968	-12.2534	-12.9162	-12.6768
BIC	<b>-9.4696</b>	-6.2336	-6.2266	-5.4169	-6.0797	-7.2077

**Table 6:** ARCH family volatility models. P-values in parenthesis below coefficients.

Analyzing the values of Akaike's Information Criteria (AIC) or the values of Bayesian Information Criterion (BIC), the conclusion is that ARCH(1) is the model that best fits the data used. The fact that the most basic specification is the one that presents a lower AIC and BIC can probably be explained by the small size of the time series, since in the absence of long high frequency data, the estimation procedure might favor the model with less parameters. Despite that, performing an ARCH-LM test, with 1 lag, on  $\frac{\hat{\varepsilon}^2}{\hat{\sigma}^2}$  results in a Q statistic of 0.7700. Comparing it to a  $\chi^2$  distribution with 1 degree of freedom, one does not reject the null hypothesis of no arch effect, indicating that the chosen model is sufficient to capture potential arch effects in the data.

## V.4 Hypothesis 2

To test the second hypothesis I added to the model the variables that showed significance in **V.1: Leverage, Return on Assets and Equity Book-to-Market Ratio**. The estimation results are summarized in **Table 8**.

	Coefficient	Std. Err.	z	P >  z
$\theta$ (ARMA)	-0.5012	0.0539	-9.29	0,000
$\omega$ (constant)	-4.6519	2.2120	-2.10	0.035
V_LVG	11.5045	5.3899	2.13	0.033
V_ROA	-23.4154	6.2730	-3.73	0,000
V_BTM	13.0198	4.2678	3.05	0.002
$\alpha_1$ (ARCH)	0.4884	0.1242	3.93	0,000

**Table 8:** ARMA(0,1) – ARCH(1) with financial variables.

The results are substantially positive and support *Hypothesis 2*. All coefficients are significant, the constant term and leverage at 5% and the remaining at 1% significance level. Most important, the financial variables influence volatility as predicted – *Leverage* and *Book to Market* have a positive sign and *Return on Assets* has a negative sign. Moreover, the inclusion of the financial variables reduced AIC from -13.5715 to -15.5264 and BIC from -9.4696 to -9.9779 and a Likelihood Ratio test on the variables added to the model shows a Q-statistic of 7.95 which, compared to a  $\chi^2$  distribution with 3 degrees of freedom, indicates overall significance of the financial variables.

## V.5 Systematic Trading Strategy

Having better forecasts than the market can be very profitable. The positive results of **Table 8** suggest the possibility for a volatility arbitrage opportunity that I will try to explore. To design a trading strategy I assume that the market uses the ARMA(0,1) ARCH(1) model to estimate volatility. This seems reasonable to assume, since it is the model that better fits historical quarterly market returns. Instead, I use the same model, but improved by the inclusion of financial variables. I also assume that the market uses returns volatility as a relevant measure of risk and that market estimations are included in

stock prices. This should be also reasonable to assume, according to Markowitz (1952) and Fama (1970). Therefore, if my model indicates a lower volatility than what it is estimated by the market I will be long on the stock. I do so, because in the current price of the stock, would be implied a higher risk profile than the one the stock actually has. Therefore, investors ‘demand’ a higher return than what would be fair. I will be out of the market in the opposite condition, to avoid incurring in excessive risk.

Using the observations included in the estimation of the models, the strategy would earn substantial returns, presenting an annual average of 9.12% net of the market for a Sharpe Ratio of 0.2. This result confirms that the model with financial variables is better, in-sample, than the simple ARMA(0,1) ARCH(1) and shows that the assumptions of my strategy hold in this market. However, an out-of-sample strategy would allow a more accurate analysis of the real forecasting power of my model. This way, I apply the same strategy only to the 16 last quarterly observations (4 years of data). The results are not so good, but still positive – an annual average of 0.92% net of the market and a Sharpe Ratio of 0.06. The strategy is better than a passive investment on the index, due to the fact that it keeps us out of the market in periods of high volatility, increasing the expected return and decreasing the standard deviation of the investment. To note that the observations to which the strategy applies were not used in the estimation process and that the variables were constructed in a way to avoid using information not available to market participants. Therefore, these results are free of any forward looking bias and reflect a true increment to the volatility model given by the financial variables.

Arbitrage on volatility is not commonly used, but the volatility forecasts can be used in investment strategies different from the one described above. A more popular way of using the volatility forecasts would be to include them in a risk parity strategy, for example.

## VI. Conclusions

My research enhances a recent link between financial accounting information and volatility modeling. I show that *Leverage* and *Equity Book-to-Market Ratio* (positively) and *Return on Assets* (negatively) are related with realized volatility and that the inclusion of those variables in an ARCH(1) model provides an increment to its explanative power. I also construct a successful investment strategy, which explores the superior forecast performance of the model with financial variables.

I show that using financial information has the advantage of improving the estimation results of a volatility model. However, a drawback of this approach should also be considered. By using information collected from financial statement figures, one is significantly reducing the number of observations. Instead of daily prices, for instance, we are forced to use quarterly observations and this can impact the choice of the correct model specification. Nevertheless, estimating volatility this way can still be useful, namely for medium/long term investments.

After converting the panel with 15,916 observations to a time series index I was left with only 46 quarterly observations and, because I only used 30 of those in my estimations, one might question the significance of my results. However, I believe they are valid even despite the sample is small. The reason is that the results were built on solid financial theories and empirical evidences discussed in section II. In fact, they are an extension of Sridharan (2015) and, therefore, are somehow supported by that author's research. Moreover, I tested *Hypothesis 1* with enough observations to support the conclusion that the variables that were later included in the ARCH(1) model should be related with volatility.

The results contribute to the volatility modeling literature in the sense that they were able to establish a relationship between a traditional ARCH volatility model and a set of financial variables. Considering that there are still many firms that do not have stock options trading on an exchange, there are several applications to a model of this nature, such as risk management, investment decisions or pricing spreads or financial instruments such as defaultable bonds.

## **VII. References**

**Alexakis, Christos; Patra, Theophano and Poshakwale, Sunil. 2010.** "Predictability of Stock Returns using Financial Statement Information: Evidence on Semi-strong Efficiency of Emerging Greek Stock Market." *Applied Financial Economics*, 20(16): 1321-1326.

**Allayannis, George and Weston, James. 2005.** "Earnings Volatility, Cash Flow volatility, and firm value." Working Paper.

**Beaver, William. 1966.** "Financial Ratios as Predictors of Failure." *Journal of Accounting Research*, 4: 71-111.

**Beaver, William; McNichols, Maureen and Rhie, Jung-Wu. 2005.** "Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy." *Review of Accounting Studies*, 10: 93-122.

**Bollerslev, Tim; Engle, Robert and Wooldridge, Jeffrey. 1988.** "A Capital Asset Pricing Model with Time Varying Conditional Covariances." *Journal of Political Economy*, 96: 116-131.

**Bonfim, Diana. 2007.** "Credit Risk Drivers: Evaluating the Contribution of Firm Level Information and of Macroeconomic Dynamics." *Banco de Portugal*, Working Paper.

- Caneghem, Tom; Campenhout, Geert and Uytbergen, Steve. 2002.** "Financial Statement Information and the Prediction of Stock Returns in a Small Capital Market: The Case of Belgium." *Brussels Economic Review*, 45(3): pp. 65-90.
- Cermeno, Rodolfo and Grier, Kevin. 2001.** "Modeling GARCH Processes in Panel Data: Theory, Simulations and Examples." *CIDE*, Working Paper.
- Chan, Louis; Lakonishok, Josef and Sougiannis, Theodore. 2001.** "The Stock Market Valuation of Research and Development Expenditures." *The Journal of Finance*, 56(6): 2431-2457.
- Chen, Nai-fu and Zhang, Feng. 1998.** "Risk and Return of Value Stocks." *The Journal of Business*, 71(4): 501-535.
- Christensen, Bent and Strunk, Charlotte. 2002.** "New Evidence on the Implies-Realized Volatility Relation." *The European Journal of Finance*, 8(2): 187-205.
- Claessen, Holger and Mitnik, Stefan. 2002.** "Forecasting Stock Market Volatility and the Informational Efficiency of the DAX-index Options Market." *Center of Financial Studies*, Working Paper No. 2002/04.
- Eklund, Trond; Larsen, Kai and Bernhardsen, Eivind. 2001.** "Model for analysing credit risk in the enterprise sector." *Norges Bank Economic Bulletin*, 3(01): 99-106.
- Engle, Robert. 1982.** "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*, 50(4): 987-1087.
- Fama, Eugene. 1970.** "Efficient capital markets: A review of theory and empirical work." *Journal of Finance*, 25(2): 427-465.
- Fama, Eugene and French, Kenneth. 1992.** "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, 47(2): 427-467.
- Fama, Eugene and French, Kenneth. 2008.** "Dissecting Anomalies." *The Journal of Finance*, 63(4): 1653-1678.

- Goodman, T.; M. Neamtiu and F. Zhang. 2013.** “Fundamental Analysis and Option Returns”, Working Paper.
- Goslim, Jonathan; Chai, Daniel and Gunasekarage, Abeyratna. 2012.** “The Usefulness of Financial Statement Information in Predicting Stock Returns: New Zealand Evidence.” *Business and Finance Journal*, 6(2): 51-70.
- Gow, Ian; Ormazabal, Gaizka and Taylor, Daniel. 2009.** “Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research.” *The Accounting Review*, 85(2): 483-512.
- Grullon, Gustavo; Lyandres, Evgeny and Zhdanov, Alexei. 2012.** “Real Options, Volatility, and Stock Returns.” *Journal of Finance*, 67: 1499-1536.
- Hui, Cho; Lo, Chi-Fai and Huang, M. 2007.** "Predictions of Default Probabilities by Models with Dynamic Leverage Ratios." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.1113726.
- Hutton, Amy; Marcus, Alan and Tehranian, Hassan. 2009.** “Opaque Financial Reports, R-Square, and Crash Risk.” *Journal of Financial Economics*, 94: 67-86.
- İlker, Gökbulutand and Mehmet, Pekkaya. 2014.** “Estimating and Forecasting Volatility of Financial Markets Using Asymmetric GARCH Models.” *International Journal of Economics and Finance*, 6(4): 23-35.
- Kaplan, Robert and Gabriel Urwitz. 1979.** "Statistical Models of Bond Ratings: A Methodological Inquiry." *Journal of Business*, 52(2): 231–261.
- Malkiel, Burton. 2003.** “The Efficient Market Hypothesis and Its Critics.” *Journal of Economic Perspectives*, 17(1): 59-82.
- Mandelbrot, Benoit. 1963.** “The Variation of Certain Speculative Prices.” *The Journal of Business*, 36(4): 394-419.

- Maricica, Moscalu and Georgeta, Vintila.** 2012. "Business Failure Risk Analysis Using Financial Ratios." *Procedia – Social and Behavioral Sciences*, 62: 728-732.
- Markowitz, Harry.** 1952. "Portfolio Selection." *The Journal of Finance*, 7(1): 77-91.
- Minton, Bernadette and Schrand, Catherine.** 1999. "The Impact of Cash Flow Volatility on Discretionary Investment and the Costs of Debt and Equity Financing." *Journal of Finance Economics*, 54(3): 423-460.
- Palepu, Krishna.** 1986. "Predicting Takeover Targets: A Methodological and Empirical Analysis." *Journal of Accounting and Economics*, 11(8): 3-35.
- Petersen, Mitchell.** 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies*, 22(1): 435-480.
- Poon, Ser-Huang and Granger, Clive.** 2003. "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature*, 41(2): 478-539.
- Roberts, SW.** 1959. "Control Chart Tests Based on Geometric Moving Averages." *Technometrics*, 1(3): 239-250.
- Sinclair, Euan.** 2013. "Volatility Forecasting Using Fundamental Information" in *Volatility Trading*, 2<sup>nd</sup> Edition. New Jersey: Wiley.
- Sridharan, Suhas.** 2015. "Volatility Forecasting Using Financial Statement Information." *The Accounting Review*, 90(5): 2079-2106.
- Sun, Bo.** 2009. "Asset Returns with Earnings Management." *International Finance Discussion Papers*, No 988.