Stock Returns, Efficiency of Beta and the Probability to Grow at an Above-Average Rate Relative to the Market: Evidence from a Logit Model

Ricardo Goulart Serra
Insper Ibmec-SP, Brazil

Luiz Paulo Lopes Fávero
University of São Paulo
FEA/USP, Brazil

Roy Martelanc
University of São Paulo
FEA/USP, Brazil

Abstract

Does the beta help to distinguish between companies that would gain an above or below-market return? Using logistic regression models, this paper aims to verify what characteristics help to determine the probability that stock prices will grow above market on a day of considerable market growth. The selected day was October 13th 2008, when S&P 500 achieved the largest growth (11.6%) since 1950. The analysis considered 461 companies listed on NYSE. The logistic analysis identified that the lagged return of 3 months and illiquidity are significant variables to determine the desired probability. This model would have classified correctly, a posteriori, 73.3%. On the other hand, the beta correctly classified not even 50% of observations.

Key words: Capital Market; Stock Return; Efficiency; Beta; Logit Model.

Introduction

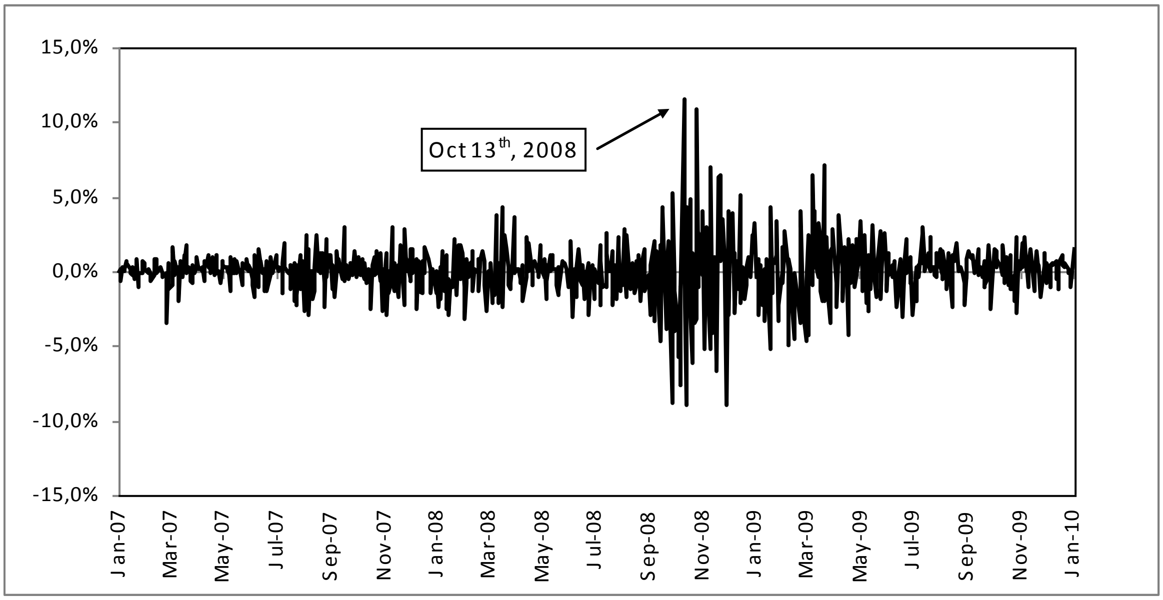
The beta is a central theme in finance and has been the focus of several studies. Some of these verify whether the beta is a good stock return estimator – alone or together with other company characteristics – or if the beta is not significant in the explanation of stock returns. Some authors believe that the beta is, to say the least, wounded, like Fama and French (1992). Others have found significance for the beta in the explanation of stock return, including Kothari, Shanken and Sloan (1995).

The motive for this study is not the search for evidence on the linear relation between return and beta, but an even more basic issue. Does the beta help to distinguish between companies with above or below-market returns? That is, on a day of a considerable market growth, would stock prices of companies with a beta over 1 grow more, and would stock prices of companies with a beta below 1 grow less than the market?

Hence, this paper aims to identify, for a day of considerable market growth, company characteristics that help to define the probability that stock prices surpass the market, with a view to checking whether the beta is part of this set of characteristics. In case it is not, how would the beta have performed in this distinction when compared with the performance of the identified set of characteristics?

October 13th 2008 was studied, which was the day with the highest returns for S&P 500 (11.6% on one single day) since January 3rd 1950. It can be observed in Graph 1 that this day is part of a critical period in the 2008 financial crisis. The sample comprised non-financial companies listed on the New York Stock Exchange (NYSE).

Graph 1: **S&P 500 Returns**



Obs.: Returns between January 3rd 2007 and January 3rd 2010. Immediately before October 13th 2008, on September 29th 2008 (-8.8%), October 7th 2008 (-5.7%) and October 9th 2008 (-7.6%), the market dropped down by more than 5.0% and, on September 30th 2008 (5.3%), it grew up by more than 5.0%. On September 13th 2008 occurred the largest growth since 1950 (11.6%).

Source: Economática® (March 23rd 2010).

A logistic regression model is proposed to check the probability that a given stock return will exceed market returns for October 13th 2008. This paper comprises four sections. While the first section presents a short literature review on the beta and what company characteristics can influence stock returns, section two describes the adopted method. The third section addresses the results, and conclusions are reserved for the final section.

1. Conceptual Review

1.1. Beta

According to Copeland, Weston and Shastri (2005), the ability to quantify risk was one of the big advances in finance theory in recent decades. This ability enhanced theoretical advances on different fronts, one of which was the development of the Capital Asset Pricing Model – CAPM (Sharpe, 1964 and Lintner, 1965), the most used method for cost of own capital calculations (Weston and Weaver, 2001).

1.2. Company Characteristics

Different authors study the company characteristics that can explain the variability of these same companies' stock returns. While Banz (1981) studied size, Bhandari (1988) observed leverage, Stattman (1980), Rosenberg, Reid and Lanstein (1985) and Chan, Hamao and Lakonishok (1991) studied the book value/market value ratio and Ball (1978) and Basu (1983) tested the earnings/price ratio. Some of these studies addressed the characteristic alone or together with other variables, including the beta.

In a compilation of earlier studies, Fama and French (1992) tested a combination of several previously studied variables: beta, size, leverage, book value/market value and earnings/price in non-financial companies. The authors concluded that, if assets are priced rationally, stock risks are multidimensional (Fama and French, 1992, p. 428). Various studies followed Fama and French' paper (1992), some contesting their results or finding different ones (Black, 1993, MacKinlay, 1995 and Kothari, Shanken and Sloan, 1995), while others supported the results for other time periods, companies (including financial institutions) or countries (Davis, 1994, Barber and Lyon, 1997, Capul, Rowley and Sharpe, 1993, Hawawini and Keim, 1997, Fama and French, 1998 and Griffin, 2002).

Other authors have studied the impact of other variables (alone or together with the previously mentioned variables): for example, Amihud and Mendelson (1986, 1991) and Amihud (2002) studied the liquidity measured by bid-ask spread or illiquidity calculated according to equation 1 and Chan, Jegadeesh and Lakonishok (1996) studied the moment and reversal.

$$\frac{\sum_{t=1}^{T} \frac{r_t}{v_t}}{T} \times 1000 \qquad (1)$$

Where r_t is the return on day t and v_t the volume in \$ on day t, calculated for time period T.

2. *Method*

The goal of this paper is to identify, for a day of considerable market growth, the company characteristics that help define the probability that stock prices will exceed market returns, so as to verify if the beta is part of this set of characteristics. Likewise, if the beta is not included, this study also intends to verify how it would have performed, for distinctive purposes, in comparison with the identified set of characteristics.

Next, the database, characteristics and the adopted model will be presented.

2.1. Database

Information on the companies listed on NYSE was taken from the information system Economática® on two different dates: April 01st 2010 and May 05th 2010. Among the 971 companies available in Economática®, 767 are non-financial. The final database contains 461 non-financial companies, which included all information needed for the analysis.

This final database was divided to constitute the analysis sample. Initially, 3 groups were formed: (a) 150 companies with higher returns, (b) 150 companies with lower returns and (c) 161 companies with intermediary returns. Then, samples (a) and (b) were each randomly subdivided into two sub-samples of 75 companies, one used in the analysis (analysis sample) and the other to validate the results (test sample).

2.2. Characteristics Used

As verified in section 1.2, different company characteristics have been used to explain expected return. Chart 1 summarizes the characteristics used in this paper.

Chart 1: **Tabulation of Characteristics Used**

Code	Characteristics
BE/ME	Book Value of Equity / Market Value of Equity
E/P	Earnings / Price
SIZE	Natural logarithm of Market Value of Equity
ROE	Earnings/Book Value of Equity
LEV	Market Value of Firm /Market Value of Equity
LR	Lagged return, in different periods: 1 week (LR_1s), 1 month (LR_1m), 2 months (LR_2m), 3 months (LR_3m) and 36 months (LR_36m)
ILLIQ	Calculated, according to Amihud (2002), considering a period of 1 year
STDV	Standard deviation of one-year returns
BETA	Beta calculated with a five-year return history, one-month return interval and market index S&P 500
BETA_SW	Beta calculated, according to Scholes and Williams (1979) method, with a one-year return history, one-day return interval and market index S&P 500

The beta was calculated considering a 5-year return history, 1-month return interval and the S&P 500 as the market index. Damodaran (1994) and Koller, Goedhart and Wessels (2005) suggest using 5 years history and monthly return interval.

As two company characteristics used in this paper were calculated with a 1-year history and daily return interval (STDV and ILLIQ), another beta was calculated with these parameters. It is known that the beta of thinly traded stocks, calculated with a 1-day return interval, can be biased downwards (Dimson, 1979) therefore the Scholes and Williams (1979) method was applied to calculate these betas.

2.3. Logistic Regression

Logistic regression is a technique developed in the 1960's to investigate the relation between metric and non-metric explanatory variables and a binary categorical dependent variable. Hence, its goal is to verify the probability that an event will occur and to identify characteristics of the elements belonging to each group determined by the categorical variable (Fávero, Belfiore, Silva and Chan, 2009).

A model is defined as logistic if the function follows the equation below:

$$f(z) = \frac{1}{1 + e^{-(z)}} \tag{2}$$

With z:

$$z = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \tag{3}$$

In which p indicates the probability that a given event of interest will occur, X represents the vector of explanatory (or independent) variables and α and β represent the model parameters. Term ln(p/(1-p)) is called logit and term (p/(1-p)) represents the odds that the event of interest will occur.

In simplified terms, function f(z) can be understood as the probability that the dependent variable equals 1, given the behavior of the explanatory variables X_1, X_2, ..., X_k. Mathematically, it can be represented as follows:

$$P(1) = f(Y = 1 \mid X_1, X_2, \ldots, X_k) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \tag{4}$$

In line with Fávero, Belfiore, Silva and Chan (2009), as α and β are unknown parameters, they need to be estimated in order to determine the probability that the event of interest will occur. In this paper, this refers to the return of a given stock being above-market on the studied day. In other words, the goal of estimating these parameters is to find a logistic function so that weighting the explanatory variables permits determining the importance of each variable for the occurrence of the event of interest, as well as calculating the probability that this event will occur.

According to the same authors, logistic regression is preferable in many situations due to the small number of premises, including the non-assumption of variance homogeneity. Moreover, different authors indicate the elimination of outliers to elaborate the logistic regression, as Tabachnick and Fidell (2006) appoint. In this study, observations with high residual values were removed from the sample for the final analysis. Therefore, Cook's distance was used, which is common to estimate the influence of a given observation in regression model (Fávero, Belfiore, Silva and Chan, 2009).

3. Results

3.1. Logistic Model

For the date under analysis, the logistic regression indicated that the variables related to the 3-month lagged return (LR_3m) and illiquidity (ILLIQ) represent good indicators to identify the probability that a stock will be part of the group of stocks with above-market returns, with a higher significance level for the former (LR_3m).

The logistic regression model is expressed in equation 5 below.

z = - 3,408 - 13,672 * LR_3m - 1.904,629 * ILLIQ (5)

As presented above, the logistic regression considered that the event of interest refers to the fact that a given company will present an above-market return. In other words, the higher the logit z, the greater the probability of achieving higher returns. Thus, company stocks with a smaller lagged return, as well as lower illiquidity, have a greater probability of presenting higher returns. The logistic regression shows to be adequate.

While the Omnibus test of the model coefficients reveals 0.000 significance, the Hosmer and Lemeshow test presents 0.751 significance. Moreover, Cox & Snell's R^2 is 0.540, Nagelkerke's R^2 is 0.719 and Pseudo R^2 is 0.559. Finally, the parameters are statistically different from zero (Wald statistics equaling 22.32 for the constant, 33.89 for the variable LR_3m and 9.11 for the variable ILLIQ). The *a posteriori* classifications are indicated in section 3.3. Five observations were removed from the sample, based on residue analysis using Cook's distance. That is why group (highest and lowest returns) sizes differ.

3.2. Beta

The beta was also used alone in order to verify if it would classify the companies well in terms of above and below-market returns.

The *a posteriori* classifications are presented in the following section.

3.3. Classification

Table 1 shows the models' *a posteriori* classification. The logistic regression managed to correctly classify 83.4% of observations in the analysis sample and 78.0% of observations in the test sample. It is also observed, for the entire sample (461 companies), that the logistic model's accuracy level (73.3%) greatly exceeded that of the betas (46.9% and 48.6% for each of the analyzed betas), which revealed to be inadequate for this classification.

The classification table of the entire sample (Table 1 (c)) indicates less symmetric accuracy, i.e. a higher accuracy level for the below-market than for the above-market returns. The market index (S&P 500) on the analyzed day rose by 11.6% while the whole sample (461 companies) median is 11.0%.

Table 1: ***A Posteriori*** Classification

Table 1 (a): Analysis Sample (145 companies)

Logistic Regression

		Expected		
		Highest Returns	Lowest Returns	**Total**
Real	Highest Returns	59 80.8%	14 19.2%	**73**
Real	Lowest Returns	10 13.9%	62 86.1%	**72**
	Total	**69**	**76**	**145**
	Accuracy	83.4%		

Table 1 (b): Test Sample (150 companies)

Logistic Regression

		Expected		
		Highest Returns	Lowest Returns	**Total**
Real	Highest Returns	57 76.0%	18 24.0%	**75**
Real	Lowest Returns	15 20.0%	60 80.0%	**75**
	Total	**72**	**78**	**150**
	Accuracy	78.0%		

Table 1 (c): Whole Sample (461 companies)

Logistic Regression

		Expected		
		Return > S&P	Return < S&P	**Total**
Real	Return > S&P	140 67.0%	69 33.0%	**209**
Real	Return < S&P	54 21.4%	198 78.6%	**252**
	Total	**194**	**267**	**461**
	Accuracy	73.3%		

Beta

		Expected		
		Return > S&P	Return < S&P	**Total**
Real	Return > S&P	106 50.7%	103 49.3%	**209**
Real	Return < S&P	142 56.3%	110 43.7%	**252**
	Total	**248**	**213**	**461**
	Accuracy	46.9%		

Beta Scholes and Williams

		Expected		
		Return > S&P	Return < S&P	**Total**
Real	Return > S&P	88 42.1%	121 57.9%	**209**
Real	Return < S&P	116 46.0%	136 54.0%	**252**
	Total	**204**	**257**	**461**
	Accuracy	48.6%		

Obs.1: Table 1 (a) presents the results based on the analysis sample, in which Highest Returns indicates the group with the highest returns and Lowest Returns that with the lowest returns.
Obs.2: Table 1 (b) presents the results based on the test sample (same terminology).
Obs.3: Table 1 (c) presents the results based on the whole sample, in which Return > S&P indicates the group with returns higher than S&P 500 and Return < S&P that with returns lower than S&P 500. Beta indicates the beta calculated with a five-year history, one-month return interval and S&P index. Beta Scholes and Williams indicates the beta calculated with the Scholes and Williams (1979) method, one-year return history and one-day return interval.

4. Conclusions

In this paper, the authors attempted to answer a quite fundamental question: on a day of considerable market growth, do companies with beta below 1 present below-market return and companies with beta above 1 present above-market return? In a broader sense, can some company characteristics, other than beta, help to determine their returns? According to the CAPM, beta should the help in this determination, and should do it alone. By studying the returns on October 13th 2008 (day of highest return in S&P 500 since 1950: 11.6%) for 461 stocks of companies listed on the New York Stock Exchange (NYSE), it was observed that the beta did not identify those company stocks that achieved above or below-market returns. A model exclusively comprising the beta correctly classified, *a posteriori*, only 46.9% of observations (48.6% for the Scholes and Williams (1979) beta). Beta would not have adequately helped to distinguish between companies which would have grown above the market and companies that would have grown below the market.

The logistic regression indicates the 3-month lagged return and illiquidity as significant variables for this distinction. The logistic model classified, *a posteriori*, 83.4% of the analysis sample, 78.0% of the test sample and 73.3% of the total sample. Companies (i) with a smaller lagged return, as well as companies (ii) with higher liquidity (lower illiquidity) have a greater probability of presenting higher returns. During the October 1987 market crash, Amihud, Mendelson and Wood (1990) documented that (i) stocks which liquidity declined more in the period from October 10th to October 19th 1987 lost more value and (ii) stocks which liquidity recovered more in the period from October 10th to October 30th 1987 enjoyed a greater recovery. The authors also documented that more liquid stocks experienced a greater price recovery. They suggest this should be interpreted as a 'flight to liquidity': "investors fearing another liquidity-related crash reallocated assets toward high-liquidity stocks" (p. 67).

Wang, Meric, Liu and Meric (2009) studied 8 crashes between 1987 and 2001, with considerable drops in the S&P 500, during a 1-day period per crash. They documented that company stocks with the highest market capitalization experience the greatest drops during the crash and the highest growth during the subsequent recovery. The authors do not hypothesize on the origin of the behavior they observe. This behavior could be explained by the overreaction hypothesis, where one determinant to the stock recovery is the return during the market decline – stocks with the most negative market decline returns experienced the largest recovery. Other overreaction explanation is the difficulty for investors to get rid of the most illiquid stocks on the day of the crash. In view of the inability to get rid of smaller companies' stocks, investors sell the biggest companies' stocks. Paradoxically, stock illiquidity would protect its price on the day of the crash. The price recovery after the crash corrects previous excesses. As the biggest companies' stocks suffered greater losses, it is natural for recovery to affect these same stocks. As the smallest companies' stock prices could not drop, there is no reason to grow during the recovery.

These explanations are in line with the present study, as a strong correlation exists between capitalization and stock liquidity. Moreover, during the week immediately before the day of the large growth – October 13th 2008, a considerable drop occurred in the market (-18.2%). The second variable to explain the high return is a negative stock return during the 3-month period before the considerable market growth event. This negative 3-month return partially incorporates the return during the week before the event, which permits to explain the subsequent growth in the same way as the overreaction hypotheses cited above.

The explanatory coefficient (r^2) between the return on the event day and (i) the return during the 3-month period before the event day is 34.3% and (ii) the return during the 1-week period before the event day is 11.4%. Thus, the week before the event contributes with 33.2% of the explanation of the return of the 3 previous month, remaining an additional 66.8% of new information in that variable. This trend reversal is consistent with the findings of Wang, Meric, Liu and Meric (2009). The study limitations include: (a) the analysis is limited to a single price growth event; (b) the analysis is limited to a 1-day period; (c) the selected day is inserted in a high market volatility period; and (d) the logistic regression was modeled to obtain the probability that a stock figures among the 50% lowest return, and not to obtain the probability that a stock present above-market returns. Future studies could analyze other dates and longer periods.

References

Amihud, Y. (2002) "Illiquidity and stock returns: cross-section and time-series effects", Journal of Financial Markets, Vol. 5, No. 1, pp. 31-56.

Amihud, Y. and Mendelson, H. (1986) "Liquidity and stock returns". Financial Analysts Journal, Vol. 42, No. 3, pp. 44-48.

Amihud, Y. and Mendelson, H. (1991) "Liquidity, asset prices and financial". Financial Analysts Journal, Vol. 47, No. 6, pp. 56-66.

Amihud, Y., Mendelson, H. and Wood, R.A.. (1990) "Liquidity and the 1987 stock market crash". Journal of Portfolio Management, Vol. 16, No. 3, pp. 65-69.

Ball, R. (1978) "Anomalies in relationships between securities' yield and yield-surrogates". Journal of Financial Economics, Vol. 6, Nos. 2&3, pp. 103-126.

Banz, R. W. (1981) "The relationship between return and market value of common stocks". Journal of Financial Economics, Vol. 9, No. 1, pp. 3-18.

Barber, B. M. and Lyon, J. D. (1997) "Firm size, book-to-market ratio, and security returns: a holdout sample of financial firms". The Journal of Finance, Vol. 52, No. 2, pp. 875-883.

Basu, S. (1983) "The relationship between earning yield, market value, and return for NYSE common stocks: further evidence". Journal of Financial Economics, Vol. 12, pp. 129-156.

Bhandari, L. C. (1988) "Debt/equity ratio and expected common stock returns: empirical evidence". Journal of Finance, Vol. 43, No. 2, pp. 507-528.

Black, F. (1993) "Beta and return". Journal of Portfolio Management, Vol. 20, No. 1, pp. 8-18.

Capaul, C., Rowley, I. and Sharpe, W. F. (1993) "International value and growth stock returns". Financial Analysis Journal, Vol. 49, No. 1, pp. 27-36.

Chan, L. K., Hamao, Y. and Lakonishok, J. (1991) "Fundamentals and stock returns in Japan". Journal of Finance, Vol. 46, No. 5, pp. 1739-1789.

Chan, L. K., Jegadeesh, N. and Lakonishok, J. (1996) "Momentum strategies". Journal of Finance, Vol. 51, No. 5, pp. 1681-1713.

Copeland, T., Weston, J. F. and Shastri, K. (2005) Financial Theory and Corporate Policy. 4th ed. Boston: Pearson.

Damodaran, A. (1994) Damodaran on Valuation: Security Analysis for Investment and Corporate Finance. New York: John Wiley & Sons.

Davis, J. L. (1994) "The cross section of realized stock returns: the pre-Compustat evidence". Journal of Finance, Vol. 49, No. 5, pp. 1579-1593.

Dimson, E. (1979) "Risk measurement when shares are subject to infrequent trading". Journal of Financial Economics, Vol. 7, No. 2, pp. 197-226.

Fama, E. F. and French, K. R. (1992) "The cross-section of expected stock returns". Journal of Finance, Vol. 47, No. 2, pp. 427-465.

Fama, E. F. and French, K. R. (1998) "Value versus growth: the international evidence". Journal of Finance, Vol. 53, No. 6, pp. 1975-1999.

Fávero, L. P., Belfiore, P., Silva, F. L. and Chan, B. L. (2009) *Análise de Dados: Modelagem Multivariada para Tomada de Decisões*. Rio de Janeiro: Elsevier.

Griffin, J. M. (2002) "Are the Fama and French factors global or country-specific?" The Review of Financial Studies, Vol. 15, No. 3, pp. 783-803.

Hawawini, G. and Keim, D. B. (1997) "The cross section of common stock returns: a review of the evidence and some new findings". *Working Paper*, University of Pennsylvania.

Koller, T., Goedhart, M. H. and Wessels, D. (2005) Valuation: Measuring and Managing the Value of Companies. 4th ed. New Jersey: John Wiley & Sons.

Kothari, S. P., Shanken, J. and Sloan, R. G. (1995) "Another look at the cross-section of expected stock returns". Journal of Finance, Vol. 50, No. 1, pp. 185-224.

Lintner, J. (1965) "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budget". Review of Economics and Statistics, Vol. 47, No. 1, pp. 13-37.

Mackinlay, A. C. (1995) "Multifactor models do not explain deviations from the CAPM". Journal of Financial Economics, Vol. 38, No. 1, pp. 3-28.

Rosemberg, B., Reid, K. and Lanstein, R. (1985) "Persuasive evidence of market inefficiency". Journal of Portfolio Management, Vol. 11, No. 3, pp. 9-16.

Scholes, M. and Williams, J. (1977) "Estimating betas from nonsynchronous data". Journal of Financial Economics, Vol. 5, No. 3, pp. 309-327.

Sharpe, W. F. (1964) "Capital asset prices: a theory of market equilibrium under conditions of risk". Journal of Finance, Vol. 19, No. 3, pp. 425-442.

Stattman, D. (1980) "Book values and stock returns". The Chicago MBA: A Journal of Selected Papers, Vol. 4, pp. 25-45.

Tabachnick, B. G. and Fidell, L. S. (2006) Using Multivariate Statistics. 5th ed. Massachusetts: Pearson Education, Inc.

Wang, J., Meric, G., Liu, Z. and Meric, I. (2009) "Stock market crashes, firm characteristics, and stock returns". Journal of Banking & Finance, Vol. 33, No. 9, pp. 1563-1574.

Weston, J. F. and Weaver, S. C. (2001) Mergers and Acquisitions. New Jersey: McGraw-Hill.