2011

Is the Basis of the Stock Index Futures Markets Nonlinear?

Heni Puspaningrum
University of Wollongong

Yan-Xia Lin
University of Wollongong, yanxia@uow.edu.au

Chandra Gulati
University of Wollongong

Publication Details

Puspaningrum, Heni; Lin, Yan-Xia; and Gulati, Chandra, Is the Basis of the Stock Index Futures Markets Nonlinear?, Proceedings of the Fourth Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia.

Is the Basis of the Stock Index Futures Markets Nonlinear?

Abstract

A considerable amount of papers use a cost-carry model in modelling the relationship between future and spot index prices. The cost-carry model defines basis, $bt;T$, at time t and maturity date of the future contract at T as $bt;T = ft - st = r(T - t)$, where ft, st and r denote the log of future prices, the log of spot index prices and the difference between interest rate and dividend rate, respectively. Using daily data time series on future contracts of the S&P 500 index and the FTSE 100 index, as well as the price levels of the corresponding underlying cash indices over the sample period from January 1, 1988 to December 31, 1998, [1] argued that there is significant nonlinearity in the dynamics of the basis due to the existence of transaction costs or agents heterogeneity. They found that the basis follows a nonlinear stationary ESTAR (Exponential Smooth Transition Autoregressive) model. However, based on the study with the S&P 500 data series from January 1, 1998 to December 31, 2009, we conclude that there is no significant difference between a linear AR(p) model and a nonlinear STAR model in fitting the data.

Keywords

autoregressive model, smooth transition autoregressive model, unit root test, co-integration

Publication Details

Puspaningrum, Heni; Lin, Yan-Xia; and Gulati, Chandra, Is the Basis of the Stock Index Futures Markets Nonlinear?, Proceedings of the Fourth Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia.

Is the Basis of the Stock Index Futures Markets Nonlinear?

Heni Puspaningrum, Yan-Xia Lin and Chandra Gulati

University of Wollongong, Wollongong, NSW, 2500, AUSTRALIA

Abstract

A considerable amount of papers use a cost-carry model in modelling the relationship between future and spot index prices. The cost-carry model defines basis, $b_{t,T}$, at time t and maturity date of the future contract at T as $b_{t,T} = f_t - s_t = r(T - t)$, where f_t, s_t and r denote the log of future prices, the log of spot index prices and the difference between interest rate and dividend rate, respectively. Using daily data time series on future contracts of the S&P 500 index and the FTSE 100 index, as well as the price levels of the corresponding underlying cash indices over the sample period from January 1, 1988 to December 31, 1998, [1] argued that there is significant nonlinearity in the dynamics of the basis due to the existence of transaction costs or agents heterogeneity. They found that the basis follows a nonlinear stationary ESTAR (Exponential Smooth Transition Autoregressive) model. However, based on the study with the S&P 500 data series from January 1, 1998 to December 31, 2009, we conclude that there is no significant difference between a linear AR(p) model and a nonlinear STAR model in fitting the data.

Key words: autoregressive model, smooth transition autoregressive model, unit root test, cointegration

1. Introduction

[1] analysed the mean reversion of future bases of S&P 500 and FTSE 100 with daily data spanned from January 1, 1988 to December 31, 1998. They concluded that the two bases follow ESTAR (Exponential Smooth Transition Autoregressive) models.

A STAR model can be written as follow:

$$\begin{aligned} b_t &= \theta_{10} + \sum_{j=1}^{p} \theta_{1j} b_{t-j} + \\ &\left[\theta_{20} + \sum_{j=1}^{p} \theta_{2j} b_{t-j}\right] G(\theta, r, b_{t-d}) + \epsilon_t \qquad (1) \end{aligned}$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d \geq 1$ is a delay parameter; $(\theta, r) \in \{\mathcal{R}^+ \times \mathcal{R}\}$ where $\mathcal{R}$ denotes the real space $(-\infty, \infty)$ and $\mathcal{R}^+$ denotes the positive real space $(0, \infty)$. The transition function $G(\theta, r, b_{t-d})$ determines the speed of adjustment to the equilibrium r. Two simple transition functions suggested by [2] and [3] are logistic and exponential functions:

$$G(\theta, r, b_{t-d}) = \frac{1}{1 + \exp\{-\theta(b_{t-d} - r)\}} - \frac{1}{2}, \qquad (2)$$

$$G(\theta, r, b_{t-d}) = 1 - \exp\{-\theta^2 (b_{t-d} - r)^2\}. \qquad (3)$$

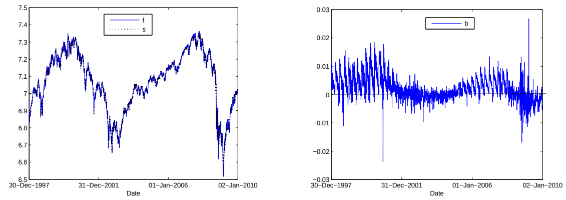
If the transition function $G(\theta, r, b_{t-d})$ is given by (2), (1) is called a logistic smooth transition autoregressive (LSTAR) model. If the transition function $G(\theta, r, b_{t-d})$ is given by (3), (1) is called an exponential smooth transition autoregressive (ESTAR) model.

[1] argued that an ESTAR model is more appropriate for modelling basis movement than a LSTAR model due to symmetric adjustment of the basis. Furthermore, there is fairly convincing evidence that distribution of the basis is symmetric, for example the evidence provided by [4] using both parametric and nonparametric tests of symmetry applied to data for the S&P 500 index. However, [1] also tested for nonlinearities arising from the LSTAR formulation, then make conclusion confirming that the ESTAR model is more appropriate for modelling basis movement than a LSTAR model.

Using current available data, we would like to know whether the basis of S&P 500 follows an ESTAR model as [1] suggested.

2. Empirical Analysis

Using daily closing prices data of future and spot index prices of the S&P 500 from January 1, 1998 to December 31, 2009, the procedures in [1] are followed. In constructing the basis, the spot price is paired up with the future contract price with the nearest maturity. Figure 1(a) shows the plots of f_t and s_t while Figure 1(b) shows the plot of b_t.



(a) (b)

Figure 1: (a) Plot of f_t and s_t; (b) Plot of b_t .

From Figure 1, the plots of f_t and s_t are almost similar indicating the basis b_t which is the difference between f_t and s_t is not large. During the data period, there are 2 major financial crises. The first is in 1999-2002 due to the South American economic crisis in Argentina, Brazil and Uruguay [1] as well as the Dot-com bubble crisis [2]. The second is the financial crisis of 2007 to the present triggered by the US subprime mortgage crisis[3]. Both financial crises are reflected in the fall of the future and spot index prices. The crises are also reflected in the basis where the basis tends to has negative value during the crisis periods.

2.1. *Preliminary Statistics*

Table 1 shows some summary statistics for the future prices f_t, the spot index prices s_t, the basis b_t and the demeaned basis mb_t. The PACF plots (not shown in this paper) suggest that both the future and spot index prices show significant spikes at the first 3 lags, but the first spikes is very strong. The PACF plot of the basis displays a slower decay of the PACF with significant spikes at the first five lags, lag 7 , lag 10 and lag 19. Box-Ljung autocorrelation tests statistics for AR(3) residuals using 20 lags for f_t and s_t are 30.0231 [0.0694] and 29.3014 [0.0820], respectively, where the figures in the parentheses are the p-values. Thus, we can accept the null hypothesis of no autocorrelation in residuals for f_t and s_t using AR(3) models and then use $p = 3$ for unit root tests. Box-Ljung autocorrelation tests statistics using 20 lags on mb_t for AR(5), AR(7) AR(10) and AR(19) residuals are 58.0468 [0.0000], 41.2758 [0.0034], 27.1426 [0.1313], 2.7141 [1.0000], respectively. From these results, $p = 10$ is enough to make the residuals become unautocorrelated for mb_t.

The standard augmented Dickey-Fuller (ADF) unit root tests reported in Table 2 shows that both f_t and s_t are I(1) while mb_t is I(0). Using other lags do not change the conclusions.

Table 1: Summary Statistics

	f_t	s_t	b_t	mb_t
Minimum	6.5160	6.5169	-0.0237	-0.0260
Maximum	7.3627	7.3557	0.0267	0.0244
Mean	7.0711	7.0688	0.0023	-6.60E-06
Variance	0.0279	0.0272	1.74E-05	1.74E-05

Notes: f_t, s_t, b_t and mb_t denote the log of the future prices, the log of the spot index prices, the basis and the demeaned basis, respectively. The demeaned basis is defined as $mb_t = b_t - \overline{b}$, where $\overline{b}$ is the mean of the basis so that the mean of mb_t is zero.

Table 2: Unit Root Tests for S&P 500

Future prices	$f_t^{(c)}$	Lags	Δf_t	Lags
	-2.1139	2	-44.072**	1
Spot Index	$s_t^{(c)}$	Lags	Δs_t	Lags
prices	-2.1255	2	-43.824**	1
Demeaned	mb_t	Lags	Δmb_t	Lags
basis	-7.1598**	9	-25.312**	8

Notes: The statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process; (c) superscripts indicate that a constant was included in the augmented Dickey-Fuller regression; "Lags " in the fourth column are the lags used in the augmented Dickey-Fuller regression for f_t, s_t, and mb_t while the last column denotes the lags used for Δf_t, Δs_t, and Δb_t; * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in [5].

Johansen cointegration test (see [6], [7]) is employed and reported in Table 3. The test uses a maximum likelihood procedure in a vector autoregression comprising f_t and s_t, with a lag length of 2 and an unrestricted constant term [4]. Both Johansen likelihood ratio (LR) test statistics clearly suggest that there are 2 cointegrating relationships between f_t and s_t, but the first cointegrating relationship shows much more significant than the second one. Financial theory based on the cost-carry model suggests that the cointegrating parameter equals unity, i.e. in this case means one unit price of f_t is cointegrated with one unit price of s_t or the first cointegrating vector β in the Johansen cointegration test results is [1,-1]. However, from Table 3, the first cointegrating vector, i.e. the first row of β', in the Johansen cointegration test results for the data is [1,-1.0124]. Imposing the restriction of the first row of β' equals [1,-1] produces the $\mathcal{X}^2$ statistics reported in the last row of Table 3. It concludes that there is not enough support for the restriction. It is quite different conclusion compared to [1] where they concluded that there is only exist one cointegrating relationship with the restriction of $[1, -1]$ can be supported.

[1] See "South American economic crisis of 2002" in http://en.wikipedia.org/wiki/South_American_economic_crisis_of_2002. Retrieved on 18/11/2010.

[2] See "Dot-com bubble" in http://en.wikipedia.org/wiki/Dot-com_bubble. Retrieved on 18/11/2010.

[3] See "Causes of the Financial Crisis of 2007-2010" in http://en.wikipedia.org/wiki/Causes_of_the_financial_crisis_of_2007-2010. Retrieved on 18/11/2010.

[4] We use a lag length of 2 because $p = 3$ is the common lag for f_t and s_t, so that in the vector autoregression, the lag length is $p-1 = 2$.

Table 3: Johansen Maximum Likelihood Cointegration Results for S&P 500

H_0	H_1	LR
	Maximum Eigenvalue LR Test	
$r = 0$	$r = 1$	296.1**
$r \leq 1$	$r = 2$	5.105*
	Trace LR Test	
$r = 0$	$r \geq 1$	301.2**
$r \leq 1$	$r = 2$	5.105*
Eigenvalues	Standardized β' eigenvectors	
	f_t	s_t
0.0902856	1.0000	-1.0124
0.00163014	0.37218	1.0000
LR-test restriction = $\mathcal{X}^2(1)$		83.801 [0.0000] **

Table 4: Linearity tests on the demeaned basis mb_t

RESET Test			11.8 [0.00]**
Lags Used			10
d	LM^G	LM^3	LM^E
p =7			
1	15.6 [0.00]**	7.1 [0.00]**	19.6 [0.00]**
2	15.6 [0.00]**	7.1 [0.00]**	19.6 [0.00]**
p =10			
1	11.0 [0.00]**	4.6 [0.00]**	14.1 [0.00]**
2	11.0 [0.00]**	4.6 [0.00]**	14.1 [0.00]**

Notes: RESET test statistics are computed considering a linear AR(p) regression with 10 lags without a constant as the constant is not significant at 5% significant level against an alternative model with a quadratic and a cubic term. The F-statistics forms are used for the RESET test, LM^G , LM^3 and LM^E and the values in parentheses are the p-values. * and ** superscripts indicate significance at 5% and 1%, respectively.

2.2. Linearity Tests

Table 4 reports linearity tests results. The first linearity test employed is a RESET test ([8]) of the null hypothesis of linearity of the residuals from an AR(10) for mb_t against the alternative hypothesis of general model misspecification involving a higher-order polynomial to represent a different functional form. Under the null hypothesis, the statistics is distributed as $\mathcal{X}^2(q)$ with q is equal to the number of higher-order terms in alternative model. Table 4 reports the result from executing RESET test statistics where the alternative model with a quadratic and a cubic terms are included. The null hypothesis is very strongly rejected considered with the p-value of virtually zero, suggesting that a linear AR(10) process for mb_t is misspecified.

The second linearity tests are based on [3]. The tests can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. The artificial regression of (1) is estimated as follow:

$$\begin{aligned} b_t &= \theta_{00} + \sum_{j=1}^{p} \Big(\phi_{0j} b_{t-j} + \phi_{1j} b_{t-j} b_{t-d} + \phi_{2j} b_{t-j} b_{t-d}^2 \\ &\quad + \phi_{3j} b_{t-j} b_{t-d}^3 \Big) + \phi_4 b_{t-d}^2 + \phi_5 b_{t-d}^3 + errors \quad (4) \end{aligned}$$

where ϕ_4 and ϕ_5 become zero if $d \leq p$. Keeping the delay parameter d fixed, testing the null hypothesis

$$H_0 : \phi_{1j} = \phi_{2j} = \phi_{3j} = \phi_4 = \phi_5 = 0, \quad \forall j \in \{1, \cdots, p\}$$

against its complement is a general test (LM^G) of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression(i.e., $\phi_{3j} = \phi_5 = 0$ if the true model is an ESTAR model, but $\phi_{3j} \neq \phi_5 \neq 0$ if the true model is an LSTAR), thus, testing the null hypothesis that

$$H_0 : \phi_{3j} = \phi_5 = 0, \quad \forall j \in \{1, \cdots, p\}$$

provides a test (LM^3) of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if the restrictions $\phi_{3j} = \phi_5 = 0$ cannot be rejected at the chosen significance level, then a more powerful test (LM^E) for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_0 : \phi_{1j} = \phi_{2j} = \phi_4 = 0 | \phi_{3j} = \phi_5 = 0, \quad \forall j \in \{1, \cdots, p\}.$$

A lag length of 7 and 10 are considered for executing the linearity tests for mb_t using the artificial regression in (4). Table 4 shows values of the test statistics LM^G, LM^3 and LM^E. The delay parameter $d \in \{1, 2, 3, 4, 5\}$ are considered [5]. However, the test statistics LM^G, LM^3 and LM^E show that different values of d do not affect the results. From Table 4, the p-values from LM^G, LM^3 and LM^E statistics are virtually zero for both $p = 7$ and $p = 10$. From the LM^G statistics, we can conclude that linearity is strongly rejected. From the LM^3 and LM^E statistics, we can conclude that a LSTAR model is much more strongly supported than an ESTAR model. It is quite different conclusion compared to [1] where they concluded the opposite one, i.e. an ESTAR model is more favoured than a LSTAR model. From Table 4, the LM^G, LM^3 and LM^E statistics for $p = 7$ are higher than those for $p = 10$. Therefore, we chose a LSTAR model with $p = 7$ and $d = 1$ for model estimation.

We also try for other lags such as $p - 1 = 6, 9, 18$, but they do not change the conclusions.

[5] We only report for $d = 1$ and $d = 2$ as the values are the same for other d

Table 5: Estimation results for the demeaned basis mb_t

	LSTAR(7)	AR(7)
$\hat{\theta}_{11}(= -\hat{\theta}_{21})$	0.3642 (0.0177)	0.3672 (0.0179)
$\hat{\theta}_{12}(= -\hat{\theta}_{22})$	0.1885 (0.0189)	0.1864 (0.0190)
$\hat{\theta}_{13}(= -\hat{\theta}_{23})$	0.0856 (0.0193)	0.0902 (0.0193)
$\hat{\theta}_{14}(= -\hat{\theta}_{24})$	0.1086 (0.0192)	0.1095 (0.0192)
$\hat{\theta}_{15}(= -\hat{\theta}_{25})$	0.0556 (0.0193)	0.0604 (0.0193)
$\hat{\theta}_{16}(= -\hat{\theta}_{26})$	0.0084 (0.0189)	0.0131 (0.0190)
$\hat{\theta}_{17}(= -\hat{\theta}_{27})$	0.0713 (0.0176)	0.0708 (0.0178)
θ	-26.0070 (8.7306)	
SSE	0.0214	0.0214
LR	0.000 [1.0000]	0.000 [1.0000]
SW	0.8912 [0.0000]**	0.8922 [0.0000]**
BL (20)	40.2877 [0.0046]**	41.3718 [0.0033]**

Notes: Figures in parentheses beside coefficient estimates denote the estimated standard errors. SSE is sum square error; LR is a likelihood ratio statistics for parameter restrictions; SW is a Shapiro-Wilk normality test for residuals; BL is a Box-Ljung autocorrelation test for residuals using 20 lags; the figures in parentheses denote the p-values.

2.3. Estimation Results

Table 5 reports comparison of model estimation results for a nonlinear LSTAR model with $p = 7$ and $d = 1$ and a linear AR(7) model. The nonlinear LSTAR model estimation uses a nonlinear least squares method in the form of (1) and (2) for mb_t. As the mean of mb_t is zero, theoretically, $\theta_{10} = \theta_{20} = r = 0$. Further restriction of $\theta_{2j} = -\theta_{1j}$ for $j = 1, \cdots, 7$ produces the likelihood ratio statistics, LR, in Table 5 concluding that the restrictions can not be rejected at the conventional 5% significance level. A linear AR(7) is also estimated as a comparison. The LR statistics comparing the LSTAR model and the AR(7) model concludes that there is no significant different between the two models. Furthermore, the parameter estimates of θ_{1j}, $j = 1, \cdots, 7$, for the two models are quite similar. Other statistics such as Shapiro-Wilk normality test Box-Ljung autocorrelation test for residuals are also similar for the two models. [1] did not make model comparison and they concluded that a nonlinear ESTAR model quite fits with the data they have.

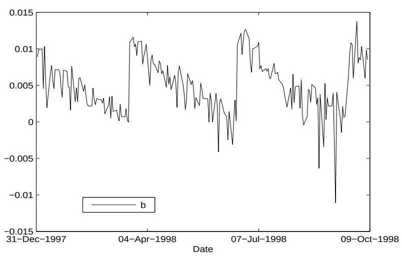


Figure 2: Plot of b_t from January 1, 1998 to October 19, 1998.

3. Conclusions

Using current available data, from January 1, 1998 to December 31, 2009, we examine the basis of S&P 500 following procedures in [1]. Even though we can conclude that there is possibility nonlinearity in the basis, there is no significant different between a nonlinear LSTAR model and a linear autoregressive model in fitting the data. It is a different conclusion compared to [1] concluding that a nonlinear ESTAR model quite fits with the data they have.

Our data has two major financial crises while the data used by [1] does not have a major financial crisis. This different data characteristic may lead to different conclusions.

We also have a concern in the way the basis is constructed. By pairing up the spot price with the future contract with the nearest maturity, it may produce artificial jumps at the time of maturity. The longer the time to maturity, the higher the difference between the future price and the spot price. For example for S&P 500, it has 4 maturity times during a year which are the third Friday in March, June, September and December. We find that at that times, there are jumps in the basis. Figure 2 shows the plot of b_t from January 1, 1998 to October 19, 1998 with jumps on the third Friday in March, June, September 1998. [1] did not discuss this issue. [9] argued that it may create volatility and bias in the parameter estimates. Therefore, the next step of this research will examine the cointegration of f_t and s_t with a time trend for each future contract.

References

[1] M. Monoyios, L. Sarno, Mean reversion in stock index futures markets: A nonlinear analysis, The Journal of Futures Markets 22 (4) (2002) 285–314.
[2] C. Granger, T. Terasvirta, Modelling Nonlinear Economic Relationship, Oxford University Press, Oxford, 1993.
[3] T. Terasvirta, Specification, estimation and evaluation of smooth transition autoregressive models, Journal of the American Statistical Association 89 (1994) 208–218.
[4] G. Dwyer, P. Locke, W. Yu, Index arbitrage and nonlinear dynamics between the S&P 500 future and cash, Review of Financial Studies 9 (1996) 301–332.
[5] W. Fuller, Introduction to Statistical Time Series, John Wiley, New York, 1976.
[6] S. Johansen, Statistical analysis of cointegrating vectors, Journal of Economic Dynamics and Control 12 (2/3) (1988) 231–254.
[7] S. Johansen, Estimation and hypothesis testing of cointegartion vectors in gaussian vector autoregressive models, Ecnometrica 59 (6) (1991) 1551–1580.
[8] J. Ramsey, Tests for specification errors in classical least-squares regression analysis, Journal of the Royal Statistical Analysis, Series B, 13 (1969) 350–371.
[9] C. Ma, J. Mercer, M. Walker, Rolling over futures contracts: A note, The Journal of Futures Markets 12 (1992) 203–217.