Marketing Letters 12:4, 351–364, 2001

Testing for Unit Roots in Market Shares*

PHILIP HANS FRANSES
Econometric Institute and Department of Marketing and Organization, Econometric Institute H11-34, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, email: franses@few.eur.nl

SHUBA SRINIVASAN
Department of Marketing, The Gary Anderson Graduate School of Management, University of California, Riverside, CA 92521-0203

PETER BOSWIJK
Department of Quantitative Economics, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands

Revised April 4, 2001; Accepted June 13, 2001

Abstract

A unique characteristic of marketing data sets is the logical consistency requirement in market share models that market shares are bounded by 0 and 1, and they sum to unity. To take account of this logical consistency requirement, we propose to test for unit roots in individual market share series within the context of a market share attraction (MCI) framework. Our paper offers new contributions in testing for unit roots in market shares. First, a novel feature of our paper is that we propose a new unit root testing methodology designed to deal with the logical consistency requirement in market share models within the context of a market share attraction (MCI) framework. A second novel component of our paper is that we demonstrate how one could use the Johansen (1995) test to identify unit roots. This is implemented using Eviews software. The Johansen test is a system-based test rather than a single equation test; it is more appropriate given the dependencies in the market share relationships. Finally, we demonstrate using simulations that our procedure works well and improves substantially on the univariate Dickey-Fuller procedure. Accordingly, our procedure leads to better unit root inference than the univariate Dickey-Fuller method; the latter is not that reliable when dealing with market shares. We conclude the paper with suggestions for future research.

Key words: unit roots, market shares, attraction model

1. Introduction

Over the past few years, there has been a significant increase in the research literature on the topic of long-run effects of marketing activity on market shares and sales (Dekimpe, Hanssens and Silva-Risso 1999; Bronnenberg, Mahajan and Vanhonacker 2000; Franses, Kloek and Lucas 1999; Srinivasan and Bass 2000). The interested reader is referred to Dekimpe and Hanssens (2000) for an excellent review of recent time series papers in marketing. In order to analyze the persistence of marketing efforts on market shares, a necessary first step is to examine whether the market share time series are stationary or evolving, see for example, Dekimpe and Hanssens (1995), Nijs et al. (2001) and Pauwels et al. (2001). For instance, if shares are mean-reverting, marketing actions only have a temporary effect on shares. On the other hand, if they are evolving, marketing actions could have a long-term effect on shares (Srinivasan, Popkowski Leszczyc and Bass 2000). One approach to investigating stationarity versus evolution in market shares amounts to testing for the presence of unit roots in market shares. The interested reader is referred to Phillips and Xiao (1998) and Maddala and Kim (1998) for lucid surveys on unit roots in time series.

A unique characteristic of marketing data sets is the logical consistency requirement in market share models that market shares are bounded by 0 and 1, and they sum to unity. This logical consistency requirement is likely to have an impact on the statistical distinction between stationarity and evolution in market shares, as the time series properties of market shares are closely related. Hence, it is important to take it into account if one is interested in studying all the brands in a product category. A useful modeling approach, which allows for these features, is the market share attraction model (or MCI model), see for example Cooper and Nakanishi (1988). In this paper, we propose a new unit root testing methodology for individual market shares within this MCI framework and provide empirical applications using scanner-based data sets for two frequently purchased consumer product categories. Our paper offers new contributions in testing for unit roots in market shares. First, a novel feature of our paper is that we propose a new unit root testing methodology designed to deal with a unique characteristic of marketing data sets, namely the logical consistency requirement in market share models. Second, we demonstrate how one could use the Johansen (1995) test, a system-based test that is more appropriate given the dependencies in market share relationships, to identify unit roots. This is implemented using Eviews software. Finally, we demonstrate using Monte Carlo simulations that a unique feature of our procedure is that it leads to better unit root inference and improves substantially on the univariate Dickey-Fuller method.

The structure of our paper is as follows. Section 2 first shows that it is inconvenient to test for unit roots in market shares by taking the market shares together in a multivariate model, as one then has to take account of the summation restriction, which makes subsequent analysis rather complicated. Next, it is argued that the MCI model implies that one examines the unit root properties of the differences between the logs of market shares. Additionally, this section demonstrates that when one knows the properties of these variables, there are ways to derive the properties of all individual market shares. Section 3 deals with testing for unit roots within the context of the MCI model, and it elaborates on the relevant application of the Johansen (1995) method. Specific attention will be paid to the appropriate inclusion of the intercepts in the testing model. Section 4 contains Monte Carlo simulations and empirical applications using data from two scanner product categories. The simulations suggest that univariate unit root tests for individual market shares are not very reliable. Section 5 presents the conclusions.

2. Unit Roots in Market Shares

We seek to develop a test for unit roots in individual market share series. We denote by $M_{i,t}$ the market share series for I brands in a category, where $i = 1, 2, \ldots, I$ and $t = 1, 2, \ldots, n$, such that $\sum_{i=1}^{I} M_{i,t} = 1$ for all t. The MCI model to be discussed below usually implies the analysis of the market shares when these are transformed by applying the natural logarithmic transformation, that is, log $M_{i,t}$. Notice that in other fields such as macroeconomics, it is the convention to follow the same procedure, where, for example, a unit root in, say, GNP usually implies that one examines a unit root in the logs of GNP. Hence, from now on we will consider the logs of market shares.

In assessing the long-run effects of marketing activity on shares, a necessary first step is to determine which variables have unit roots and which variables do not, see for example Banerjee, Dolado, Galbraith and Hendry (1993). A consequence of a time series variable having a unit root is that the test statistics for parameter restrictions (such as t-ratios and Wald tests) do not have asymptotic standard normal or χ^2 distributions. For further discussion, we will sometimes call a variable with a unit root a variable that is integrated of order 1 [I(1)], and a time series that is stable as a variable that is integrated of order 0 [I(0)].

Given that the market shares sum to unity, it is inappropriate to analyze the properties of the market shares in separate individual models. Indeed, as the market shares sum to unity, each of these models should then take account of this summation restriction. The same argument holds for an analysis of the logs of the market shares. In the appendix to this paper, we show that an approximate linearization results in

$$\log M_{I,t} \approx c_I - m_1 y_{1,t} - \ldots - m_{I-1} y_{I-1,t}, \tag{1}$$

where c_I is some constant, $y_{k,t}$ denotes log $M_{k,t}$ − log $M_{I,t}$, and where the m_1 to m_{I-1} denote the values of the market shares at a particular point in time. Note that these values are unlikely to be constant over time, due in part to the effects of temporary promotional activity, and hence, at any point in time they may take different values. This equation is already a prelude to a discussion below, where we will indicate that the properties of $y_{k,t}$ are informative for the properties of the individual log market shares. As it naturally holds that $m_1 + m_2 + \cdots + m_I = 1$, the equation can also be written as

$$m_1 \log M_{1,t} + m_2 \log M_{2,t} + \ldots + m_I \log M_{I,t} \approx c, \tag{2}$$

for some constant c. Hence, the properties of the logs of the market shares are also interdependent. For example, (2) indicates that it cannot occur that $I-1$ log market shares are stationary, while the I-th log market share is nonstationary. This is due to the well-known result that the sum (or difference) of two $I(0)$ series is also $I(0)$, and the sum (or difference) of an $I(1)$ and an $I(0)$ series is $I(1)$. Equation (2) also tells us that the maximum number of unit roots for the system with I log market shares is $I-1$, as it indicates the presence of a long-term stable relationship between the I log market shares. Usually, such a long-term relationship amongst potentially evolving variables is coined a cointegration relationship.

One may now think that a multivariate time series model for the $I-1$ (instead of I) logs of market shares, thereby taking account for contemporaneous correlation, would perhaps be useful. However, due to the restriction in (2), only in very specific cases can one then derive the properties of the log market share that is left out. For example, in case of three variables, suppose a unit root analysis of log $M_{1,t}$ and log $M_{2,t}$ results in a stable linear relationship. It may now be that this relationship has weights m_1 and m_2. In that case, (2) tells us that log $M_{3,t}$ should be stable. However, if this is not the case, this variable should have a unit root.

In sum, unit root analysis of the individual log market shares should be preferably performed by considering other log market shares. This section further contains a discussion of an MCI type model for the I market shares, and it illustrates that this model implies that the subsequent unit root analysis focuses on analyzing the differences between pairs of logs of market shares. In the next subsection we will argue that the results for these differences, and a few additional but straightforward computations, can lead to an inference on the unit root properties of the individual log market shares.

2.1. Unit Roots in the MCI Model

The MCI model assumes that market shares are defined in terms of I unobserved positive attraction variables $A_{i,t}$, $i = 1, 2, \ldots, I$, that is,

$$M_{i,t} = \frac{A_{i,t}}{\sum_{j=1}^{I} A_{j,t}}. \tag{3}$$

Given this definition, it is evident that each $M_{i,t}$ lies between 0 and 1, and that the summation-to-unity restriction holds. The notion that the attraction variables are unobserved is important for the subsequent discussion, as a host of various specifications are possible. These specifications all lead to different parameter restrictions concerning unit root properties, as will become clear below.

When the focus is on the univariate analysis of each of the I market shares, a potentially useful but simple specification of $A_{i,t}$ is for example

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) M_{i,t-1}^{\lambda_i}, \tag{4}$$

where the $\varepsilon_{i,t}$ variables are standard white noise variables, that is, they are uncorrelated over time, with common mean 0 and variance-covariance matrix Σ. It is possible to restrict this Σ, see for example Franses and Paap (1999), but for convenience no such restrictions are imposed here.

For parameter estimation, it is usually convenient to consider the $I-1$ variables $M_{i,t}/M_{I,t}$, thereby cancelling the common unobserved denominator $\sum_{j=1}^{I} A_{j,t}$. Taking natural logarithms on both sides then yields the $I-1$ equations

$$\log M_{i,t} = \mu_i^* + \lambda_i \log M_{i,t-1} + \log M_{I,t} - \lambda_I \log M_{I,t-1} + \varepsilon_{i,t}^*, \tag{5}$$

for $i = 1, 2, \ldots, I-1$, where μ_i^* equals $\mu_i - \mu_I$ and $\varepsilon_{i,t}^*$ equals $\varepsilon_{i,t} - \varepsilon_{I,t}$. Notice that $\varepsilon_{i,t}^*$ is again a white noise process, that is, $\varepsilon_{i,t}^*$ is uncorrelated with $\varepsilon_{i,t-k}^*$ for $k = 1, 2, \ldots$. The individual parameters μ_i, $i = 1, 2, \ldots, I$, are not identified, and the variance-covariance matrix for $\varepsilon_{i,t}^*$ is an unrestricted matrix $\sum^*$ of dimension $I-1$. Furthermore, note that the parameter λ_I appears in each of the $I-1$ equations. Due to all this, the parameters in (5) should be estimated using an Iterative Seemingly Unrelated Regression [ISUR] technique (which is available in for example Eviews 4.0), where the parameter for log $M_{I,t-1}$ is restricted to be equal across the equations, see also Franses and Paap (1999). The resultant estimates are the Maximum Likelihood estimates, and hence they are consistent and efficient.

To examine the unit root properties of the I (log) market shares, where it is assumed that each series can have at most a single unit root, one can focus on each equation separately, while of course the parameters are jointly estimated using the ISUR method. By the assumption of at most a single unit root, we rule out the possibility of explosive behavior in market shares (which is likely to happen if there are two unit roots). Given the logical consistency restrictions on the market shares, this seems a plausible assumption. Each of the $I-1$ equations can now be viewed as one of the two equations of a bivariate system containing log $M_{i,t}$ and log $M_{I,t}$, where the second equation for log $M_{I,t}$ is left unspecified for the moment. For each pair of equations, one can distinguish four relevant cases. The first is the case where $\lambda_i = 1$ and $\lambda_I = 1$. In that case, (5) reduces to

$$\log M_{i,t} - \log M_{i,t-1} = \mu_i^* + \log M_{I,t} - \log M_{I,t-1} + \varepsilon_{i,t}^*. \tag{6}$$

Given that $\varepsilon_{i,t}^*$ is white noise and that both variables appear in first-differenced form, it trivially follows that log $M_{i,t}$ and log $M_{I,t}$ cannot both be stationary. Put otherwise, each of them may be integrated of order 1 [I(1)]. Clearly, when $\lambda_i = 1$ and $|\lambda_I| < 1$, log $M_{i,t}$−log $M_{i,t-1}$ is a function of a stationary variable, and hence the log $M_{i,t}$ series itself is [I(1)]. Given that one can always choose a brand i as the benchmark brand (instead of brand I), the third case where $\lambda_I = 1$ and $|\lambda_i| < 1$ implies that log $M_{I,t}$ is I(1). Finally, the fourth case, where $|\lambda_i| < 1$ and $|\lambda_I| < 1$, naturally implies that both (log) market shares are I(0). In sum, the values of the λ_i and λ_I parameters imply unit root properties of the corresponding log market shares.

The above-sketched approach is rather problematic for at least two reasons. The first is that there are cross-equation parameter restrictions. One consequence is that in case one wants to examine all log market shares at the same time, one may end up testing for 0 versus I unit roots. This would require developing new asymptotic theory for the test statistics, which is caused by the fact that the $I-1$ equation system corresponds with a VAR(1) model for I log market shares with several specific parameter restrictions. In fact, one then should resort to so-called panel unit root techniques, see Phillips and Moon (1999).

A second reason which makes the above approach difficult to use in practice is that there are many possible ways to parameterize a function for the attractions, and hence that there are many possible reduced form models. For example, suppose one wants to consider

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) M_{i,t-1}^{\lambda_i} M_{j,t-1}^{\lambda_j}, \tag{7}$$

for some brand j, that is, the attraction of each brand i also depends on that of a brand j, then the $I-1$ equations look rather different, and also other parameter restrictions than those discussed earlier lead to unit root properties. Interestingly enough, it is not easy to test which reduced form model is more adequate, as these tests and their asymptotic distributions in turn depend again on the unit root properties of the individual variables.

In sum, a unit root analysis of a set of $I-1$ reduced form equations leads to serious complications. As the attraction model naturally leads to an analysis of the variables $\log M_{i,t} - \log M_{I,t}$, it seems perhaps most useful to start an empirical study with these variables. To cover a wide variety of possible specifications, it seems most easy to start with an unrestricted vector autoregression [VAR] for the $I-1$ differences between the log market shares. In the next subsection we will discuss the consequences of taking this perspective.

2.2. *Implied Properties for Individual Market Shares*

In this subsection, we will argue that knowledge of the properties of $\log M_{i,t} - \log M_{I,t}$ can lead to insights into the unit root properties of the individual log market shares.

As any of the I brands can be the base-brand, (1) can be written such that any brand appears on the left-hand side. For example, when $I = 3$, we have the three approximate equations

$$\log M_{1,t} \approx c_1 - m_2 \log \frac{M_{2,t}}{M_{1,t}} - m_3 \log \frac{M_{3,t}}{M_{1,t}}, \tag{8}$$

$$\log M_{2,t} \approx c_2 - m_1 \log \frac{M_{1,t}}{M_{2,t}} - m_3 \log \frac{M_{3,t}}{M_{2,t}}, \tag{9}$$

and

$$\log M_{3,t} \approx c_3 - m_1 \log \frac{M_{1,t}}{M_{3,t}} - m_2 \log \frac{M_{2,t}}{M_{3,t}}, \tag{10}$$

where of course the restriction holds that $m_1 + m_2 + m_3 = 1$. Hence, the unit root properties of the log ratios of market shares imply unit root properties of the individual log market shares.

From (2), one can deduce that if there are R stable relationships amongst the $I-1$ variables $y_{1,t}$ to $y_{I-1,t}$, there are $R+1$ such relationships amongst the log market shares. This is because the $(R+1)$-th stable relationship amongst the log market shares is simply given by (2). Hence, when the entire system is stable, that is $R = I-1$, then all I log market shares are stable. Additionally, when $R = 0$, then all log ratios are I(1) and there is no cointegration amongst these variables. Using (1) and (2), one can conclude that then all log market shares are I(1), that is, have a unit root. When $0 < R < I-1$, one needs to add a second step to the cointegration analysis, which involves tests for linear restrictions on the cointegration parameters.

Perhaps this is all best illustrated using an example. Consider again the example with 3 brands, and take brand 3 as the benchmark brand. If there are two stable relationships in a multivariate time series model for $y_{1,t}$ and $y_{2,t}$, then $R = 2$. Given (10), this implies that log $M_{3,t}$ is stationary. As $y_{j,t}$ equals log $M_{j,t}$ minus log $M_{3,t}$, for $j = 1, 2$, the two other log market shares should also be stationary. Secondly, if there are no stable relations in the bivariate system, that is, $R = 0$, then log $M_{3,t}$ must be I(1) as well. Due to (2), log $M_{1,t}$ and log $M_{2,t}$ cannot be I(0) at the same time. Neither can only one of the log market shares be I(1), due to either (8) or (9). For the intermediate case with $R = 1$, there are several possibilities. If this cointegration relationship corresponds with the unity vector (10), then $y_{1,t}$ is stationary. As $y_{2,t}$ is I(1), we can deduce from (10) that log $M_{3,t}$ is I(1). As $y_{1,t}$ is I(0), this automatically implies that log $M_{1,t}$ should be I(1) too. For log $M_{2,t}$, however, it is possible to be either I(1) or I(0). The only way to find out its properties is to take another brand as the benchmark and repeat the analysis. Naturally, the value of R will not change by doing so.

3. Testing for Unit Roots

In this section we provide some guidelines for testing for unit roots using the Johansen (1995) method in unrestricted VAR models. We pay specific attention to the specification of the testing model, as there may seem to be some confusion about its proper form.

Without loss of generality, consider the VAR(1) model

$$Y_t = \Phi_1 Y_{t-1} + e_t, \tag{11}$$

for an $(I-1)\times 1$ times series Y_t containing $y_{1,t}$ through $y_{I-1,t}$, where e_t is an $(I-1)\times 1$ vector white noise series. For cointegration analysis it is convenient to write (11) in the so-called equilibrium correction format, that is,

$$\Delta_1 Y_t = \Pi Y_{t-1} + e_t, \tag{12}$$

where $\Pi = \Phi_1 - E_{I-1}$, and where δ_1 denotes the first differencing operator, and where E_{I-1} denotes an $I-1$-dimensional identity matrix. The matrix Π contains information on cointegrating relations between the $I-1$ elements of Y_t. In cointegration analysis it is common to write (12) as

$$\Delta_1 Y_t = \alpha\beta' Y_{t-1} + e_t, \tag{13}$$

where α and β are $(I-1)\times R$ full rank matrices. When $0<R<I-1$, there are R cointegrating relations between the $I-1$ variables, see Engle and Granger (1987) and Johansen (1995).

The maximum likelihood cointegration test method, developed in Johansen (1988), tests the rank of the matrix Π using the reduced rank regression technique based on canonical correlations. For model (13) this amounts to calculating the canonical correlations between $\Delta_1 Y_t$ and Y_{t-1}. This gives the eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{I-1}$ and the corresponding eigenvectors $\hat{\beta}_1, \ldots, \hat{\beta}_{I-1}$. A reliable test for the rank of Π is the likelihood ratio [LR] test statistic Q

$$Q = -n \sum_{i=R+1}^{I-1} \log(1-\hat{\lambda}_i). \tag{14}$$

The null hypothesis is that there are at most R cointegration relations. Asymptotic theory for Q is given in Johansen (1995), and the critical values for this Q for model (13) are given in Table 15.1 in Johansen (1995). Additionally, the procedure gives standard errors around the parameters $\hat{\beta}$ and these can be used to see if there are unity vectors amongst $\hat{\beta}_1, \ldots, \hat{\beta}_{I-1}$.

Notice that the model in (13) assumes that the $I-1$ time series do not have a trend, and that the cointegrating relations $\beta' Y_t$ have zero equilibrium values. This may however not be a reasonable assumption for differences across log market shares or for the log market shares themselves. Hence, it seems relevant to expand (13) by somehow incorporating an intercept and trend in the VAR model. This expansion should be such that the data display the same kind of behavior under the various hypotheses.

There are now two cases which are most relevant in practice. The first case concerns the case where none of the $I-1$ time series seems to display a deterministic trend. First, the restriction that the cointegrating relations $\beta' Y_t$ in (13) all have an equilibrium value which is exactly equal to zero does not seem plausible. Hence, it is more appropriate to extend (13) as follows, that is,

$$\Delta_1 Y_t = \alpha(\beta' Y_{t-1} - \mu_1) + e_t. \tag{15}$$

This representation also implies that in case there are no cointegrating relations, the data do not have a trend.

To compute the LR statistic, one should now calculate the canonical correlations between $\Delta_1 Y_t$ and $(Y_{t-1}, 1)'$. The relevant asymptotic theory is given in Johansen (1995). The critical values of the corresponding LR test appear in Table 15.2 in Johansen (1995). Notice that this case corresponds with Option 2 in the relevant routine in EViews.

The second case concerns the case when (some or all) series display trending patterns. One then should consider

$$\Delta_1 Y_t = \mu_0 + \alpha(\beta' Y_{t-1} - \mu_1 - \delta_1 t) + e_t. \tag{16}$$

This model allows the individual time series to have trends by not restricting μ_0 to zero, while the cointegrating relations attain their equilibrium values at $\mu_1 + \delta_1 t$. In very special cases, all parameters in δ_1 may equal 0, but it is safe not to assume this beforehand.

To compute the LR statistic, one should calculate the canonical correlations between demeaned first differenced series and demeaned $(Y_{t-1}, t)'$. The relevant asymptotic theory is again given in Johansen (1995). The critical values of the LR test appear in Table 15.4 in Johansen (1995). This second case corresponds with Option 4 in the relevant routine in EViews.

In case one *a priori* assumes that $\delta_1 = 0$ in (16), one implicitly assumes that there are links between the deterministic growth patterns across the $I-1$ individual time series. This assumption has an impact on the value of the LR test statistic and on its asymptotic distribution. The relevant theory is given in Johansen (1995), and the critical values appear in Table 15.3 in Johansen (1995). This case corresponds with (the default) Option 3 in EViews. However, as mentioned above, the assumption that $\delta_1 = 0$ may not be a sensible assumption for many economic data, and also it is difficult to test without having precise knowledge of the number of cointegrating relations.

Finally, when the VAR model order is in excess of 1, the above calculations need to be adjusted by also considering regressions of the first-differenced variables. The main practical issue is of course the specification of the lag order of this VAR. A commonly applied procedure involves the use of a model selection criterion, such as AIC or BIC, together with diagnostic tests for residual autocorrelation.

4. Empirical Illustrations

In this section the method proposed in the previous section is considered in a simulation experiment and it is applied to two sets of market shares.

4.1. Monte Carlo Simulations

To see whether our procedure yields useful inference, in contrast to a straightforward application of the Dickey-Fuller test to each of the market share series, consider the following data generating process, that is,

$$\log A_{1,t} = 0.5 \ \log A_{1,t-1} + \varepsilon_{1,t} \tag{17}$$

$$\log A_{2,t} = \log A_{2,t-1} + \varepsilon_{2,t} \tag{18}$$

$$\log A_{2,t} + \log A_{3,t} = 0.5(\log A_{2,t-1} + \log A_{3,t-1}) + \varepsilon_{3,t}, \tag{19}$$

where the $\varepsilon_{i,t}$ series, $i = 1, 2, 3$, are mutually independent standard white noise series. From this, we calculate $\log M_{i,t} = \log A_{i,t} - \log \sum_{i=1}^{3} A_{i,t}$ and $y_{i,t} = \log M_{i,t} - \log M_{1,t}$ $= \log A_{i,t} - \log A_{1,t}$. It easily follows that $y_{2,t}$ and $y_{3,t}$ are both I(1), and that they are cointegrated with cointegrating vector (11) Hence, $\log M_{1,t}$ is I(0), whereas $\log M_{2,t}$ and $\log M_{3,t}$ are I(1). The rank R is therefore equal to 1.

First, we consider the familiar Dickey and Fuller (1981) test procedure, where we use the F-version by testing the joint significance of the lagged level and the intercept. We use the 5% significance level for this and our test. Some experimentation suggests that four lagged difference need to be included. When we rely on 1000 replications, we find that for a sample size of 100 observations, the test correctly indicates that $\log M_{1,t}$ is I(0) in only 31% of the cases, and that $\log M_{2,t}$ and $\log M_{3,t}$ are I(1) in 91% and 91% of the cases, respectively. When we increase the sample size to 500 data points, these percentages become 27%, 88% and 87%, respectively. Hence the Dickey-Fuller test does not seem to be very reliable, and worse, its performance deteriorates for larger samples.

Next, we consider our procedure based on the log ratios. For the present data generating process, we should find that $R = 1$. Again, we use 1000 replications. For the sample size of 100, we find that R is estimated to equal 1 in 57% of the cases, while it is 0 and 2 in 38% and 5% of the cases, respectively. When we increase the sample size to 500, these percentages become 0%, 95% and 5%. Hence, our procedure works quite well and seems to improve substantially on the univariate Dickey-Fuller procedure.

4.2. Two Real-life Data Sets

We now apply our method to two FMCG scanner data sets for illustrative purposes. These sets concern four brands in the ketchup category. The same data have been analyzed in several marketing research studies, see Srinivasan and Bass (2000) for references. As none of the log market shares seem to display a deterministic trend, we consider the test equation as in (15). We experimented with the second test regression, but found no salient differences in the outcomes. We also have results for many other sets of data, but the two considered here suffice for illustrative purposes.

For the first set, we use A. C. Nielsen weekly scanner data for the period 1986–1988 ($n = 104$) for the ketchup category. The four brands are Heinz, Hunts, Del Monte and a rest category. This rest category is taken as the base-brand. We consider a VAR model for the three variables measuring the differences between the log market shares. Model selection criteria and diagnostic tests suggest that a VAR model with first order dynamics suffices to describe the data. The LR test statistics for $R = 0$ versus $R > 0$, $R = 1$ versus $R > 1$, and $R = 2$ versus $R > 2$ take the values 96.07, 46.49 and 17.28, respectively, and these should be compared with the 5% critical values 34.91, 19.96 and 9.24, respectively. The three null hypotheses get rejected convincingly, and hence given that $R = 3$, it can be concluded that the four market shares in the ketchup category are all stationary at the 5% significance level.

The second set of market shares are A. C. Nielsen store movement data, and they concern the same brands of the same product category as in the first set, although now the sample concerns 104 observations for the years 1992 to 1994. Again, we take the rest category as the base-brand. The residual diagnostics and model selection criteria indicate that a first order VAR can be considered for the three log ratios of market shares. The LR test statistics for $R = 0$ versus $R > 0$, $R = 1$ versus $R > 1$, and $R = 2$ versus $R > 2$ take the values 101.33, 56.23 and 18.52, respectively, and these should again be compared with the 5% critical values 34.91, 19.96 and 9.24, respectively. Hence, the four market shares are found to be stationary at the 5% significance level.

5. Concluding Remarks

In this paper, we proposed a new testing procedure for diagnosing stationarity versus evolution in market shares within the context of a market share attraction [MCI] model, accommodating the logical consistency constraint that market shares lie between 0 and 1 and that they sum to unity. In testing for unit roots, care must be exercised to take account of the fact that the market shares of brands within a single product category are related. An important conclusion from our paper is that one should consider the differences of the log market shares with a base-brand first, and next, one should derive the properties of individual log market shares. Simulation results suggest that our procedure is more useful than the univariate Dickey-Fuller method.

There are limitations, however. First, while we considered univariate log market shares in an MCI model, the case of relating evolving brands with explanatory variables is more complicated, as it may include many more statistical decisions. Second, in order to enable parameter estimation, the model assumed a transformation of market shares to log market shares. A question that arises is what can one infer about the properties of levels of the market shares based on these tests. This is not a trivial matter, as it is unclear if exponential functions of I(0) or I(1) variables are again I(0) and I(1), see Park and Phillips (1999). Third, the unit root testing methodology we proposed did not explicitly take into account structural breaks in market shares. Future research should address extensions to the unit root tests accommodating structural breaks within the MCI framework.

Limitations aside, an important managerial issue that has been increasingly addressed in research in marketing is the long-term share and sales impact of manufacturers' marketing actions. Our paper proposes a new approach for testing for unit roots in market shares and can help researchers in this important area.

Appendix

Let $\{M_{i,t}, i = 1, \ldots, I, t = 1, \ldots, T\}$ denote time series of market shares of I brands at time t, such that $M_{i,t} \geq 0$ and $\sum_{i=1}^{I} M_{i,t} = 1$. The attraction model is given by

$$M_{i,t} = \frac{A_{i,t}}{\sum_{i=1}^{I} A_{i,t}},$$

where $A_{i,t} \geq 0$, and $(\log A_{1,t}, \ldots, \log A_{I,t})'$ follows a vector time series process. It is useful to consider the $I-1$ relative log-market shares $\{y_{i,t}, i = 1, \ldots, I-1\}$ defined by

$$y_{i,t} = \log \frac{M_{i,t}}{M_{I,t}} = \log \frac{A_{i,t}}{A_{I,t}} = \log A_{i,t} - \log A_{I,t},$$

so that a linear vector time series process for $(\log A_{1,t}, \ldots, \log A_{I,t})'$ implies a linear vector time series process for $y_t = (y_{1,t}, \ldots, y_{I-1,t})'$. The choice of brand I as the benchmark is arbitrary and irrelevant: any other choice of the benchmark leads to a vector of log-market shares which is a simple linear transformation of y_t.

The inverse transformation from y_i,t to $M_{i,t}$ follows from $M_{i,t} = \exp(y_{i,t})M_{I,t}$ and $M_{I,t} = \left(1 - \sum_{i=1}^{I-1} M_{i,t}\right)$, so that

$$\frac{1 - M_{I,t}}{M_{I,t}} = \frac{\sum_{i=1}^{I-1} M_{i,t}}{M_{I,t}} = \sum_{i=1}^{I-1} \exp(y_{i,t}),$$

and hence

$$M_{I,t} = \frac{1}{1 + \sum_{i=1}^{I-1} \exp(y_{i,t})},$$

$$M_{i,t} = \frac{\exp(y_{i,t})}{1 + \sum_{i=1}^{I-1} \exp(y_{i,t})}, i = 1, \ldots, I-1.$$

This shows that the $(I-1)$-vector y_t fully determines the I-vector of market shares.

Let $m_{i,t} = \log M_{i,t}$. From the previous analysis it follows that $m_t = (m_{1,t}, \ldots, m_{I,t})'$ is a non-linear transformation of the linear time series process y_t. A first-order Taylor series approximation may yield some indication of the time series behaviour of m_t, especially with respect to the stationarity or non-stationarity of m_i,t. The matrix of derivatives $G(y_t)$ of the function $m(y_t)$ contains

$$G_{ij}(y_t) = \frac{\partial m_{i,t}}{\partial y_{j,t}} = -M_{j,t}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, i-1, \; i+1, \ldots, I-1,$$

$$G_{ii}(y_t) = \frac{\partial m_{i,t}}{\partial y_{i,t}} = 1 - M_{i,t}, \quad i = 1, \ldots, I-1.$$

Therefore, a linearization of $m(y_t)$ in a point $\bar{y}$, corresponding to log-market shares $\bar{m} = \log \bar{M}$, is

$$m_{i,t} \approx \bar{m}_i + y_{i,t} - \bar{y}_i - \sum_{j=1}^{I-1} \bar{M}_j (y_{j,t} - \bar{y}_j), \quad i = 1, \ldots, I-1,$$

$$m_{I,t} \approx \bar{m}_I - \sum_{j=1}^{I-1} \bar{M}_j (y_{j,t} - \bar{y}_j)$$

If we take $\bar{y} = 0$, such that $\bar{M}_i = 1/I$, then this simplifies to

$$m_{i,t} \approx -\log I + y_{it} - \frac{1}{I} \sum_{j=1}^{I-1} y_{j,t}, \quad i = 1, \ldots, I-1,$$

$$m_{I,t} \approx -\log I - \frac{1}{I} \sum_{j=1}^{I-1} y_{j,t}.$$

In both cases it is quite easy to see that $\sum_{i=1}^{I} \bar{M}_i m_{i,t} \approx$ constant.

Acknowledgements

We thank Frank Bass, Bart Bronnenberg, Richard Paap, Peter Popkowski Leszczyc and two anonymous referees for their helpful comments. This paper contains a few simulation and empirical results for illustrative purposes. Further simulation and empirical results can be obtained from the corresponding author.

References

Banerjee, Anindya, Juan Dolado, John W. Galbraith, and David F. Hendry. (1993). *Co-integration, Error-correction, and the Econometric Analysis of Nonstationary Data*. Oxford: Oxford University Press.

Bronnenberg, Bart J., Vijay Mahajan, and Wilfred R. Vanhonacker. (1999). "The Emergence of Market Structure in New Repeat-Purchase Categories: The Interplay of Market Share and Retailer Distribution," *Journal of Marketing Research*, 37, 16–31.

Cooper, Lee G., and M. Nakanishi. (1988). *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*. Boston: Kluwer Academic Publishers.

Dekimpe, Marnik G., and Dominique M. Hanssens. (1995). "Empirical Generalizations About Market Evolution and Stationarity," *Marketing Science*, 14, G109–G121.

Dekimpe, Marnik G., and Dominique M. Hanssens. (2000). "Time Series Models in Marketing: Past, Present and Future," *International Journal of Research in Marketing*, 17, 183–193.

Dekimpe, Marnik G., Dominique M. Hanssens, and Jorge M. Silva-Risso. (1999). "Long-Run Effects of Price Promotions in Scanner Markets," *Journal of Econometrics*, 89, 269–291.

Dickey, David A., and Wayne A. Fuller. (1981). "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49, 1057–1072.

Engle, Robert F., and Clive W. J. Granger. (1987). "Co-integration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

Franses, Philip Hans, and Richard Paap. (1999). Testing Market Share Attraction Models, Research report 9918 of the Rotterdam Institute for Business Economic Studies.

Franses, Philip Hans, Teun Kloek, and Andre Lucas. (1999). "Outlier Robust Analysis of Long-Run Marketing Effects for Weekly Scanning Data," *Journal of Econometrics*, 89, 293–315.

Johansen, Søren. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

Johansen, Søren. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Maddala, George S., and In-Moo Kim. (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.

Nijs, Vincent, Marnik Dekimpe, Jan-Benedict E.M. Steenkamp, and Dominique M. Hanssens. (2001). "The Category Demand Effects of Price Promotions," *Marketing Science*, 20, 1–22.

Park, Joon Y., and Peter C. B. Phillips. (1999). "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15, 269–298.

Pauwels, Koen, Dominique M. Hanssens, and S. Siddarth. (2000). The Long-term Effect of Price Promotions on Category Incidence, Brand Choice and Purchase Quantity, Working paper, Anderson Graduate School of Management, UCLA.

Phillips, Peter C. B., and Hyungsik R. Moon. (1999). "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–1111.

Phillips, Peter C. B., and Zhijie Xiao. (1998). "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12, 423–469.

Srinivasan, Shuba, and Frank M. Bass. (2000). "Cointegration Analysis of Brand Sales and Category Sales: Stationarity and Long-run Equilibrium in Market Shares," *Applied Stochastic Models in Business and Industry*, 16, 159–177.

Srinivasan, Shuba, Peter Popkowski Leszczyc, and Frank M. Bass. (2000). "Market Share Response and Competitive Interaction: The Impact of Temporary, Evolving and Structural Changes in Prices," *International Journal of Research in Marketing*, 17, 281–305.