CHAPTER 4

RESEARCH METHODOLOGY

This thesis derives a testable implication from the rational speculative bubble model, introduces a test for the bubbles based on the statistical theory of duration dependence, and then provides evidence of nonrandom behavior in monthly real returns that is consistent with the presence of the bubbles. **The implication suggests that the probability that a run (sequence of observations of the same sign) of positive abnormal returns ends should decline with the length of the run (positive duration dependence or negative hazard function)**. Duration dependence tests address nonlinearity by allowing the parameters (probability of ending a run) to vary depending on the length of the run and on whether the run is of positive or negative abnormal returns. Additionally, duration dependence is more unique to bubbles than attributed such as autocorrelation, skewness, or kurtosis. For example, time-varying risk premiums (i.e., **Fama and French (1988)**) and fads (i.e.,**Poterba and Summers (1988)**) could both induce autocorrelation. Skewness could result from asymmetric fundamental news and leptokurtosis could be a consequence of the batch arrival of information (i.e., **Tauchen and Pitts (1983)**).

4.1) Estimation Procedure

To investigate an evidence of bubble, the research methodology need to be separated into 2 major steps:

Step 1 : To identify the positive and negative abnormal return

Step 2 : To test hypothesis of duration dependence for runs of positive and negative abnormal returns

Step 1 : To identify the positive and negative abnormal returns

To perform the duration dependence test, signs of abnormal returns are needed to divide observations into two states (positive and negative sign). In the base case that follows, the abnormal returns are defined as the residuals from the following regression,

$$(15) \quad R_t^{vw} = \beta_0 + \beta_1 (TERM)_{t-1} + \beta_2 (D/P)_{t-1} + \beta_3 \left(R_{t-1}^{vw}\right) + \beta_4 \left(R_{t-2}^{vw}\right) + \beta_5 \left(R_{t-3}^{vw}\right) + \in_t^{vw}$$

where R^{vw}_t is the real continuously compounded monthly returns on the value-weighted portfolios of SET Index.

TERM is the different between the maximum lending rate and the average lending rate

D/P is the value-weighted SET portfolio's dividend yield calculated by dividing the sum of the prior 12 monthly dividends by the current price.

$\in^{vw}_t$ is the abnormal returns.

After getting the residuals from the above regression, we have to count the number of positive and negative residuals and use them as proxy of positive or negative abnormal returns, respectively. For example, a return series of four positive abnormal returns followed by three negative, two positives and finally four negative abnormal returns is transformed into two data sets: a set for runs of positive abnormal returns with values of 4 and 2 and a set for runs of negative abnormal returns with values of 3 and 4. (see Chapter 3)

Step 2 : To test hypothesis of duration dependence for runs of positive and negative abnormal returns

In this step, the test of duration dependence will be applied for runs of positive and negative abnormal returns separately in order to find an evidence of bubbles by following the bellowed steps:

4.2) Hypothesis Testing

According to the conceptual framework, if price contain bubbles, then run of observed positive abnormal returns would exhibit duration dependence with an inverse relationship between the probability of a run ending and the length of the run. On the contrary; if price contain no bubble, then run of observe positive abnormal returns exhibit no duration dependence which means that the abnormal returns are serially independent. Therefore, the following hypothesis can be obtained:

H_0	:	No duration dependence	(No Bubble)
H_1	:	Positive duration dependence	(Contain Bubble)

4.2.1) The Sample Hazard Rate

The sample hazard rate represents the conditional probability that a run ends at *i*, given that it lasts until *i*.

$$h_i = N_i / (M_i + N_i) \qquad (16)$$

where h_i is the conditional probability that a run ends at i, given that it lasts until i

N_i is the count of runs of length i

M_i is the count of runs with a length greater than i

According to the definitions, there is a relationship between duration dependence and sample hazard rate. For example, if price contain bubbles, then run of observed positive abnormal returns will exhibit duration dependence with an inverse relationship between the probability of a run ending and the length of the run, which should be consistent with the decreasing sample hazard rate when the length of run increases. On the other hand, if price contain no bubble, then run of observe positive abnormal returns exhibit no duration dependence means the abnormal returns are serially independent, which should be consistent with the constant sample hazard rate when the length of run increases. Therefore, the sample hazard rate can be used to represent duration dependence. In this thesis we will use the sample hazard rate as the proxy of duration dependence and we can obtain the following hypothesis:

H_0 : The hazard rate should be constant.
(No bubble and no duration dependence)

H_1 : The hazard rate should decrease.
(Contain bubble and duration dependence)

4.2.2) The Hazard Function

In addition, to perform tests of duration dependence, a functional form must be chosen for the hazard function. The tests of duration dependence in this thesis are based on the logistic transformation of the log of *i*

(17) $$h_i = \frac{1}{1 + e^{-(\alpha + \beta Lni)}}$$

In term of the model, the null hypothesis of no duration dependence is that $\beta = 0$ (constant hazard rate or geometric density function). The bubble alternative suggests the probability of a positive run ending should decrease with the run length or that the value of the slope parameter, β, is negative (decreasing hazard rate) as shown by the following;

H_0 : $\beta = 0$

(Constant hazard rate and no bubble and no duration dependence)

H_1 : $\beta < 0$

(Decreasing hazard rate and contain bubble and duration dependence)

4.3) Regression Analysis

4.3.1) Regression Estimation

For ease of exposition, we write the above functional form of hazard function as

(18) $$h_i = \frac{1}{1 + e_i^{-z}}$$ where $Z_i = \alpha + \beta \text{Ln } i$.

Equation (18) represents what is known as the (cumulative) logistic distribution function.

It is easy to verify that as Z_i ranges from $-\infty$ to $+\infty$, h_i ranges between 0 and 1 and that h_i is nonlinearly related to Z_i , thus satisfying the two requirements of logit model. But it seems that in satisfying these requirements, we have created an estimation problem because h_i is nonlinear not only in Ln i but also in the α and β as can be seen clearly from (17). This means that we cannot use the familiar OLS procedure to estimate the parameters.

If h_i , the probability that a run ends at i is given by (18), then (1- h_i), the probability that a run does not end at i , is

(19) $$1 - h_i = \frac{1}{1 + e_i^{z}}$$

Therefore, we can write

(20) $$\frac{h_i}{1 - h} = \frac{1 + e_i^{z}}{1 + e_i^{-z}} = e_i^{z}$$

Now $h_i/(1- h_i)$ is simply the **odds ratio** in favor of a run ends at i – the ratio of the probability that a run ends at i to the probability that a run does not end at i. Thus, if h_i = 0.8, it means that odds are 4 to 1 in favor of the a run ends at i.

Now if we take the natural log of (20), we obtain a following result

(21) $$L_i = Ln \frac{h_i}{(1 - h_i)} = Z_i$$

$$= \alpha + \beta Lni$$

That is, L, the log of odds ratio, is not only linear in Ln i, but also linear in parameters. L is called the **logit**, and hence the name **logit model** for models like (21)

Notice these features of logit model.

1. As h goes from 0 to 1, the logit L goes from $-\infty$ to $+\infty$. That is, although the probability lies between 0 and 1, the logits are not so bounded.
2. Although L is linear in Ln i, the probabilities themselves are not.
3. The interpretation of the logit model is as follows: β, the slope, measures the changes in L for a unit change in Ln i, that is it tells how the log-odds in favor of a run ends at i changes as Ln i changes by a unit. The intercept α is the value of the log-odds in favor of a run ends at i if Ln i is zero.

Estimation of the Logit Model

For estimation purposes, we write (21) as follows:

(22) $$L_i = Ln \frac{h_i}{(1 - h_i)} = \alpha + \beta\ Lni$$

There are 5 steps in estimating the logit regression as shown by the followings:

1) For each run length (*i*), compute the estimated probability that a run ends at *i* given that it lasts until *i*.

$$h_i = \frac{N_i}{(M_i + N_i)}$$

where N_i is the count of runs of length *i*

M_i is the count of runs with a length greater than *I*

2) For each run length (*i*), obtain the logit as

$$L = Ln \frac{h_i}{(1 - h_i)}$$

3) To resolve the problem of heteroscedasticity, transform as follows:

$$\sqrt{w_i} L_i = \alpha \sqrt{w_i} + \beta \sqrt{w_i} Lni + \sqrt{w_i} \mu_i$$

$$L_i^* = \alpha \sqrt{w_i} + \beta Lni^* + V_i$$

Where the weights $w_i = (M_i + N_i) h_i(1-h_i)$; L^*_i = transformed or weighted L_i ; $Ln\ i^*$ = transformed or weighted $Ln\ i$; and v_i = transformed error term.

4) Estimate $L_i^* = \alpha\sqrt{w_i} + \beta Lni^* + V_i$ by Ordinary Least Square (OLS).

5) Establish confidence intervals and test hypothesis in the usual OLS framework.

4.3.2) Logit Regression Estimation

Under a suitable description of the data set, hazard function can be estimated as a logit regression. In the logit formulation, the independent variable is the log of the current length of the run (i) and the dependent variable is 1 (0) if the run ends (does not end) in the nest period. The Likelihood Ratio Test (LRT) of $\beta = 0$ is asymptotically distributed χ^2 with one degree of freedom.

- Independent Variable → log (i)
- Dependent Variable → 1 if the run ends in the next period
  → 0 if the run does not end in the next period