

## Application of Multiple Imputations to Freight Transportation Survey Data: A Case Study of Commodity Flow Survey

Terdruk Rongviriyapanich and Akachut Suppiyatrakul

Department of Civil Engineering,

King Mongkut's University of Technology, North Bangkok 1518,  
Piboonsongkram Rd., Bangsue Bangkok, 10800, Thailand

---

**Abstract: Problem statement:** Freight transportation data was indispensable input to transportation planning. In Thailand, efforts had been put to collect freight movement data by conducting road side survey and commodity flow survey. The results of these surveys did not produce consistent volume of shipment due to limited sampling coverage and non-response. Nevertheless, freight distribution patterns, which were derived from these surveys, were favorably consistent with each other.

**Approach:** The objective of this study was propose an approach to improving quality of the commodity flow survey data in terms of total shipment weight. Our scope of study was limited to consumer goods and food stuffs. Multiple imputations were performed to correct non-response. The shipment weight was again adjusted by taking into account of the probability of no shipment in a particular quarter. **Results:** Comparison between the adjusted weight and road side survey data showed that the discrepancies in total weight of significantly reduced. **Conclusion:** Total shipment weights of the CFS after the adjustments are compares to those of road side survey. Plausible result is obtained for the case of consumer goods, while that of food stuffs is still notably different.

---

**Key words:** Commodity flow survey, Multiple Imputation (MI), Missing Completely At Random (MCAR), shipment weight

---

### INTRODUCTION

Freight transportation plays an important role in economic development. It is used to move raw materials and products from origins to destinations. Efficient freight transportation management is a key success factor of country. Government is responsible to the development of infrastructure to promote efficiency and cost reduction in freight transportation. In order to build infrastructure in the most cost effective manner, public sector needs to analyze freight transportation demand and supply. Freight movement data is an indispensable input to transportation planning. Data collection of freight transportation, however, is particularly costly and time consuming. As a result, given budget limitations, we have to estimate the population parameters by using sample data.

Various survey methods may be employed to collect freight transportation data. For example,

Commodity Flow Survey (CFS) is an approach to collecting data from shippers, while road side interviews may be used to gather data from drivers. Each survey technique has its own advantages and drawbacks. CFS, for instance, offers rich detailed data on origin-destination and mode share, which cannot be easily obtained by other techniques. It is an indispensable input to truck origin-destination matrix estimation, which is very useful for freight transportation planning. However, it needs cohesive cooperation from respondents as the survey period is designed to cover all quarters in a year. Road-side survey, on the contrary, requires less cooperation from interviewees due to availability of weight in motions at the survey locations. Nonetheless, road side survey suffers from incomplete coverage of survey locations and double counting, which can be somehow corrected (Kuwahara and Sullivan, 1987).

---

**Corresponding Author:** Terdruk Rongviriyapanich, Department of Civil Engineering, King Mongkut's University of Technology, North Bangkok 1518, Piboonsongkram Rd., Bangsue Bangkok, 10800, Thailand Fax: +66-2-513-2500

In Thailand, related agencies have long relied on freight transportation database and model, which were established in 1990s (OECD, 1997). Changes in demography, economy and transportation network have made this model become obsolete. Thus, in recent years Department of Land Transport and National Statistical Office independently conducted freight movement surveys. Department of Land Transport (2009) opted to conduct road side interviews in 10 selected provinces, which are major economic activity centers of Thailand. Given budget limitations, the agency expected that these survey locations would cover a majority of freight movement volume in the country.

National Statistical Office, on the other hand, opted to conduct shipper survey, following the framework of CFS Survey 2002 in the U.S. (Bureau of Transportation Statistics, 2007). The survey's goals are to estimate the characteristics associated with the goods shipments. The sampling methodology is theoretically sound but the survey results turned out to be underestimated in comparison to other data sources (SSHC, 2008). This is not surprising, given that a considerable number of respondents reported that there was no outbound shipment activity for all survey quarters. Zmud (2006) also mentioned high level of unit non response in the 2002 CFS conducted in the United States.

It may be stated that, at present, freight survey data in Thailand need to be refined due to limited sampling coverage in the case of road side survey and non response in the case of CFS. The CFS data are selected as datasets in this study due to their relatively solid statistical framework and richness of the data. Hence, this study is aimed to investigate an approach to improving Thailand CFS data by correcting non-response. Parametric regression-based multiple imputation is chosen as it was found to be comparable to non parametric methods. We focused only on two commodity types namely food stuffs and consumer products due to low seasonality of the shipment volume. Lack of reliable freight transportation data makes it impossible to test validity of our approach. Anyway, comparison to the other data source should give us some insights into benefits of the proposed method.

The study starts with a brief summary of Thailand 2007 CFS data. Next comparison with other data source provides an evidence of needs for adjustment of the CFS data. Subsequent of the study

focus on multiple imputation as a tool for correcting non response. Finally, comparison between adjusted results with the road side survey data is discussed.

**Commodity flow survey in Thailand:** SSHC (2001) conducted Thailand Commodity Flow Survey to collect freight movement data, which are necessary for transportation infrastructure planning and development. The sampling methodology is based on that of the 2002 CFS conducted in the United States. The survey covered establishments in various industries including agriculture, mining, manufacturing, wholesale and retail trade. In order to avoid double counting, only outbound shipments were sampled from all target industries except agriculture. The reason is that it would be very costly to conduct survey at farms. Thus, agricultural products, such as rice or sugarcane, were sampled from inbound shipments of the establishments in agro-industrial sector. Population is stratified by geography, industry and size of establishment. Establishments in each stratum were sampled based on a three-stage sample design (i.e., establishments, reporting weeks and shipments). Shipment data collected in the survey include:

- Total number of shipments in the specified week
- Value and weight of shipments
- Type of product
- Origin - Destination
- Mode of transport

**Response rate and pattern:** CFS data can be classified into 2 categories namely response and non-response. Response is defined as data, of which respondent reports shipment in at least one quarter. Non-response is defined as data, of which respondent reports no shipment in all quarters. We opted to classify the data as non response due to the fact that it is very unlikely that an establishment having no shipment in any quarter throughout a year. Response rates vary with product types and survey periods. The response rate of foodstuffs is about 53%, while that of consumer goods is around just 18% as illustrate in Fig. 1. At present, non response has been treated as zero shipment, resulting in considerably low estimates of shipment weight as will be discussed later.

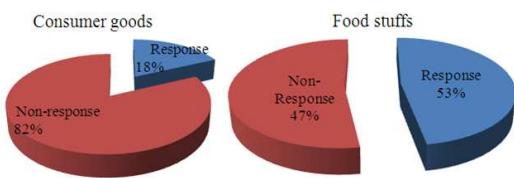


Fig. 1: Response rates of the CFS data

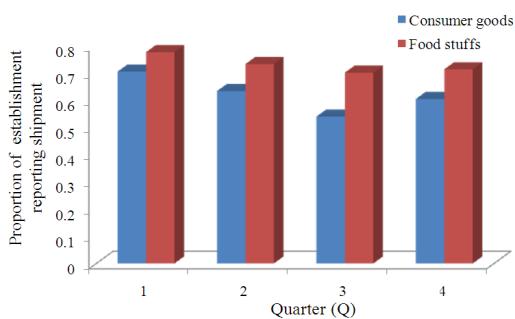


Fig. 2: Proportion of the response data reporting some shipment in each quarter

Table 1: Response pattern of the response data

No. of responded quarters	Quarter of shipments	Consumer goods		Foodstuffs	
		No. of establishments	%*	No. of establishments	%*
1	1	70	48	89	44
	2	28	19	42	21
	3	15	11	31	15
	4	32	22	41	20
2	12	31	38	41	32
	13	9	11	12	9
	14	14	17	19	15
	23	7	9	14	11
	24	7	9	7	5
	34	13	16	36	28
3	123	17	21	52	28
	124	24	29	54	29
	134	11	13	34	18
	234	30	37	48	25
4	all	140	100	462	100
	Total	448	982		

\*: Represents the percentage of total for each case of number of responded quarters

Table 2: Comparison of total shipment weight of the CFS with road side survey (Million ton)

Commodity	CFS	Road side survey
Foodstuffs	6.53	37.00
Consumer goods	0.86	10.13

Response data were further analyzed to examine response behavior of the respondents. The proportions

of establishments having shipment slightly reduced in quarter 2 and 3, then rebounded in quarter 4 as illustrated in Fig. 2. A longitudinal analysis of response behavior over the quarters shows that respondents are more likely to report shipment in quarter 1, as can be seen by the largest share of quarter 1 in every row of Table 1. Given low seasonality of the products, this pattern suggested that establishments tend to provide non response in the later quarters. Nonetheless, transition of states between response and non response should be reversible.

If response had always changed to non response irreversibly, the response rate would have dropped continuously. A rebound of response rate in quarter 4 indicated a possibility of transition from non-response to response.

**Comparison of the CFS with road side survey data:** The CFS data were compared to road side survey data in terms of total shipment weight and distance distribution. Lack of reliable source of secondary data in Thailand makes it impossible to validate the survey results. As a result, this comparison was done to empirically examine commonness and differences between these surveys, which can be used to define the assumption of this study.

**Weight:** the total shipment weight of foodstuffs and consumer goods of the CFS and road side survey as illustrated in Table 2. Although it is not unusual to see high variability in the results of this kind of survey, it should be apparent that estimates from the CFS data are underestimated. Under estimation is not surprising as non response has been treated as zero shipment and the response rates are considerably low as mentioned above. The comparison in Table 2 gives us some idea on how much the CFS data is underestimated.

**Distance distribution:** Distance distribution of the shipments, derived from the CFS and road side survey data and were compared in order to examine reliability of the results. Given that road side interviews were conducted in only 10 provinces, only CFS data in the corresponding provinces were used for the comparison. Road side survey data is believed to provide unbiased estimates of trip length distribution due to the fact that vehicles were randomly selected at the survey locations. Thus, it serves as a benchmark for testing distance distribution of the CFS data (Hirun and Sirisoponsilp, 2010).

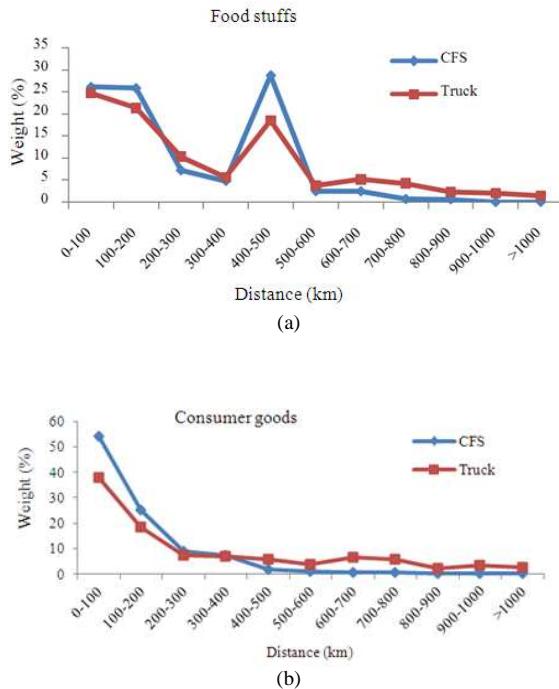


Fig. 3: Comparison of freight distribution of the CFS and roadside survey data (a) test statistic = 9.1,  $\chi^2_{(0.05, 10)} = 18.3$  (b) test statistic = 16.2  $\chi^2_{(0.05, 10)} = 18.3$

Statistical test, using chi-square test, was performed to see if these distributions are different. The trip length distributions derived from these surveys are not significantly different at confidence level of 95% as illustrated in Fig. 3. This suggested that, although being underestimated, the distance distribution of the CFS data is reasonable to some extent. As a result, amusingsness of the CFS data may be treated as random, irrespective of shipment distance.

The methodology of this study is summarized in Fig. 4. First, the CFS data are classified into 2 types namely response and non-response. Next, the response data are classified into 2 types of commodity, which are the focus of this study, namely foodstuffs and consumer goods.

Multiple imputations are performed based on parametric regression, in which macroeconomic data and establishment-specific data are used as independent variables. In the next step, no shipment pattern of response data is examined by using Markov chain process.

Probabilities of no shipment in each quarter are then used to adjust the original CFS and imputed

weights. Finally, total adjusted weights are obtained and compared to those of road side survey.

**Multiple imputations:** Multiple Imputation (MI) is a technique for handling data sets with missing values. It involves imputing  $m$  values for each missing item and creating  $m$  ( $m > 1$ ) complete datasets. Across these complete data sets, the observed values are the same, but the missing values are filled in with different imputations to reflect uncertainty levels (Little and Rubin, 2002; Landerman *et al.*, 1997). This approach retains the advantage of single imputation by using a complete data set and thus allowing standard statistical methods to be utilized. However, by allowing more than one value on a missing variable to be estimated, MI corrects for sampling variability and thus improves upon single imputation techniques. In addition, random error in the imputation process yields approximately unbiased estimates of all parameters, which no deterministic method can perform. Also, repeated imputation allows for good estimates of the standard errors. Rubin (1987) and Rubin and Schenker (1986) have demonstrated that even in extreme cases where the proportion of missing information constitute about a third of the data set, no more than 5 replicates of the model provides efficient estimates.

**Assumption:** Multiple imputations is performed under the following assumptions:

- Missing At Random (MAR), which requires that the cause of the missing data is unrelated to the missing values, but may be related to the observed values of other variables, is assumed in this study. Most imputation methods assume MAR as it is less restrictive than Missing Completely At Random (MCAR)
- MI assumes that the variables are jointly multivariate normal
- Parametric regression-based multiple imputation is chosen

MI obviously is an approximation, as few data sets have variables that are all continuous and unbounded, much less multivariate normal. Yet researchers have found it to study as well as more complicated alternatives specially designed for categorical or mixed data sets (Schafer, 1997; Rubin and Schenker, 1991).

**Regression analysis:** In order to apply regression-based imputation, we have to determine the relationship between dependent variable and independent variables.

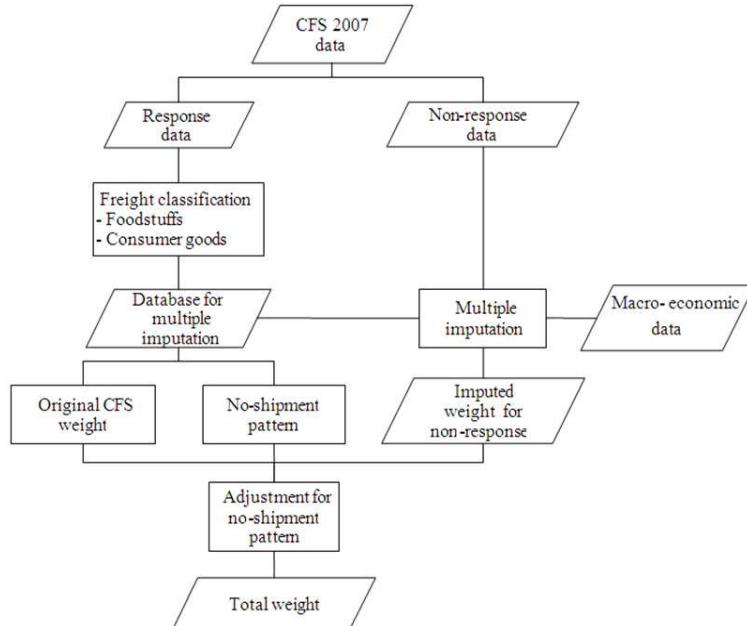


Fig. 4: Methodology

Shipment weight is the dependent variable, while independent variables must be chosen among the number of employees in the establishment or socio-economic data such as gross provincial product, population and income of the province, in which the establishment is located. Regression analysis yielded the relationships between dependent and independent variables as follows in Eq. 1:

$$(Y) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) \quad (1)$$

Where:

- $Y_c, Y_f$  = Shipment weights of consumer goods foodstuffs
- $X_1$  = Number of employees in the establishment
- $X_2$  = Population of the province in which the establishment is located
- $\beta, \theta$  = Regression parameters

Log-log relationship was found to produce better results compared to the linear one. One of its advantages is it always produces positive weight, while linear relationship may result in negative value. In addition, log-log relationship is also more robust to cope with wider range of data comparing to the linear form.

The multiple imputation results of consumer goods and foodstuffs as illustrate in Table 3. The imputation results are based on 10 complete datasets, which were used to determine the averages and standard errors. It should be noted that the probability of no shipment has not been taken into account in the multiple imputation.

**Adjustment for no shipment pattern:** As we mentioned earlier that a considerable number of respondents reported that there was no outbound shipment from their establishments in the specified week. The respondents, who are not willing to provide data on their shipments, may intentionally state that there was no shipment in that period. However, some of them may be true as it is possible that no shipment is made in a whole week. Since we have no reliable information to cross check the validity of the answer to this question, it is necessary to make certain assumption on the probability of no shipment in a given period.

We believe that the possibility of an establishment having no shipment in all quarters should be very low. Hence, in this study it is assumed that such data are non-response. In order to correct the non-response, we need to examine probability of no shipment of the response, which are defined by an establishment reporting shipments in at least one quarter.

Table 3: Multiple Imputation result (Million ton)

Quarter	Consumer goods		Foodstuffs	
	Mean	S.E.	Mean	S.E.
1	2.81	1.21	5.09	1.32
2	3.94	2.17	3.4	0.86
3	2.21	0.97	4.02	0.83
4	3.61	1.20	4.35	1.35
Total	12.57	2.11	16.86	2.18

Some of these establishments may report shipment in all quarters, while others may report no shipment in some quarters. Accordingly, we shall extract pattern of the probability of no shipment from these establishments.

Shipment pattern of an establishment in a given quarter is assumed to follow a stationary Markov chain process (Breiman, 1986). Three states are assumed to exist, one of them is having some shipments,  $S_{t,i} = 1$  and the others are having no shipment,  $S_{t,i} = 0$  and non-response,  $S_{t,i} = 2$ . Initial probabilities of these states are assumed to equal the observed data in the first quarter. Initial probability of state 1,  $p(1)$ , is set to equal to a proportion of establishments having shipments in quarter 1. Initial probability of state 0,  $p(0)$ , is set to equal to a proportion of establishments having no shipment in quarter 1. Since we have no reliable information on non-response, initial probability of non response in the first quarter,  $p(2)$ , is assumed to be zero. Due to relatively low seasonality of consumer goods and food stuffs in Thailand (Big C Super Center, 2010), it is assumed for sake of simplicity that the number of establishments reporting no shipment in each quarter should be identical. Nevertheless, observed data indicated that the number of establishments reporting no shipment grows in successive quarters compared to that of quarter 1. The increase should be attributed to the propensity of false reporting of no shipment to avoid prolonged survey procedures.

Transition probabilities between the three states are assumed to be time independent and may be given as follows Eq. 2:

$$P = \begin{Bmatrix} P_{0 \rightarrow 0} & P_{0 \rightarrow 1} & P_{0 \rightarrow 2} \\ P_{0 \rightarrow 0} & P_{1 \rightarrow 1} & P_{1 \rightarrow 2} \\ P_{2 \rightarrow 0} & P_{2 \rightarrow 1} & P_{2 \rightarrow 2} \end{Bmatrix} \quad (2)$$

The probability of each state in quarter  $t+1$  can be expressed as follows Eq. 3-5:

$$\begin{aligned} (S_{t+1,i} = 0) &= p(S_{t,i} = 0) * p_{0 \rightarrow 0} \\ &+ p(S_{t,i} = 1) * p_{1 \rightarrow 0} + p(S_{t,i} = 2) * p_{2 \rightarrow 0} \end{aligned} \quad (3)$$

$$\begin{aligned} (S_{t+1,i} = 1) &= p(S_{t,i} = 0) * P_{0 \rightarrow 1} \\ &+ p(S_{t,i} = 1) * p_{1 \rightarrow 1} + p(S_{t,i} = 2) * p_{2 \rightarrow 1} \end{aligned} \quad (4)$$

$$\begin{aligned} (S_{t+1,i} = 2) &= p(S_{t,i} = 0) * P_{0 \rightarrow 2} \\ &+ p(S_{t,i} = 1) * p_{1 \rightarrow 2} + p(S_{t,i} = 2) * p_{2 \rightarrow 2} \end{aligned} \quad (5)$$

Where:

- $P$  = Transition probability matrix
- $(S_{t,i} = 0)$  = Probability of establishment  $i$  having no shipment in quarter  $t$
- $(S_{t,i} = 1)$  = Probability of establishment  $i$  having some shipments in quarter  $t$
- $(S_{t,i} = 2)$  = Probability of establishment  $i$  false report no shipment in quarter  $t$

Time independent transition probabilities between the three states can be estimated from the observed data by using Eq. 6 maximum likelihood estimation (Zamani and Ismail, 2010):

$$\text{Max } L = \prod_{t=2}^4 \prod_{i=1}^4 P_{t,i} \quad (6)$$

Taking logarithm to the above Eq. 7, we obtained:

$$\text{Max } \ln(L) = \sum_{t=2}^4 \left( \sum_{i=1}^N \ln(p_{t,i}) \right) \quad (7)$$

The right side of (7) can be rearranged as follow Eq. 8:

$$\sum_{t=2}^4 [H_t \cdot \ln[P(S_{t,i} = 1)] + NH_t \cdot \ln[p(S_{t,i} = 0)] + (NH_t - NH_1) \cdot \ln[p(S_{t,i} = 2)]] \quad (8)$$

Where:

- $L$  = Likelihood function
- $p_{t,i}$  = Probability of a state of establishment  $i$  in quarter  $t$
- $t$  = quarter (including quarter 2-4)
- $N$  = total number of responses
- $H_t$  = number of establishments reporting some shipment in quarter  $t$
- $NH_t$  = number of establishments reporting no shipment in quarter  $t$

In addition, we have 3 constraint Eq. 9:

$$\sum_{j=0}^2 p_{i \rightarrow j} = 1 \text{ for } i = 0, 1, 2 \quad (9)$$

Thus, the objective function may be rewritten as Eq. 10:

$$\ln(L) - \left( \sum_{j=0}^2 P_{i \rightarrow j} - 1 \right) \quad (10)$$

Taking derivatives with respect to  $P_{i \rightarrow j}$ , we obtain Eq. 11-12:

$$\frac{n_{i,j}}{P_{i \rightarrow j}} - \lambda_i = 0 \quad (11)$$

$$P_{i \rightarrow j} = \frac{n_{i,j}}{\lambda_i} \quad (12)$$

Lagrange multiplier,  $\lambda_i$ , is given by Eq. 13:

$$\lambda_i = \sum_{j=0}^2 n_{ij} \quad (13)$$

Obviously,  $P_{i \rightarrow j} \propto n_{ij}$ . Transition probabilities, obtained from the above formulation, can be summarized as follows Eq. 14:

$$P_f = \begin{pmatrix} 0.351 & 0.607 & 0.042 \\ 0.186 & 0.763 & 0.051 \\ 0.186 & 0.436 & 0.378 \end{pmatrix}, P_c = \begin{pmatrix} 0.336 & 0.587 & 0.077 \\ 0.278 & 0.627 & 0.095 \\ 0.273 & 0.306 & 0.421 \end{pmatrix} \quad (14)$$

Transition probabilities of food stuffs,  $P_f$  and consumer goods,  $P_c$ , exhibit a comparable pattern. Probability of staying in no shipment state is approximately equivalent to one third in both cases. As we earlier expected, possibility of an establishment having no shipment in all quarters is very low; equivalent to just around 1%. Probability of an establishment having shipments in two consecutive quarters is in a range of 0.60-0.76, implying that cooperative respondents are more likely to consistently provide data. On the other hand, non-cooperative respondents are also more likely to be non response repetitively as reflected by a probability of 0.38-0.42. Another interesting result is that the probability of going into state 2, which is having no shipment, is equivalent either for state 0 or 1, which is very intuitive.

Adjustment of shipment weight for both response and non response data is necessary in order to take into account of false reporting. Response data need an adjustment due to our assumption that a number of establishments false reporting no shipment in quarter 2-4. The probabilities of non response in these quarters were used to adjust shipment weights. Adjusted shipment weights of the response data are summarized in Table 4. Shipment weight in quarter 1 remains unchanged since we assumed that the data in this quarter is free of non response, while those in the remaining quarters marginally increase.

Shipment weight imputation was obtained by imputing weight for all establishments. Thus, the results need an adjustment to take into account of the probabilities of no shipment. Adjusted shipment weight in each quarter is summarized in Table 5. After the adjustment, imputed weights reduced by 20-30%.

Finally, total shipment weights of the two products, as shown in Table 6, were compared with the road side survey data. It should be mentioned that the CFS survey was conducted one year prior to the road side survey. So it should not be surprising that the results would be somehow different. The total shipment weights of consumer products are almost equivalent, while that of food stuffs differ at an extent of 33% this result is attributed to the fact that response rate of consumer goods is very low. Once non response has been taken into account, shipment weight considerably increases. One of advantages of multiple imputation technique is an ability to determine standard error of the estimates.

Table 4: Adjustment for no shipment: Response data (Million ton)

Quarter	Original response		Adjusted response	
	Consumer goods	Food stuffs	Consumer goods	Food stuffs
1	0.15	1.49	0.15	1.49
2	0.23	1.50	0.26	1.59
3	0.10	1.83	0.13	2.02
4	0.24	1.34	0.28	1.46
Total	0.72	6.17	0.82	6.57

Table 5: Adjustment for no shipment: Imputed data (Million ton)

Quarter	Original MI		Adjusted MI	
	Consumer goods	Food stuffs	Consumer goods	Food stuffs
1	2.97	6.59	2.09	5.12
2	4.17	4.91	2.94	3.81
3	2.32	5.85	1.63	4.55
4	3.84	5.69	2.71	4.42
Total	13.3	23.04	9.38	17.9

Table 6: Total shipment weight after adjust (Million ton)

Quarter	Consumer goods		Foodstuffs	
	Mean	S.E.	Mean	S.E.
1	2.25	0.85	6.62	1.03
2	3.20	1.56	5.41	0.66
3	1.76	0.71	6.57	0.63
4	2.99	0.90	5.88	1.03
Total	10.2	1.52	24.47	1.67

Given high variability of this kind of survey, it may be reasonable to state that the proposed approach helps enhance the CFS data quality by taking into account of non response.

## CONCLUSION

Commodity flow survey was conducted in Thailand in 2007 in order to collect freight movement data of the country. Initial expansion of the survey data produced underestimated shipment weight comparing to other data source. Nevertheless, trip length distribution of the CFS seemed to agree with that of road side survey. Investigating response rate and pattern leaded to a hypothesis that transition between response and non response should be reversible. Non response data, defined as a unit reporting no shipment in all quarters, were the target for a correction by performing multiple imputations. A parametric regression-based multiple imputation was employed to correct non response under the assumption of Missing At Random (MAR). Shipment weight was found to correlate with the number of employees in the establishment and population in province.

Transition probabilities between states are assumed to follow Markov chain process and can be estimated by the observed data under certain assumptions. Three states are assumed to exist, one of them is having some shipments and the others are having no shipment and non-response respectively. Non response is defined as an establishment, which false reports no shipment in a particular quarter. Given low seasonality of the products in this study, proportion of establishments having no shipment in each quarter is assumed to be constant. Transition probability matrix of food stuffs and consumer goods turned out to be quite similar. It also confirmed our prior expectation of rare possibility of an establishment having no shipment in all quarters.

Multiple imputation results need to be adjusted for no shipment, based on the probabilities which were derived previously. Adjusted multiple imputation

results reduced by 20-30% comparing to original. Under the same assumption, non response is also believed to mix in the response data. Consequently, by taking into account of non response, original shipment weights in each quarter of the CFS were also adjusted. Total shipment weights of the CFS after the adjustments were compared to those of road side survey. Plausible result was obtained for the case of consumer goods, while that of food stuffs was still notably different.

It should be noted that in this study we have not addressed the problem of under reporting, in which respondent intentionally understate the number of shipments and/or shipment weights. It was assumed that the number of shipments and shipment weights as reported by respondents are true. It is likely that the seemingly low estimates of total shipment weight of food stuffs may be attributed to the under reporting. Due to unavailability of secondary data for cross checking, this problem would be a challenging topic for future study.

## ACKNOWLEDGEMENT

This study was financially supported by a research grant from Faculty of Engineering, King Mongkut's University of Technology North Bangkok. We also would like to express our gratitude to National Statistical Office and Department of Land Transport for provision of the Thailand CFS and road side survey data.

## REFERENCES

- Big C Super Center, 2010. Financial Statement. Big C Supercenter Public Company Limited. <http://bigc.listedcompany.com>
- Breiman, L., 1986. Probability and Stochastic Processes: With a View toward Applications. 2nd Edn., Scientific Press, Palo Alto, Calif., ISBN: 0894260766, pp: 324.
- Bureau of Transportation Statistics, 2007. Commodity Flow Survey. Washington, D.C.
- Department of Land Transport, 2009. Truck Freight OD Survey, Thailand.
- Hirun, W. and S. Sirisoponsilp, 2010. Analysis of interregional commodity flows. *Am. J. Eng. Applied Sci.*, 3: 728-733. DOI: 10.3844/ajeassp.2010.728.733
- Kuwahara, M. and E.C. Sullivan, 1987. Estimating origin-destination matrices from roadside survey data. *Trans. Res. Part B: Methodol.*, 21: 233-248. DOI: 10.1016/0191-2615(87)90006-3

- Landerman, L.R., K.C. Land and C.F. Pieper, 1997. An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociol. Methods Res.*, 26: 3-33. DOI: 10.1177/0049124197026001001
- Little, R.J.A. and D.B. Rubin, 2002. Statistical Analysis with Missing Data. 2nd Edn., Wiley, New York, ISBN-10: 0471183865, pp: 381.
- OECD, 1997. Maritime Transport Statistics. OECD. <http://www.oecd.org>
- SSHC, 2001. National Statistical Office. 1st Edn., National Statistical Office, Office of the Prime Minister, Bangkok, ISBN-10: 9742387001, pp: 62.
- SSHC, 2008. The 2007 Commodity Flow Survey. 1st Edn., National Statistical Office, Bangkok, ISBN: 9741130082, pp: 366.
- Rubin, D.B. and N. Schenker, 1986. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Am. Stat. Assoc.*, 81: 366-374.
- Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. 99th Edn., John Wiley and Sons, New York. ISBN: 047108705X, pp: 258.
- Rubin, D.B. and N. Schenker, 1991. Multiple imputation in health-care databases: An overview and some applications. *Stat. Med.*, 10: 585-598. DOI: 10.1002/Sim.4780100410
- Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. 1st Edn., Chapman and Hall, London, ISBN: 0412040611, pp: 430.
- Zamani, H. and N. Ismail, 2010. Negative binomial-lindley distribution and its application. *J. Math. Stat.*, 6: 4-9. DOI: 10.3844/jmssp.2010.4.9
- Zmud, J.P., 2006. Commodity flow survey: Improving methods to enhance data quality and usefulness. Proceedings of the Commodity Flow Survey Conference, Jul. 7-8, National Academy of Sciences, Boston Massachusetts, pp: 98-112.