Discussion Papers In Economics And Business

Trading volume and serial correlation in stock returns: a threshold regression approach

Shoko Morimoto and Mototsugu Shintani

Discussion Paper 10-28

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Trading volume and serial correlation in stock returns: a threshold regression approach

Shoko Morimoto and Mototsugu Shintani

Discussion Paper 10-28

December 2010

Graduate School of Economics and
Osaka School of International Public Policy (OSIPP)
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Trading volume and serial correlation in stock returns: a threshold regression approach

Shoko Morimoto[*] and Mototsugu Shintani[†‡]

Abstract

We extend the analysis of Campbell *et al.* (1993) on the relationship between the first-order daily stock return autocorrelation and stock market trading volume by allowing abrupt and smooth transition structures using lagged stock returns as a transition variable. Using U.S. stock market data, we find the evidence supporting the nonlinear relationship characterized by a stronger return reversal effect on a high-volume day combined with low lagged stock returns.

JEL Classification: C22, G12
Keywords: TAR, STAR, Stock return autocorrelation, Trading volume

[*]Graduate School of Economics, Osaka University, Toyonaka, Osaka 560-0043 Japan
[†]Department of Economics, Vanderbilt University, Nashville, TN 37235 USA
[‡]Corresponding author. E-mail: mototsugu.shintani@vanderbilt.edu

1 Introduction

The first-order autocorrelation in daily stock returns tends to be lower when the aggregate stock market trading volume is higher. Using both the stock price index and individual stock price series in the U.S., Campbell *et al.* (1993, CGW) find a significantly negative effect of volume on the autoregressive coefficients of stock returns. They explain their finding using a model in which risk-averse "market makers" accommodate selling pressure from "liquidity" or "noninformational" traders in exchange for the reward of a higher expected stock return. They argue that a stock price decline on a high-volume day is more likely the result of exogenous selling pressure by noninformational traders, and will be followed by price increases on subsequent days. In contrast, a stock price decline on a low-volume day may be caused by the arrival of public information on lower future cash flows (or fundamentals) with the lower possibility of price reversals.

In this paper, we extend the empirical analysis of CGW by introducing an additional nonlinear structure where the serial correlation of stock returns does not depend only on the size of trading volume but also on the sign of lagged stock returns. To this end, we consider variants of the threshold autoregressive (TAR) model and smooth transition autoregressive (STAR) model using past stock returns as a transition variable. The former assumes abrupt transition, while the latter assumes smooth transitions between two alternative effects of volume on the serial correlation of stock returns. We use updated U.S. data on stock price index returns and trading volume, and examine whether the findings of CGW are robust to these extensions.

There are a number of reasons why we may expect that the relationship between trading volume and serial correlation in stock returns depends on the sign of lagged stock returns, which is conveniently described by TAR/STAR models. First, if both types of investors hold stock for more than one period, their behavior in the face of liquidity shock obviously depends on the past performance of the stock returns. Second, negative returns increase the risk in the following period measured by volatility because of an increased debt-to-equity ratio (leverage effect). Since risk-averse market makers demand higher expected returns for riskier assets, the sign of lagged stock returns will have an effect on the serial correlation of stock returns. Third, when market declines are larger, there is a greater likelihood that margin accounts will be liquidated. Thus, noninformational traders are expected to be more active following the negative stock return periods than after the positive stock return periods.

It should be noted that our analysis is also related to some prior work that considers the asymmetric impact of stock market shocks on stock returns. De Bondt and Thaler (1989) finds that stock market overreaction effects among losers are much stronger than among winners. Koutmos (1998), Nam *et al.* (2001) and Chiang *et al.* (2007) find that stock indexes incorporate negative shocks faster than positive shocks in many advanced nations.

2 Model

The stock return, r_t, is often assumed to follow an autoregressive model because of a partial adjustment of past price to its market fundamentals. To capture the dependence of serial correlation of stock returns on volume, CGW include the product of lagged trading volume, v_{t-1}, and the lagged stock return, r_{t-1}, as an additional regressor in the autoregression. The benchmark CGW regression takes the form of

$$r_t = \alpha + \beta r_{t-1} + \gamma v_{t-1} r_{t-1} + \varepsilon_t, \tag{1}$$

and

$$r_t = \alpha + (\sum_{i=1}^{5} \beta_i D_{it}) r_{t-1} + \gamma v_{t-1} r_{t-1} + \varepsilon_t, \tag{2}$$

where D_{it}'s are five day-of-the-week dummies, and ε_t is an error term with mean zero and a finite variance. In our analysis, we extend the CGW regression models (1) and (2) to the following TAR/STAR models,

$$r_t = \alpha + \beta r_{t-1} + \gamma_1 v_{t-1} r_{t-1} + \gamma_2 v_{t-1} r_{t-1} F(z_t) + \varepsilon_t, \tag{3}$$

and

$$r_t = \alpha + (\sum_{i=1}^{5} \beta_i D_{it}) r_{t-1} + \gamma_1 v_{t-1} r_{t-1} + \gamma_2 v_{t-1} r_{t-1} F(z_t) + \varepsilon_t, \tag{4}$$

where $F(z_t)$ is a transition function, which takes a value between 0 and 1 depending on the transition variable z_t. Here, the transition variable, z_t, represents the past performance of the stock returns, e.g., the moving averages of lagged stock returns. Our model reduces to the benchmark CGW regression when $\gamma_2 = 0$.

If the relationship between the serial correlation of stock returns and trading volume is determined only by the sign of past stock market returns, an abrupt transition can be introduced by employing a transition function of the form, $F(z_t) = \mathbf{1}[z_t > 0]$, where $\mathbf{1}[A]$ is an indicator function that takes a value 1 if A is true and a value 0, otherwise. In such a case, the coefficient on the product of volume and the stock return, γ_1, represents the dependence of the first-order return autocorrelation on the trading volume for the case with negative past stock market returns. With positive past stock market returns, the coefficient on the product is represented by $\gamma_1 + \gamma_2$.

We can further extend the model to allow for the smooth transition by employing the logistic transition function, $F(z_t) = [1 + \exp(-\delta z_t)]^{-1}$, where $\delta(> 0)$ is a scale parameter that controls the rate of the transition. Note that the logistic transition function nests the indicator function since the former approaches the latter as $\delta \to \infty$. However, for simplicity, we simply refer to the model with the indicator transition function as the TAR model and the one with the logistic transition function as the STAR model.

3 Data

For the stock return series, we use daily log returns defined as $r_t = 100 \times \log(P_t/P_{t-1})$ where P_t is a value-weighted index of stocks traded on the New York Stock Exchange and American Stock Exchange (NYSE/ASE), from the Center for Research in Security Prices (CRSP) at the University of Chicago. Our data covers the period from 7/1/1963 to 12/31/2009. In addition to the full sample period, we consider a shorter sample period through 9/30/1987, which focuses on the period prior to the stock market crash of October 1987. We refer to this subsample as the CGW sample, as it roughly corresponds to the main sample period used in the analysis of CGW (sample A in their notion). For the stock market trading volumes, we use the CRSP data on a value-weighted number of shares traded daily on the NYSE/ASE. We follow CGW and use a triangular moving average of growth rates, by subtracting a one-year backward moving averages from the log trading volumes (multiplied by 100).[1]

For the transition variable, which represents the past stock market performance, we consider both a simple lagged return, $z_t = r_{t-1}$, and a lagged five-day moving average of stock returns, $z_t = \sum_{j=1}^{5} r_{t-j}/5$. Descriptive statistics of returns, their five-day moving averages, and detrended volumes are reported in Table 1. A comparison of the CGW sample in panel A and full sample in panel B shows that the variation of returns is larger with the full sample period, but no obvious difference is observed for trade volume series.

4 Results

We first estimate the benchmark CGW regression model (1) and (2) using the ordinary least squares (OLS). The estimation results of two specifications are presented in Table 2. The panel A of the table shows results based on the CGW sample. The estimate of the coefficient γ is negative and significant at the 1 percent significance level for both with and without five day-of-the-week dummies. Our result is thus consistent with the result reported in Table 2 of CGW (page 912). Inclusion of day-of-the-week dummies contributes to a somewhat higher γ estimate in absolute value, an increase in the R^2 statistic and reduction in the sum of squared residuals. The full sample results reported in panel B show the lower estimate of γ in absolute value and the smaller R^2 statistic, compared to the CGW sample results. However, it is important to note that the significantly negative estimate of γ, the main finding by CGW, remains the same even if the sample period is extended for more than twenty years.

Next, we estimate the TAR model with $F(z_t) = \mathbf{1}[z_t > 0]$ in (3) and (4). The model can again be estimated by the OLS since the additional regressor, $v_{t-1}r_{t-1}\mathbf{1}[z_t > 0]$, is observable. The results are presented in Table 3 for two alternative transition variables

[1]Thus our volume measure is the detrended log volume rather than the detrended log turnover.

and for two sample periods. When $z_t = r_{t-1}$ is used for the transition (or threshold) variable, the original coefficient on the product of volume and the stock return, γ_1, is negative and statistically significant at the 1 percent level. However, at the same time, the coefficient on the additional regressor, γ_2, turns out to be positive and statistically significant. This result suggests that the price reversals on a higher trading volume day are more evident when the past lagged returns are negative. This finding holds for both the CGW sample and full sample. When five-day moving averages, $z_t = \sum_{j=1}^{5} r_{t-j}/5$, are used for the transition variable, results are very similar to the case of $z_t = r_{t-1}$, except for the full sample case with day-of-the-week dummies where the estimate of γ_2 is positive but not statistically significant.

Finally, we estimate the STAR model with $F(z_t) = [1 + \exp(-\delta z_t)]^{-1}$ in (3) and (4). Here, for each specification, the transition variable is normalized to have a unit sample variance for the purpose of the unit-free interpretation of the scale parameter δ (see van Dijk *et al.*, 2000). The model is estimated by the nonlinear least squares (NLS) method, and the results are presented in Table 4 for two alternative transition variables and for two sample periods.[2] Since the NLS estimate corresponds to the OLS for a fixed value of δ, both the R^2 statistic and the sum of squared residuals are also reported in the table. The results for γ_1 and γ_2 in the estimated STAR model are not distinguishable from those in the estimated TAR model. While the standard error is very large, the estimate of the scale parameter δ is also large for all cases, suggesting that the shape of transition function is similar to the abrupt transition function in the TAR model.[3] However, reduction in the sum of the squared residuals suggests some improvement in terms of the model fit over the TAR model.

Unlike the case of TAR models with known threshold values (zero in our case), the linear hypothesis cannot be tested by the significance of γ_2 in the case of STAR models, because δ is not identified if γ_2 is zero. To conduct a formal specification test of our STAR

[2]Initial values are first obtained by running OLS regressions with the regressor $v_{t-1}r_{t-1}[1 + \exp(-\delta z_t)]^{-1}$ for fixed δ's from 300 equally spaced grids. The Newton method is then employed to minimize in the least square criterion for each of 300 initial values to obtain the final NLS estimate.

[3]A large standard error with a large scale parameter estimate is commonly observed in the estimation of the STAR model. See van Dijk *et al.* (2000) for the reasoning.

model, we employ the test proposed by Teräsvirta (1994). The test without day-of-the-week dummies is based on an auxiliary regression of the form,

$$\begin{aligned} r_t =& \alpha + \beta r_{t-1} + \gamma v_{t-1} r_{t-1} + \left[\phi_{11} r_{t-1} + \phi_{12} v_{t-1} r_{t-1}\right] z_t \\ &+ \left[\phi_{21} r_{t-1} + \phi_{22} v_{t-1} r_{t-1}\right] z_t^2 + \left[\phi_{31} r_{t-1} + \phi_{32} v_{t-1} r_{t-1}\right] z_t^3 + \varepsilon_t. \end{aligned} \tag{5}$$

Under the null hypothesis of linearity against the STAR model, $\phi_{ij} = 0$ holds for all $i = 1, ..., 3$ and $j = 1, 2$. For the test with day-of-the-week dummies, βr_{t-1} in (5) is replaced by $(\sum_{i=1}^{5} \beta_i D_{it}) r_{t-1}$. The results of the F test are reported in Table 5. For all cases, the linearity is significantly rejected which justifies the use of the STAR model.

5 Conclusion

We investigated the relationship between the first-order daily stock return autocorrelation and stock market trading volume using threshold and smooth transition autoregressive models. We found that, consistent with the finding by Campbell *et al.* (1993), a stock price decline on a high-volume day tends to be followed by a return reversal compared to that on a low-volume day for the extended series. Furthermore, we found statistically significant evidence of an additional nonlinear relationship where serial correlation structure also depends on past stock returns. In particular, we found stronger return reversal effect on a high-volume day if the lagged stock returns are negative.

Acknowledgements

We are grateful to Zi-Yi Guo, Keiich Hori, and seminar participants at Osaka University for helpful comments and suggestions.

References

[1] Campbell, J. Y., Grossman, S. J. and Wang, J. (1993) Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics*, 108, 905-39.

[2] Chiang, T. C., Chen, C. W. S. and So, M. K. P. (2007) Asymmetric return and volatility responses to composite news from stock markets, *Multinational Finance Journal*, 11, 179-210.

[3] De Bondt, W. F. M. and Thaler, R. H. (1989) A mean-reverting walk down Wall Street, *Journal of Economic Perspectives*, 3, 189-202.

[4] Koutmos, G. (1998) Asymmetries in the conditional mean and the conditional variance: evidence from nine stock markets, *Journal of Economics and Business*, 50, 277-290.

[5] Nam, K., Pyun, C. S. and Avard, S. L. (2001) Asymmetric reverting behavior of short-horizon stock returns: an evidence of stock market overreaction, *Journal of Banking and Finance*, 25, 807-824.

[6] Teräsvirta, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association*, 89, 208-218.

[7] van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002) Smooth transition autoregressive models: a survey of recent developments, *Econometrics Reviews,* 21, 1-47.

Table 1: Descriptive statistics

	Mean	Median	SD	Min	Max	Obs
A: 7/1/1963-9/30/1987						
Return	.04	.06	.77	-4.44	5.16	6,095
Return (5-day MA)	.04	.07	.41	-2.01	2.53	6,095
Volume	.07	.06	.22	-1.10	1.04	6,095
B: 7/1/1963-12/31/2009						
Return	.04	.07	.98	-18.80	10.90	11,707
Return (5-day MA)	.04	.07	.46	-5.57	3.64	11,707
Volume	.06	.06	.22	-1.42	1.04	11,707

Notes: Return series are log stock returns expressed in percentage. '5-day MA' represents the five-day moving averages of returns. Volume series are the log trading volumes detrended by subtracting the one-year backward moving averages. 'SD,' 'Min,' 'Max' and 'Obs' are the standard deviation, minimum, maximum and number of observations, respectively.

Table 2: Benchmark model

Day-of-the-week dummy	β	γ	R^2	SSR
A: 7/1/1963-9/30/1987				
No	.285***	$-.291^{***}$	.061	$3,384$
	(.015)	(.049)		
Yes	—	$-.340^{***}$	.075	$3,330$
		(.049)		
B: 7/1/1963-12/31/2009				
No	.106***	$-.153^{***}$	.007	$11,214$
	(.012)	(.032)		
Yes	—	$-.136^{***}$	.017	$11,096$
		(.032)		

Notes: Numbers in parentheses are standard errors. Statistically significant estimates at the 1%, 5%, and 10% levels are shown with asterisks, ***, **, and *, respectively. 'SSR' is the sum of squared residuals.

Table 3: TAR model

Transition variable (z_t)	Day-of-the-week dummy	β	γ_1	γ_2	R^2	SSR
A: 7/1/1963-9/30/1987						
r_{t-1}	No	$.280^{***}$	$-.545^{***}$	$.377^{***}$	.063	$3,374$
		(.015)	(.079)	(.092)		
r_{t-1}	Yes	—	$-.610^{***}$	$.400^{***}$	.078	$3,320$
			(.079)	(.091)		
$\sum_{i=1}^{5} r_{t-i}/5$	No	$.284^{***}$	$-.458^{***}$	$.257^{***}$	.062	$3,379$
		(.015)	(.075)	(.087)		
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	—	$-.509^{***}$	$.259^{***}$	.076	$3,325$
			(.075)	(.087)		
B: 7/1/1963-12/31/2009						
r_{t-1}	No	$.103^{***}$	$-.220^{***}$	$.164^{***}$	.008	$11,205$
		(.012)	(.039)	(.053)		
r_{t-1}	Yes	—	$-.185^{***}$	$.117^{**}$	.017	$11,092$
			(.039)	(.054)		
$\sum_{i=1}^{5} r_{t-i}/5$	No	$.104^{***}$	$-.181^{***}$	$.113^{*}$	.007	$11,210$
		(.012)	(.035)	(.058)		
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	—	$-.150^{***}$	.055	.017	$11,096$
			(.035)	(.058)		

Notes: $F(z_t) = \mathbf{1}[z_t > 0]$. Numbers in parentheses are standard errors. Statistically significant estimates at the 1%, 5%, and 10% levels are shown with asterisks, ***, **, and *, respectively. 'SSR' is the sum of squared residuals.

Table 4: STAR model

Transition variable (z_t)	Day-of-the-week dummy	β	γ_1	γ_2	δ	R^2	SSR
A: 7/1/1963-9/30/1987							
r_{t-1}	No	.280***	$-$.547***	.378***	41.34	.063	$3,370$
		(.015)	(.079)	(.092)	(1842.75)		
r_{t-1}	Yes	—	$-$.612***	.401***	42.23	.078	$3,315$
			(.079)	(.092)	(1741.65)		
$\sum_{i=1}^{5} r_{t-i}/5$	No	.284***	$-$.450***	.239***	161.26	.062	$3,374$
		(.015)	(.075)	(.087)	(1961.55)		
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	—	$-$.498***	.236***	161.26	.076	$3,321$
			(.075)	(.087)	(1972.91)		
B: 7/1/1963-12/31/2009							
r_{t-1}	No	.103***	$-$.220***	.164***	36.83	.008	$11,200$
		(.012)	(.039)	(.053)	(2956.47)		
r_{t-1}	Yes	—	$-$.185***	.117**	36.83	.017	$11,088$
			(.039)	(.054)	(4124.93)		
$\sum_{i=1}^{5} r_{t-i}/5$	No	.104***	$-$.180***	.108*	58.11	.007	$11,206$
		(.012)	(.036)	(.059)	(668.42)		
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	—	$-$.150***	.053	58.11	.017	$11,092$
			(.036)	(.059)	(1354.30)		

Notes: $F(z_t) = [1 + \exp(-\delta z_t)]^{-1}$. Numbers in parentheses are standard errors. Statistically significant estimates at the 1%, 5%, and 10% levels are shown with asterisks, ***, **, and *, respectively. 'SSR' is the sum of squared residuals.

Table 5: Linearity test

Transition variable (z_t)	Day-of-the-week dummy	F statistic	p-values
A: 7/1/1963-9/30/1987			
r_{t-1}	No	52.91	0.00
r_{t-1}	Yes	23.83	0.00
$\sum_{i=1}^{5} r_{t-i}/5$	No	54.34	0.00
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	25.40	0.00
B: 7/1/1963-12/31/2009			
r_{t-1}	No	27.00	0.00
r_{t-1}	Yes	17.60	0.00
$\sum_{i=1}^{5} r_{t-i}/5$	No	35.65	0.00
$\sum_{i=1}^{5} r_{t-i}/5$	Yes	17.73	0.00

Notes: Teräsvirta's (1994) test for linearity against the logistic STAR model.