Names of authors:
Ryuji Wakuta

University address, telephone number, e-mail address:
Kyoto Sangyo University
Motoyama, Kamigamo
Kyoto, JAPAN 603-8555
telephone: +81-75-705-0910
e-mail: wakuta@cc.kyoto-su.ac.jp

Acknowledgment
This work was supported by JSPS KAKENHI Grant Number JP 16K21470

Can we compare visit frequencies among Japanese coffee shops?
Comparative study using self-answered data and panel data with the NBD Dirichlet model

Abstract

This study aims to reveal the research method that most precisely evaluates the visit frequencies of competitive Japanese coffee shops. Specifically, this study answers two central questions. The first question is whether the Japanese coffee shop market can be approximated using the negative binomial distribution (NBD) Dirichlet model when we use self-answered data collected from consumer questionnaires as well as panel data gathered from consumers' receipts. The second question is whether the Japanese coffee shop market can be more precisely approximated with panel data than self-answered data.

According to Reichheld and Sasser (1990), service companies such as credit cards and retail stores should increase their customer retention rates to improve profitability because the cumulative transaction costs will tend to decrease. Therefore, many retail stores, including coffee shops, have tried to increase the frequency of their customers' visits. In contrast, Sharp (2010) pointed out that service companies should not simply increase their customer retention rate because the customer retention rate they observe is influenced by their market penetration. Sharp et al. (2002) revealed that market penetration could influence the customer retention rate if the market could be approximated using the NBD Dirichlet model. Moreover, Heiens and Larry (2014) found that the coffee shop market in the state of Kuwait could be approximated with the NBD Dirichlet model using self-answered data collected from questionnaires. However, little research uses data similar to sales results data to confirm whether the NBD Dirichlet model can approximate service markets in non-western countries. According to Wright et al. (2002), the NBD Dirichlet model is constrained because panel data is needed for model estimation.

This study analyses data collected using two different methods. One is a simple questionnaire in which consumers are asked to recall their past visit frequencies and visit experiences at each coffee shop over one week. The other is panel data, developed by asking consumers to hand in their receipts from the coffee shops they actually visited over five weeks. This study examines whether the Japanese coffee shop market can be approximated using the NBD Dirichlet model through analysis of each type of data.

As a result, this study reveals the following two findings. The first finding is that the Japanese coffee shop market can be approximated using the NBD Dirichlet model when we analyse self-answered data from consumer questionnaires as well as panel data based on their receipts. The second finding is that the Japanese coffee shop market can be more precisely approximated when we use panel data than when we use self-answered data. Therefore, both research methods are useful for comparing visit frequencies among Japanese coffee shops. However, we should not compare the values of visit frequencies among Japanese coffee shops, but should regard their values as values influenced by their market penetration. Marketers of coffee shops should eliminate the influence of market penetration when making comparisons.

Key words

NBD Dirichlet model, Coffee shops, Visit frequency, Panel data, Self-answered data

Introduction and Objectives

Larger brands not only have more buyers, but these buyers also tend to buy more often. In contrast, smaller brands not only have fewer buyers, but those buyers tend to buy those brands less often, resulting in smaller brands being punished twice. Many marketing scholars such as Ehrenberg et al. (1990) and Sharp (2010) call this phenomenon 'double jeopardy'.

According to McPhee (1963), double jeopardy happens in the following two ways. Of the many people who choose well-known coffee shop A, if asked, nearly all will say it is their favourite (because few are even aware of the more obscure coffee shop B). Of the few people who are aware of B, at most half will say it is their favourite since most of them will also be aware of the well-known A. They therefore split their vote.

Using the negative binomial distribution (NBD) Dirichlet model, Ehrenberg et al. (1990) revealed that double jeopardy emerged in daily commodity markets. Because they were able to approximate the American instant coffee market using the NBD Dirichlet model, they found larger instant coffee brands not only have more buyers, but those buyers tend to buy more often. The NBD Dirichlet model can approximate various repertoire markets, such as laundry detergent (Uncles et al., 1995; Ehrenberg and Goodhardt, 2002), shampoo (Yang et al., 2007), sportswear (Dawes, 2009), and even automobiles (Bennet and Graham, 2010).

Sharp et al. (2002) found the NBD Dirichlet model could approximate subscription service markets as well as repertoire markets. They found the NBD Dirichlet model could approximate the Australian and New Zealand credit card service market less precisely than the gas station market. Sharp (2010) found the bank market to be the same.

Sharp et al.'s (2002) finding is very important because based Reichheld et al.'s (1990) finding, many service company marketers believe they should increase customers' purchase frequency to improve their firm's profitability. According to Reichheld et al. (1990), service companies such as credit cards and retail stores should increase their customer retention rate to improve profitability because cumulative transaction costs will tend to decrease. Therefore, many retail stores, including coffee shops, have attempted to increase their customers' visit frequencies. However, if the NBD Dirichlet model can approximate such markets, marketers of service companies should evaluate their customer retention rate very carefully because their firm's market penetration may influence their customer retention rate.

Recently, Heiens et al. (2014) found that the coffee shop market in the State of Kuwait could be approximated using the NBD Dirichlet model. Therefore, marketers of coffee shops in the State of Kuwait should evaluate their customers' visit frequencies very cautiously, because the possibility remains that market penetration influenced visit frequency. For example, according to our research, the average visit frequency of Japanese coffee shop H was 2.08 and that of Japanese coffee shop D was 2.26. Normally, many marketers would believe coffee shop H should improve their customers' visit frequencies to a level similar to that of coffee shop D. Is this true? It may be not true if the market penetration of coffee shop D is quite a bit higher than that of coffee shop H.

Unfortunately, Heiens et al. (2014) used self-answered data from a questionnaire. Therefore, it is still unknown whether the coffee shop market can be approximated using the NBD Dirichlet model by collecting panel data based on records of consumers' actual purchase behaviour. Will it lead to the same results as in Heiens et al. (2014)?

This study aims to reveal the research method that will most precisely evaluate visit frequencies at competitive Japanese coffee shops. Specifically, this study answers two central questions. The first question is whether the Japanese coffee shop market can be approximated using the NBD Dirichlet model when we use self-answered data collected from consumers' questionnaires as well as panel data gathered from consumers' receipts. The second is whether the Japanese coffee shop market can be more precisely approximated when we use panel data rather than self-answered data.

This study analyses data collected with two methods. One is a simple questionnaire in which consumers are asked to recall their past visit frequencies and visit experiences of each coffee shop during one week, a replication of Heiens et al.'s (2014) study. The second is panel data gathered by asking consumers to hand in their receipts from the coffee shops they actually visited over five weeks. After collecting each set of data, this study examines whether the Japanese coffee shop market can be approximated using the NBD Dirichlet model.

Conceptual Framework and Literature Review

1. NBD Dirichlet Model

The NBD Dirichlet model has three assumptions. First, the category purchase rate is assumed to have a negative binomial distribution. Second, the purchases of individual brands are assumed to have a Dirichlet multinomial distribution. Finally, the two distributions are assumed to be otherwise independent.

According to Johnson et al. (1993), the probability density function for the negative binomial distribution can be described as Equation (1). Equation (1) has two parameters: γ is the shape parameter and β is the scale parameter. The category purchase rate K consists of the category purchase rate k for each individual consumer.

$$f_{\gamma,\beta}(k) = \frac{\Gamma(\gamma+k)}{\Gamma(\gamma)k!}\frac{\beta^k}{(1+\beta)^{(\gamma+k)}} \quad k = 0,1,2,... \quad (1)$$

According to Johnson et al. (1997), the probability density function for the Dirichlet multinomial distribution can be described as Equation (2). Equation (2) has h parameters, one for each brand. These parameters are $\alpha_1, \alpha_2, ..., \alpha_h$ where each is positive. The purchase rates of individual brands $R_1, R_2, ..., R_h$ consists of the brand purchase rates $r_1, r_2, ..., r_h$ for each individual consumer.

$$f_{\alpha_1,\alpha_2,...,\alpha_h}(r_1, r_2, ..., r_h \mid r_1 + r_2 + ... + r_h = k) = \frac{\Gamma\left(\sum_{j=1}^{h}\alpha_j\right)k!}{\Gamma\left(\sum_{j=1}^{h}\alpha_j + k\right)}\prod_{j=1}^{h}\frac{\Gamma(\alpha_j + r_j)}{r_j!\Gamma(\alpha_j)} \quad (2)$$

According to Rungie and Goodhardt (2004), the NBD Dirichlet model can be described as Equation (3). This model brings together the two probability density functions in Equations (1) and (2).

$$f_{\gamma,\beta,\alpha_1,\alpha_2,...,\alpha_h}(r_1, r_2, ..., r_h) = f_{\gamma,\beta}(k) f_{\alpha_1,\alpha_2,...,\alpha_h}(r_1, r_2, ..., r_h \mid r_1 + r_2 + ... + r_h = k) \quad (3)$$

Ehrenberg et al. (1990) found that larger instant coffee brands not only have more buyers, but their buyers tend to buy more often. As mentioned previously, they found the American instant coffee market could be approximated using the NBD Dirichlet model (Goodhardt et al., 1984). In addition to this finding, Ehrenberg et al. (1990) explained that double jeopardy existed without any kind of marketing activities by any firms. They referred to the following explanation in McPhee (1963).

Suppose there are just two restaurants in a town, and that 100 people who know both restaurants regard them as being of equal utility (equal in quality, service, value for money,

accessibility, etc). You would normally predict the visit frequency of each restaurant as 50%. However, suppose that one (which we call A) of the two restaurants is known by 80 people in the town and the other (which we call B) is known by 50 people in the town. 30 people in the town are aware of both restaurants.

If people are asked about their visit frequency to each restaurant, double jeopardy is bound to occur. 50 (= 80 - 30) people who are aware of only "A" restaurant will answer, "I visit A 100%". 20 (= 50 - 30) people who are aware of only "B" restaurant will answer, "I visit B 100%". And 30 people who are aware of both restaurants will answer, "I visit A 50% and B 50%". The average visit frequency of "A" is 81.25% $= (50 \times 100 + 30 \times 50) \div (50 + 30)$ and the average visit frequency of "B" is 70% $= (20 \times 100 + 30 \times 50) \div (20 + 30)$. These predictions are not the same as the normally expected equal rates. This occurs not only because fewer people eat at the more obscure restaurant, but also because they do not go there as frequently since about half the time they go to the one more widely known (which they regard as of equal merit). In contrast, most of the many who patronize the widely known one go there relatively often because they are not even aware of the obscure one.

2. Approximation accuracy in services markets

The NBD Dirichlet model was first used to approximate daily commodity markets. For example, Uncles et al. (1995) and Ehrenberg et al. (2002) revealed the NBD Dirichlet model could approximate the laundry detergent market. Yang et al. (2007) demonstrated it could approximate the shampoo market, while Dawes (2009) showed it could approximate the sportswear market. In addition to retail goods, Bennet et al. (2010) found it could even approximate the automobile market.

The NBD Dirichlet model also has been successfully applied to store choice in many studies. For example, Keng et al. (1998) revealed the NBD Dirichlet model could approximate the market for retail supermarkets. Meyer-Waarden and Benavent (2006) found it could approximate the retail mall market. Moreover, Barwise and Ehrenberg (1987) and Ehrenberg (1988) showed the NBD Dirichlet model could approximate the TV program market.

Recently, Sharp et al. (2002) found the NBD Dirichlet model could approximate subscription service markets as well as repertoire markets. They found the NBD Dirichlet model could approximate the Australian and New Zealand credit card service markets less precisely than the gas station market. Sharp (2010) also found it could approximate the bank market, while Baker et al. (2016) revealed it could approximate the professional sports market.

The finding of Sharp et al. (2002) has very important implications, because many marketers of service companies believe they should increase customers' visit frequency to improve their firm profitability, consistent with Reichheld et al. (1990). According to Reichheld et al. (1990), service companies such as credit card companies and retail stores should increase their customer retention rate to improve profitability because cumulative transaction costs will tend to decrease. Therefore, many retail stores, including coffee shops, have attempted to increase their customers' visit frequencies. For example, many coffee shops offer loyalty programs. However, if the NBD Dirichlet model can approximate such markets, marketers of service companies should evaluate their customer retention rate very carefully because their firm's market penetration may influence their customer retention rate.

To see whether the NBD Dirichlet model can approximate such markets, Wakuta (2015) evaluated the approximation accuracy in service markets by reanalysing Sharp et al.'s (2002) results. According to Wakuta (2015), Sharp et al. (2002) interpreted credit card markets as markets that could be approximated by the NBD Dirichlet model because credit card markets satisfied the following three conditions (see Table 1).

The first condition was that the MAPE (mean absolute percentage error) of market penetration was under 19.17% (Wright et al., 2002). The MAPE was lower than Sudman's (1964) 45% mean overstatement of recall share. The second condition was that the MAPE of purchase frequency was under 21.23%.

MAPE usually expresses accuracy as a percentage and is defined by the formula:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

where A_t is the actual value and F_t is the forecast value. It is a typical measure of forecast accuracy used to compare predicted market statistics to actual observations.

The third conditions were associated with the w(1 – b) model (Eherenberg et al., 1990). The w(l - b) model relates b_i (the proportion of consumers buying brand i at least once in the period analysed) and w_i (how often on average they buy it) to the corresponding values b_j and w_j for brand j. It states that $w_i(l - b_i) = w_j (l - b_j) = k$, where k is a constant (estimated as the average value of w(l - b) for all itemized brands). The third condition was that the standard deviation of w(1 - b) be under 22.335.

Table1. Approximation accuracy in Sharp et al.'s (2002) data

Market	N	MAPE		w(1-b)
		Market Penetration	Purchase Frequency	Std.
Credit Card (2001) NZ	7	19.17%	12.78%	17.578
Credit Card (2001) AUS	12	15.86%	21.23%	22.335

3. Difference between panel data and self-answered data

If the NBD Dirichlet model can approximate coffee shop markets, marketers of coffee shop companies should not simply evaluate their customers' average visit frequencies because the frequencies will be influenced by their market penetration. In contrast, if the NBD Dirichlet model cannot approximate coffee shop markets, marketers of coffee shop companies can evaluate their customers' average visit frequencies because they will not be influenced by their market penetration. In that case, we can simply compare average visit frequencies among competing coffee shops.

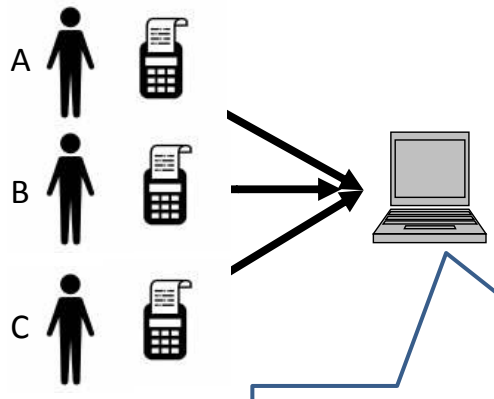
Heiens et al. (2014) solved this issue. They found that the coffee shop market in the state of Kuwait could be approximated using the NBD Dirichlet model. Therefore, marketers for coffee shops in the state of Kuwait should evaluate their customers' average visit frequencies very cautiously and eliminate the influence of market penetration. We cannot simply compare average visit frequencies among the competing coffee shops in the state of Kuwait.

However, we still do not know whether the coffee shop market can be approximated with the NBD Dirichlet model using panel data based on the records of consumers' actual purchase behaviour because Heiens et al. (2014) used self-answered data from a questionnaire. They used a survey of 618 consumers of coffee shops in the state of Kuwait. Wright et al. (2002) revealed there is not as much difference between panel data and self-answered data in supermarkets, retail fuel companies, and department stores. However, according to Wright et al. (2002), the NBD Dirichlet model is constrained because panel data is needed for model estimation.

How does the difference between panel data and self-answered data emerge? Suppose there are three consumers, A, B, and C. Consumer A visited Doutor (coffee shop) twice and Starbucks (coffee shop) twice during certain weeks. Consumer B visited Starbucks five times during the same weeks. Consumer C visited Starbucks once, Doutor once, and Tully's (coffee shop) once during the

same weeks. If we use panel data by collecting all their receipts, we will precisely estimate market penetration and the average visit frequency of each coffee shop.

As Figure 1 shows, Starbucks had 3 customers (A, B, and C), Tully's had 1 person (C) and Doutor had 2 people (A and C). Therefore, market penetration of Starbucks was 100% ($=3\div3$), that of Tully's was 33% ($=1\div3$) and that of Doutor was 67% ($=2\div3$). Starbucks was visited 8 times ($=2+5+1$), Tully's was visited 1 time ($=0+0+1$) and Doutor was visited three times ($=2+0+1$). Therefore, the average visit frequency of Starbucks was 2.67 ($=8\div3$), that of Tully's was 1.00 ($=1\div1$) and that of Doutor was 1.50 ($=3\div2$).



No.	ID	Starbucks	Tully's	Doutor
1	A	2	0	2
2	B	5	0	0
3	C	1	1	1

Number of Customers	3	1	2
Market Penetration	100%	33%	67%
Number of Visits	8	1	3
Average Visit Frequency	2.67	1.00	1.50

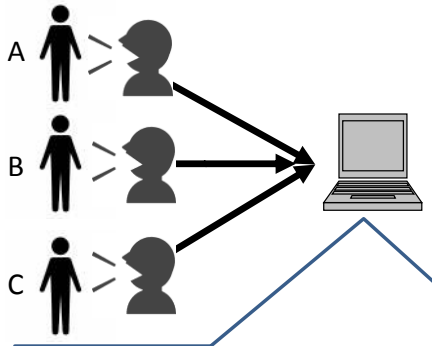
Figure 1: Panel data

On the other hand, we will not be able to estimate market penetration and average visit frequency of each coffee shop as precisely if we use self-answered data based on consumers' recollections of visits. This is because it is difficult for consumers to recall their visit frequencies of all coffee shops they visited over many weeks. Therefore, many studies have asked consumers two questions to measure their past visit frequencies more precisely. One question is about their last selection; for example, "What was the last coffee shop you visited?" The second question is about their visit frequencies, such as "How many times did you visit the selected shop during that particular time?"

If we use self-answered data and ask consumers these questions, the results will be as seen in Figure 2. First, consumers would be asked "Did you ever visit Starbucks during the particular weeks?", "Did you ever visit Tully's during the particular weeks?" and "Did you ever visit Tully's during the particular weeks?" Consumer A will answer "yes", "no", and "yes". Consumer B will answer "yes", "no", and "no". Consumer C will answer "yes", "yes", and "yes". Therefore, the number of Starbucks consumers would be 3 people (A, B, and C), Tully's would be 1 person (C) and Doutor's would be 2 people (A and C). Therefore, the market penetration of Starbucks was 100% ($=3\div3$), that of Tully's was 33% ($=1\div3$), and that of Doutor was 67% ($=2\div3$), the same as in the panel data.

Second, consumers would be asked "Which was the last coffee shop you visited?"

Consumer A would answer "Starbucks" because he (or she) visited Doutor twice and Starbucks twice during the particular weeks, as mentioned above. Consumer B would answer "Starbucks". Consumer C will answer "Tully's" because he (or she) visited Starbucks once, Doutor once, and Tully's once during the same weeks as mentioned above. Then, the consumers would be asked "How many times did you visit the selected shop during the particular time?" Consumer A will answer "Starbucks twice". Consumer B will answer "Starbucks 5 times". Consumer C will answer "Tully's just one time". The number of visits to Starbucks would be 7 $(=2+5+0)$, that of Tully's would be 1 $(=0+0+1)$ and that of Doutor would be 0 $(=0+0+0)$ because no one referred to Doutor. As a result, the average visit frequency to Starbucks would be 2.33 $(=7\div3)$, that of Tully's would be 1.00 $(=1\div1)$ and that of Doutor would be 0.00 $(=0\div2)$. These results are not same as the results using panel data.



No.	ID	Did you ever visit… Starbucks	Tully's	Doutor	Which was the last coffee shop you visited?	How many times did you visit the selected shop?
1	A	Yes	No	Yes	Starbucks	2
2	B	Yes	No	No	Starbucks	5
3	C	Yes	Yes	Yes	Tully's	1

Number of Customers	3	1	2
Market Penetration	100%	33%	67%
Number of Visits	7	1	0
Average Visit Frequency	2.33	1.00	0.00

Figure 2: Self-answered data

Method

1. Research Hypotheses

In this section, we suggest research hypotheses and how to collect data. We discuss the research hypotheses first and then explain the data collection.

To confirm that the Japanese coffee shop market can be approximated using the NBD Dirichlet model when we use self-answered data collected from consumers' questionnaires as well as panel data gathered from consumers' receipts, we used the approximation accuracy in the data of Sharp et al. (2002). Specifically, we considered the following three indicators for markets that can be approximated using the NBD Dirichlet model. First, the MAPE of market penetration is below 19.17%. Second, the MAPE of visit frequency is under 21.23%. Third, the standard deviation of w(1 - b) is under 22.335.

If the Japanese coffee shop market can be approximated with the NBD Dirichlet model when we use self-answered data, the approximation accuracy measured by these three indicators will be higher than the approximation accuracy using Sharp et al.'s (2002) data. This is hypothesis 1 (H1) in this study. We test H1 in Study 1 using a simple questionnaire in which consumers are asked to recall past visit frequencies and visit experiences at each coffee shop during one week.

H1: The Japanese coffee shop market can be approximated with the NBD Dirichlet model using self-answered data.
H1-1: The MAPE of market penetration is under 19.17%
H1-2: The MAPE of visit frequency is under 21.23%
H1-3: The standard deviation of w(1 - b) is under 22.335

If the Japanese coffee shop market can be approximated with the NBD Dirichlet model when we use panel data, the approximation accuracy measured by the three indicators will also be higher than that of the data in Sharp et al. (2002). This is hypothesis 2 (H2) in this study. We test H2 in Study 2 using panel data where consumers are asked to hand in their receipts from the coffee shops they actually visited over five weeks.

H2: The Japanese coffee shop market can be approximated with the NBD Dirichlet model using panel data.
H2-1: The MAPE of market penetration is under 19.17%
H2-2: The MAPE of visit frequency is under 21.23%
H2-3: The standard deviation of w(1 - b) is under 22.335

To confirm that the Japanese coffee shop market can be more precisely approximated with panel data than with self-answered data, we also used the approximation accuracy of the data in Sharp et al. (2002). If the Japanese coffee shop market can be more precisely approximated with panel data than self-answered data, the approximation accuracy in Study 2 will be higher than that in Study 1. This is hypothesis 3 (H3) in this study. We test H3 in Study 3 in which we compare the three indicators.

H3: The Japanese coffee shop market can be more precisely approximated with panel data than self-answered data
H3-1: The MAPE of market penetration in Study 2 is lower than that in Study 1.
H3-2: The MAPE of visit frequency in Study 2 is lower than that in Study 1.
H3-3: The standard deviation of w(1 - b) in Study 2 is lower than that in Study 1.

2. Data collection

This study analyses data collected using two different methods. One is a simple questionnaire in which consumers are asked to recall their past visit frequencies and visit experiences at each coffee shop during one week. This is Study 1. The second is panel data where consumers are asked to hand in their receipts from the coffee shops they actually visited during the five weeks. This is Study 2.

In both cases, the data was gathered by the research company, Soft Brain Field Co., Ltd. In Study 1, the research company recruited people to answer our questionnaire, delivered our questionnaire to them through the Internet, and gathered their responses through the Internet. They were asked the following four questions. First, they were asked "Did you ever visit a coffee shop from 23rd June 2014 to 29th June 2014 (one week)?" We confirmed that all of them answered yes. Second, they were asked to list the coffee shops they visited during that same period. Third, they were asked "Which is the last coffee shop you visited?" The final question was "How many times did you visit the selected shop during that same period?"

In Study 2, we used panel data collected by the research company, which gathered records from consumers by asking them to hand in their receipts from the coffee shops they actually visited. We used panel data based on consumers' receipts from 26th May 2014 to 29th June 2014 (five weeks).

Findings

1. Results of Study 1

We collected 330 questionnaire responses. Insufficient responses were eliminated using the list-wise method. Table 2 shows that the respondents comprised 142 males and 188 females. A total of 226 were married and 104 were unmarried.

Table 2: Sample for Study 1

		N	%
Gender	Male	142	43%
	Female	188	57%
Age	20-29	7	2%
	30-39	63	19%
	40-49	134	41%
	50-59	104	32%
	60-	22	7%
Married		226	68%
Unmarried		104	32%
N		330	100%

Table 3: Market penetration and visit frequency for Study 1

ID	Coffee Shop	Market Penetration O	Market Penetration T	Visit Frequency O	Visit Frequency T	w(1-b)
1	A	0.16	0.12	1.08	1.41	0.91
2	B	0.13	0.11	1.13	1.38	0.98
3	C	0.07	0.06	1.05	1.26	0.98
4	D	0.06	0.07	1.48	1.29	1.39
6	E	0.03	0.03	1.32	1.21	1.28

The market penetration of each coffee shop is computed using aggregated data from responses to the second question. We first aggregated the number of consumers of each coffee shop. We then divided the number of visits by the total number of consumers. The results are shown in Table 3 as *Market Penetration O*.

The average visit frequency of each coffee shop is calculated using the data from the third and fourth questions. First, we aggregated the number of visits of each coffee shop, and then divided the number of visits by the number of consumers of each coffee shop. The results are reported in Table 3 as *Visit Frequency O*.

The theoretical values of the NBD Dirichlet model are calculated using Kearns's (2009) software. The theoretical values for market penetration are presented in Table 3 as *Market Penetration T*, while the theoretical values of visit frequency are shown as *Visit Frequency T*. As Table 3 indicates, we also calculated w(1-b); w is *Visit Frequency O* and b is *Market Penetration O*.

To test hypothesis 1, we examined the approximation accuracy. We calculated three indicators, reported in Table 4. The first indicator was the MAPE of market penetration. It was 18.99% and therefore under 19.17%, supporting H1-1. The second indicator was the MAPE of purchase frequency. At 16.62%, it was below 21.23%, supporting H1-2. The third indicator was the standard deviation of w(1 - b), which was 17.238, under 22.335. H1-3 was also supported. As a result, hypothesis 1 was supported. The Japanese coffee shop market can be approximated with the NBD Dirichlet model using self-answered data. This result is accordance with the results of Heiens et al. (2014).

Table 4: Approximation accuracy of Study 1

Market	N	MAPE		w(1-b)
		Market Penetration	Visit Frequency	Std.
Coffee Shop (2014) based on Questionnaire	5	18.99%	16.62%	17.238

2. Results of Study 2

We collected 1,673 receipts from 440 individuals. Table 5 shows that the sample comprised 181 males and 259 females. A total of 302 were married and 138 were unmarried. 440 individuals were asked to hand in to the research company, Soft Brain Field Co., Ltd., their receipts from the coffee shops they actually visited from 26th May, 2014 to 29th June, 2014 (five weeks).

Table 5: Sample for Study 2

		N	%
Gender	Male	181	41%
	Female	259	59%
Age	20-29	9	2%
	30-39	83	19%
	40-49	183	42%
	50-59	130	30%
	60-	35	8%
Married		302	69%
Unmarried		138	31%
N		440	100%

The market penetration of each coffee shop is computed using the aggregated data from those receipts. We first counted each coffee shop's number of customers, then divided each shop's number of customers by 440, the total number of consumers. These are reported in Table 6 as *Market Penetration O*.

The average visit frequency of each coffee shop was also calculated using data from their receipts. First, we counted the number of visits of each coffee shop. We then divided the number of visits by the number of customers of each coffee shop. This is presented in Table 6 as *Visit*

Frequency O. The number of coffee shops (N=15) in Table 6 is greater than that (N=5) in Table 3 because we could include all visit experiences of each person.

Table 6: Market penetration and visit frequency for Study 2

ID	Coffee Shop	Market Penetration O	Market Penetration T	Visit Frequency O	Visit Frequency T	w(1-b)
1	A	0.34	0.36	2.23	2.10	1.46
2	B	0.31	0.31	2.04	2.01	1.41
3	C	0.27	0.28	2.03	1.95	1.48
4	D	0.26	0.30	2.26	1.98	1.67
5	E	0.14	0.14	1.87	1.76	1.61
6	F	0.12	0.13	1.90	1.74	1.68
7	G	0.11	0.10	1.44	1.70	1.28
8	H	0.06	0.07	2.08	1.67	1.96
9	I	0.05	0.04	1.26	1.64	1.19
10	J	0.04	0.03	1.39	1.63	1.33
11	K	0.04	0.04	1.88	1.64	1.81
12	L	0.04	0.04	1.65	1.64	1.58
13	M	0.03	0.03	1.40	1.63	1.35
14	N	0.03	0.03	1.43	1.63	1.38
15	O	0.03	0.03	1.67	1.63	1.62

The theoretical values of the NBD Dirichlet model are calculated using Kearns's (2009) software. The theoretical values of market penetration are presented in Table 6 as *Market Penetration T*, while the theoretical values of visit frequency are *Visit Frequency T*. As Table 6 shows, we also calculated w(1-b); w is *Visit Frequency O* and b is *Market Penetration O*.

To test hypothesis 2, we examined the approximation accuracy by calculating three indicators, as seen in Table 7. The first indicator was the MAPE of market penetration, which was 11.20%, under 19.17%. Therefore H2-1 was supported. The second indicator was the MAPE of visit frequency. It was 10.78%, under 21.23%, supporting H2-2. The third indicator was the standard deviation of w(1 - b). At 13.312, it was under 22.335, supporting H2-3. As a result, hypothesis 2 was supported. The Japanese coffee shop market can also be approximated with the NBD Dirichlet model using panel data. This result was also in accordance with the results of Heiens et al. (2014).

Table 7: Approximation accuracy of Study 2

Market	N	MAPE		w(1-b)
		Market Penetration	Visit Frequency	Std.
Coffee Shop (2014) based on Receipt Data	15	11.20%	10.78%	13.312

3. Results of Study 3

Study 3 examined whether the approximation accuracy in Study 2 was better than that in Study 1. We compared the three indicators in Tables 3 and 7 to test hypothesis 3.

The first indicator was the MAPE of market penetration, which was 11.20% in Table 7 and 18.99% in Table 3. Therefore, H3-1 was supported. The second indicator was the MAPE of visit frequency. This indicator was 10.78% in Table 7 and 16.62% in Table 3; therefore, H3-2 was also supported. The third indicator was the standard deviation of w(1 - b). At 13.312 in Table 7 and 17.328 in Table 3, H3-3 was also supported. As a result, hypothesis 3 was supported. This means the Japanese coffee shop market can be more precisely approximated using panel data than self-answered data.

Conclusion

This study aims to reveal the research method that will precisely evaluate visit frequencies of Japanese competitive coffee shops. In particular, this study answers two central questions. The first question is whether the Japanese coffee shop market can be approximated using the NBD Dirichlet model when we use self-answered data collected from consumers' questionnaires as well as when we use panel data gathered from consumers' receipts. The second question is whether the Japanese coffee shop market can be more precisely approximated using panel data than self-answered data. As a result, this study reveals the following conclusions.

In Study1, we confirmed that the Japanese coffee shop market can be approximated using the NBD Dirichlet model when we analyse self-answered data from consumers' questionnaires. This result was accordance with the results of Heiens et al. (2014).

According to Study 2, we found that the Japanese coffee shop market can be approximated with the NBD Dirichlet model when we analyse panel data based on consumers' receipts. Therefore, we find that both research methods are useful for comparing visit frequencies among Japanese coffee shops. However, we should not directly compare the values of visit frequencies among Japanese coffee shops but should regard these values as values influenced by market penetration. Marketers of coffee shops should eliminate the influence of market penetration when making comparisons.

Based on the results of Study 3, we found that the Japanese coffee shop market can be more precisely approximated using panel data than self-answered data. This result was consistent with the results of Wright et al. (2002).

Limitations and Further Research

This study has two limitations. First, we might not be able to collect all receipts of individuals in Study 2. As mentioned above, 440 individuals were asked to hand in to the research company their

receipts from the coffee shops they actually visited. However, we could not distinguish individuals who did not submit receipts from individuals who did not visit coffee shops. The collected receipts might be slightly different from the visit experiences.

Second, the results of this study may not be generalizable. This study specifically focuses on the Japanese coffee shop market. According to Heiens et al. (2014), there are some differences in the contexts of coffee shops in Kuwait compared to western countries. The Japanese coffee shop market may also be somewhat different. We should re-examine the results in various contexts.

Managerial Implications

This study has two managerial implications. First, the results suggest that marketers of coffee shops should not directly compare the values of visit frequencies among Japanese coffee shops but should regard these values as values influenced by market penetration. Coffee shop marketers should eliminate the influence of market penetration when making comparison.

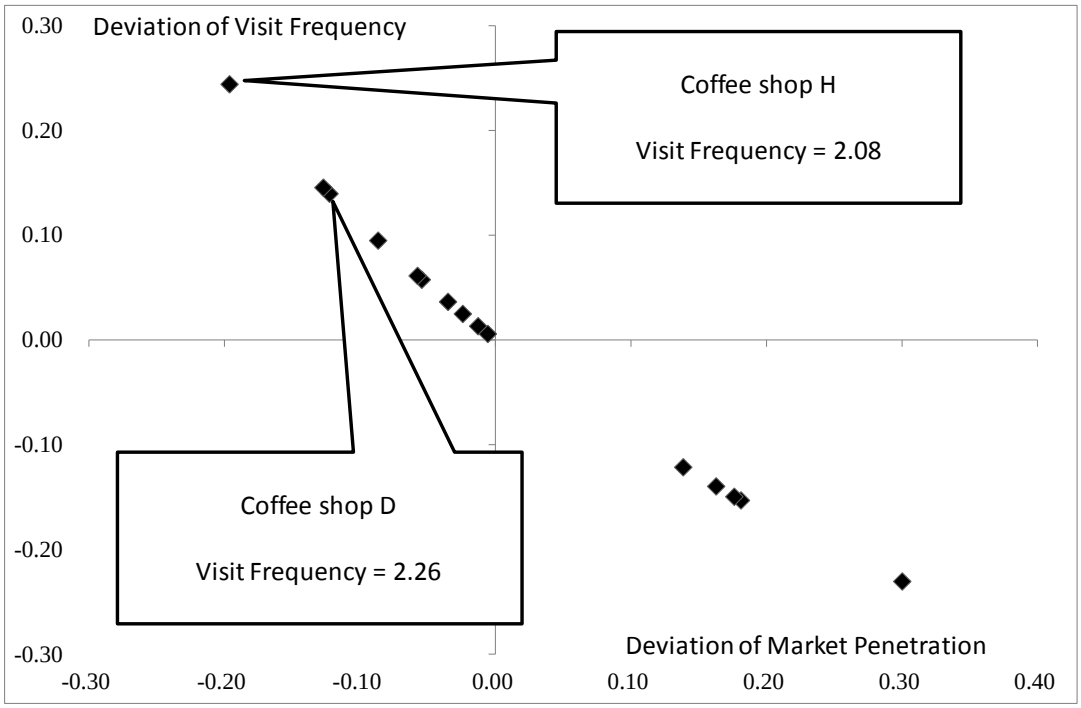


Figure 3: Deviation of visit frequency

Second, this study suggests that coffee shop marketers should recognise the market position of their own firm. For example, according to our research, the average visit frequency of coffee shop H was 2.08. On the other hand, the average visit frequency of coffee shop D was 2.26. Therefore, many marketers may believe that coffee shop H should improve its visit frequency. However, as Figure 3 shows, the deviation of visit frequency for Coffee shop H was higher than that for Coffee shop D. Therefore, coffee shop D's marketer should improve visit frequency.

Bibliography

Baker, B., McDonald, H. and Funk, D. (2016), "The Uniqueness of Sport: Testing against Marketing's Empirical Laws," *Sport Management Review*, 19(4), 378-390.

Barwise, T., and Ehrenberg, A. (1987), "The Liking and Viewing of Regular TV Series," *Journal of Consumer Research*, 14(1), 63-70.

Bennet, D. and Graham, C. (2010), "Is Loyalty Driving Growth for the Brand in Front? A

Two-Purchase Analysis of Car Category Dynamics in Thailand," *Journal of Strategic Marketing*, 18(7), 573-585.

Dawes, J. (2009), "Brand Loyalty in the UK Sportswear Market," *International Journal of Market Research*, 51(4), 449-463.

Ehrenberg, A. (1988), *Repeat-Buying: Theory and Applications, 2nd ed.*, Oxford University Press.

Ehrenberg, A., and Goodhardt, G. (2002), "Double Jeopardy Revisited, Again," *Marketing Research*, 14(1), 40-42.

Ehrenberg, A., Goodhardt, G. and Barwise, P. (1990), "Double Jeopardy Revisited," *Journal of Marketing*, 54, 82-91.

Goodhardt, G. J., Ehrenberg, A. S., and Chatfield, C. (1984), "The Dirichlet: A Comprehensive Model of Buying Behaviour," *Journal of the Royal Statistical Society*, Series A (General), 621-655.

Johnson, N., Kotz, S. and Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, John Wiley & Sons, Inc.

Johnson, N., Kotz, S. and Kemp, A. (1993), *Univariate Discrete Distributions 2nd ed.*, John Wiley & Sons, Inc.

Kearns, Z. (2009), "DIRICHLET VB.xls," *Marketing Bulletin*, 20, Technical Note 2.

Keng, A., Uncles, M., Ehrenberg, A. and Barnard, N. (1998), "Competitive Brand-Choice and Store-Choice among Japanese Consumers," *Journal of Product & Brand Management*, 7(6), 481-494.

Heiens, R. and Larry, P. (2014), "Double Jeopardy Patterns in a Middle Eastern Retail Services Setting," *Asian Journal of Business Research*, 4(1), 16-29.

McPhee, W. (1963), *Formal Theories of Mass Behavior*, The Free Press.

Meyer-Waarden, L., and Benavent, C. (2006), "The Impact of Loyalty Programmes on Repeat Purchase Behaviour," *Journal of Marketing Management*, 22(1-2), 61-88.

Reichheld, F. and Sasser, E. (1990), "Zero Defections: Quality Comes to Services," *Harvard Business Review*, 68(5), 105-111.

Rungie, C. and Goodhardt, G. (2004), "Calculation of Theoretical Brand Performance Measures from the Parameters of the Dirichlet Model," *Marketing Bulletin*, 15, Technical Note 2, 1-19.

Sharp, B., Wright, M. and Goodhardt, G. (2002), "Purchase Loyalty is Polarised into either Repertoire or Subscription Patterns," *Australasian Marketing Journal*, 10(3), 7-20.

Sharp, B. (2010), *How Brands Grow*, Oxford University Press.

Sudman, S. (1964), "On the Accuracy of Recording of Consumer Panels: 2," *Journal of Marketing Research*, 1, 69-83.

Uncles, M., Ehrenberg, A. and Hammond, K. (1995), "Patterns of Buyer Behavior: Regularities, Models and Extensions," *Marketing Science*, 14(3), G71-G78.

Wakuta, R. (2015), "External Validity of the NBD Dirichlet Model: A Case of Credit Card Services in Japan," *The Kyoto Gakuen University Review Faculty of Business Administration*, 24(2), 27-49. (Written in Japanese)

Wright, M., Sharp, A. and Sharp, B. (2002). "Market Statistics for the Dirichlet model: Using the Juster Scale to Replace Panel Data," *International Journal of Research in Marketing*, 19(1), 81-90.

Yang, Z., Bi Z. and Zhou, N. (2007), "The Double Jeopardy Phenomenon and the Mediating Effect of Brand Penetration between Advertising and Brand Loyalty," *Journal of Advertising Research*, 45(2), 211-221.