Effect of Recall Duration on Reporting of Household Expenditures

An Experimental Study in Ghana

Chris Scott and Ben Amenuvegbe
SDA Working Paper Series

Editorial Board

Chairman
Ismail Serageldin
Director, Occidental and Central Africa Department
World Bank

Members
Ramesh Chander, Statistical Adviser, World Bank
Dennis de Tray, Research Administrator, World Bank
Yves Franchet, Director General, Statistical Office of the European Communities
Ravi Kanbur, Editor of World Bank Economic Review and World Bank Research Observer and Senior Adviser, SDA Unit, World Bank
Gabriel Kariisa, Chief Economist, African Development Bank
F. J. C. Klinkenberg, Director, Directorate General for Development, Commission of European Communities
Jacques Loup, Coordinator of Assistance to Developing Countries, United Nations Development Programme
E. M. Morris-Hughes, Chief, Nutrition Planning, Assessment and Evaluation Service, Food and Agriculture Organization
F. Stephen O’Brien, Chief Economist, Africa Region, World Bank
Graham Pyatt, Professor of Economics, University of Warwick
Paul P. Streeten, Director, World Development Institute, Boston University
Victor E. Tokman, Director, Employment and Development Department, International Labour Office
R. van der Hoeven, Senior Adviser, United Nations Children’s Fund

Editor
Michel Noël
Chief, SDA Unit
World Bank

Managing Editors
Marco Ferroni
Senior Economist, SDA Unit
World Bank
Christiaan Grootaert
Senior Economist, SDA Unit
World Bank
Effect of Recall Duration on Reporting of Household Expenditures

An Experimental Study in Ghana

Chris Scott and Ben Amenuvegbe

The World Bank
Washington, D.C.
The findings, interpretations, and conclusions expressed in this paper are entirely those of the author and should not be attributed in any manner to the World Bank, to its affiliated organizations, or to members of its Board of Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of the data included in this publication and accepts no responsibility whatsoever for any consequence of their use.

In order to present the results of research with the least possible delay, the manuscript has not been edited in accordance with the procedures appropriate to formal printed texts, and the World Bank accepts no responsibility for errors.

The material in this publication is copyrighted. Requests for permission to reproduce portions of it should be sent to Director, Publications Department, at the address shown in the copyright notice above. The World Bank encourages dissemination of its work and will normally give permission promptly and, when the reproduction is for noncommercial purposes, without asking a fee. Permission to photocopy portions for classroom use is not required, though notification of such use having been made will be appreciated.

The complete backlist of publications from the World Bank is shown in the annual Index of Publications, which contains an alphabetical title list (with full ordering information) and indexes of subjects, authors, and countries and regions. The latest edition is available free of charge from Publications Sales Unit, Department F, The World Bank, 1818 H Street, N.W., Washington, D.C. 20433, U.S.A., or from Publications, The World Bank, 66 avenue d'Iéna, 75116 Paris, France.

ISSN 1014-739X

Christopher Scott is a private statistical consultant based in London, United Kingdom, and Ben Amenuvegbe is a statistical advisor for the Vitamin A Supplementation Trial in Northern Ghana.

Library of Congress Cataloging-in-Publication Data

Scott, Christopher, 1927--
Effect of recall duration on reporting of household expenditures: an experimental study in Ghana / Chris Scott and Ben Amenuvegbe.
 p. cm. — (Social dimensions of adjustment in Sub-Saharan Africa, ISSN 1014-739X)
Includes bibliographical references.
HD7067.8.G37 1990
339.47'09667—dc20 90-12403
CIP
Integration of social and poverty concerns in the structural adjustment process in Sub-Saharan Africa is a major driving force behind the design of the World Bank's adjustment lending program in the Region. To further the goal, the Social Dimensions of Adjustment (SDA) Project was launched in 1987, with the United Nations Development Programme and the African Development Bank as partners. Since then many other multilateral and bilateral agencies have supported the project financially as well as with advice. The task presents a formidable challenge because of the severity of economic and social constraints in Africa and the intrinsic difficulty of tracing the links between economic policies and social conditions and poverty. It is essential to have a continuous professional dialogue between all concerned parties, so that the best ideas get discussed by the best minds, and become, as quickly as possible, available for implementation by policymakers. This is the aim of the SDA working paper series.

To fulfill its mission, the SDA Project operates on different levels. Conceptually, contributions need to be made which advance our understanding of how the economic crisis in Africa on the one hand and the adjustment response on the other hand affect the living conditions of people. Empirically, major improvements are needed in our knowledge of the social dimensions of life in Africa, how they change, and whether all groups in society participate effectively in the process of economic development. Gaining this knowledge will demand new efforts in data collection and policy oriented analysis of these data. Most importantly, policy actions are needed in the short term to absorb undesirable side-shocks stemming from the adjustment process so that the poor and disadvantaged are not unduly hurt, and in the long term to ensure that these groups fully participate in the newly generated growth. The SDA Project's mandate is to operate, in a concerted way, in all three domains: concepts, data, actions. This working paper series will report progress and experience in all three areas. I encourage every reader's active participation in the series and the work it reports on. It is meant to be a forum not only for exchange of ideas but even more importantly to advance the cause of sustainable and equitable growth in Africa.
The SDA Project has been launched by the UNDP Regional Programme for Africa, the African Development Bank, and the World Bank in collaboration with other multilateral and bilateral agencies. The objective is to strengthen the capacity of governments in the Sub-Saharan African Region to integrate social dimensions in the design of their structural adjustment programs. The World Bank is the executing agency for the Project. Since the Project was launched in July 1987, 30 countries have formally requested to participate in the Project.

The Project aims to respond to the dual concern in countries for immediate action and for long-term institutional development. In particular, priority action programs are being implemented in parallel with efforts to strengthen the capacity of participating governments (a) to develop and maintain statistical data bases on the social dimensions of adjustment, (b) to carry out policy studies on the social dimensions of adjustment, and (c) to design and follow up social policies and poverty alleviation programs and projects in conjunction with future structural adjustment operations.

The working paper series “Social Dimensions of Adjustment in Sub-Saharan Africa” aims to disseminate in a quick and informal way the results and findings from the Project to policymakers in the countries and the international academic community of economists, statisticians, and planners, as well as the staff of the international agencies and donors associated with the Project. In the light of the three terrains of action of the Project, the working paper series consists of three subseries dealing with (a) surveys and statistics, (b) policy analysis, and (c) program design and implementation.

The Surveys and Statistics subseries focuses on the data collection efforts undertaken by the SDA Project. As such, it will report on experiences gained and methodological advances made in the undertaking of household and community surveys in the participating countries to ensure an effective cross-fertilization in the participating countries. The subseries would also include “model” working documents to aid in the implementation of surveys, such as manuals for interviewers, supervisors, data processors, and the like, as well as guidelines for the production of statistical abstracts and reports.

The Policy Analysis subseries will report on the analytical studies undertaken on the basis of both existing and newly collected data, on topics such as poverty, the labor market, health, education, nutrition and food security, the position of women, and other issues that are relevant for assessing the social dimensions of adjustment. The subseries will also contain papers that develop analytical methodologies suitable for use in African countries.

Another subseries, Program Design and Implementation, will report on the development of the conceptual framework and the policy agenda for the project. It will contain papers on issues pertaining to policy actions designed and undertaken in the context of the SDA Project in order to integrate the social dimensions into structural adjustment programs. This includes the priority action programs implemented in participating countries, as well as medium- and long-term poverty alleviation programs and efforts to integrate disadvantaged groups into the growth process. The focus will be on those design issues and experiences which have a wide relevance for other countries as well, such as issues of cost-effectiveness and ability to reach target groups.
Table of Contents

Executive Summary 1

1. Introduction and Background 2

2. Literature Review 3

3. Methodology of Investigation 5
   A. Study design 5
   B. Field work 7
   C. Data preparation and processing 8

4. Analysis of Results 9
   A. Main results 9
   B. Weekly and monthly cycles 10
   C. Seasonal variation 11

5. Factors Affecting Recall 13
   A. Time trends within interviewers and within respondents 13
   B. Alternatives to the recall lapse hypothesis 14
   C. Change of respondent 14
   D. Expenditure characteristics 15
   E. Respondent characteristics 16

6. Conclusions 17

Appendix: Schedule of field visits 18

References 19
Executive Summary

An experimental study was carried out in Ghana with a sample of 135 households each of which was interviewed 11 times at varying time intervals. Recall of expenditures on 13 frequently purchased items was compared according to the length of the recall period. Clear evidence of recall loss was seen, with average daily expenditures falling consistently in association with increasing length of recall period from 1 day to 1 week. There was little or no further fall in going to 2 weeks. For a 1 year recall period data were erratic.

Examining recall in the 1 day to 1 week range, weekly and monthly cyclic variations in expenditure were eliminated by re-weighting. Start-up bias was also ruled out, and in principle the recall periods were always bounded. In these circumstances the association of recall lapse with recall duration became even clearer with an average loss of 2.9% for every day added to the recall duration. Between different expenditure items the extent of the decline was associated positively with mean frequency of purchase of the item.

The authors favor the interpretation that the 1-day recall data are the most accurate and argue the urgent need for further studies to explore the limits of the recall effect and possible means of palliating it.
1. Introduction and Background

A crucial issue in the design of household budget surveys is the length of the recall period for reporting transactions. It is known that the longer the recall period the greater the likely recall error, but on the other side a longer period makes it possible to cover a larger sample of transactions for a given number of interviews, and hence for a given field cost. Every survey designer has to face this conflict and choose a solution believed to be optimal. The problem is particularly acute in third world surveys, where widespread illiteracy seems to rule out the only alternative, the diary method.

In most African surveys food expenditures, at least, have been recorded by interviewers visiting the household either every day or every second day.\(^1\) In Asia, recall periods of one week or one month have been more common (UNESCAP, 1985). The World Bank’s Living Standards Measurement Survey (LSMS) project, already applied in 8 developing countries, uses a fortnightly recall period, together with a 1-year estimate based on normative reporting.

These wide variations reflect the almost total absence of evidence for developing countries on the level of recall error and its relation to recall duration. Moreover, in balancing recall error against sampling error there is a similar gap in knowledge of the main factor affecting the latter: the regularity of consumer purchases in third world countries.

The range of variation found in the solutions adopted, from one day to one month, implies of course an approximately corresponding cost range, hence 30-fold. The practical importance of this issue is thus so extreme, in relation to costs of third world survey work, that the need to settle the question on an empirical basis seems almost self-evident. It can reasonably be argued that this is the most urgent design issue facing third world survey workers today.

This study arose from the needs of the Ghana Living Standards Survey (GLSS), a multi-purpose household survey sponsored by the World Bank and implemented by the Ghana Statistical Service, in which a key element is the estimation of detailed household expenditures. The GLSS design, worked out by the LSMS unit at the World Bank, called for a 2-week recall period for each household, a period chosen essentially for logistic motives rather than in response to any evidence of recall efficiency. The World Bank’s Social Dimensions of Adjustment Project is also currently proposing widespread use of the same design in African countries. The question of the viability of the 2-week recall thus becomes an urgent issue.

The present authors, already associated with the GLSS as consultants, prepared a minimal research proposal to test the issue and were able to obtain a research grant from the UK Overseas Development Administration to conduct the experiment in Ghana.\(^2\)

The design calls for an experimental survey, independent of the GLSS although sponsored and supported by the Ghana Statistical Service. Its essential purpose was to investigate recall efficiency as a function of the length of the recall period, for expenditures on common articles by African households.

Notes

1. In an exhaustive bibliography covering all household budget surveys conducted in Africa up to 1970, Scott (1971) found 146 out of 177 surveys in which daily recording was used for food expenditures at least.

2. We wish to thank the ODA for financial support and the Ghana Statistical Service for sponsoring the study and providing logistic support. The authors take sole responsibility for the conclusions expressed in this report.
2. Literature Review

There is fairly extensive literature on recall lapse in surveys. However, the number of experimental studies in which recall efficiency is compared for different recall durations is not very large. Obviously, the nature of the information to be recalled is a crucial factor. Confining attention to household expenditure/income/consumption, we have identified 15 studies in which recall periods of different durations have been compared experimentally, none of them recent. Most, though not all, of these report substantial memory loss with increasing duration, in some cases as much as 50% or more. A striking finding reported by Politz (1958) is worth citing as an example of the obstinacy of the recall problem: Politz arranged for interviewers to observe shoppers in a supermarket and record their purchases, then follow them home for an immediate interview. The number of items reported as purchased fell below the observation by 9% and the expenditure fell below the observation by 6%.

Among the studies mentioned above, just 6 relate to household expenditure/consumption/income in developing countries. Ghosh (1953) examined frequency of purchase for 9 common expenditure items and found a steadily falling average as the recall period went from 1 day to 1 week to 1 year. Mahalanobis & Sen (1954) compared reported expenditure for a weekly and a monthly recall period. The latter gave consistently lower results, with a greater discrepancy for items purchased less frequently (up to 30% for sugar). In the Ghana study (Ghana, 1961) daily recording was compared with weekly, the latter yielding 20% less by value. Dutta Roy & Mabey (1968), also in Ghana but using a more formal experimental design, found a 30% fall in expenditure in going from daily to weekly reporting for 4 commonly purchased food items. Scott (1971, 1972), using this same survey and comparing daily with weekly recall but analyzing frequency of purchase rather than amount of expenditure, found a decline of 15% for the same 4 items and a similar figure for the pooled group of 10 items analyzed. Note that this discrepancy between the effects for frequency and amount implies that heavier expenditures (within a given item) were less well remembered. In addition Scott reports a recall loss of 12% of transactions for the average of his 6 most frequently purchased items and 38% for the 4 least frequently purchased. In a similar experiment in Botswana, Scott (1971) reports a loss of 71% in going from daily to weekly recall; however, the implementation of this experiment was acknowledged to be unsatisfactory and it can be argued that the results should be set aside. Finally, Plewis (1972), reporting an experiment in Malawi, found no difference between daily and weekly reporting. However, in this experiment the weekly reporting period covered the same days as the 7 daily reporting periods for the same respondent, so that the two types of response are not independent; in addition, the sample was excessively small, with an 88% relative standard error for the estimate for daily reporting.

In the studies by Ghosh and Mahalanobis & Sen an attempt was made to check the errors against objective methods. Ghosh placed an observer in the local shops who had to record all purchases for a year. Mahalanobis & Sen tested recall of consumption of certain selected foods for 1-week and 1-month recall and checked responses by weighing the household’s stocks before and after. When this was found impractical because stocks were too large, the household was asked to set aside a much smaller stock from which all consumption was to be taken. Ghosh’s observational data were consistently lower, sometimes much lower, even than the 1-year recall reports. But considering the way in which they were obtained, it is difficult to have confidence in them as a reliable criterion. Mahalanobis and Sen found their check data agreeing on the whole more closely with the monthly (lower) figures than the weekly but there are anomalies in the figures and one can question whether households always obeyed the instruction to take their consumption only from the special stock. These writers, as well as (apparently) Zarkovich, appear to favor long recall periods and attribute the apparent recall decline to over-reporting on short periods. This may account for the rather consistent use of longer periods in Asian surveys, where daily interviewing is practically unknown.
However the evidence for this interpretation of the decline is decidedly weak.

The Ghana 1968 study and the Botswana study were poorly supervised and implemented and the Malawi study is inconclusive. The need for a more rigorous African study was clear.

Note

3. Methodology of Investigation

A. Study design

In most survey experiments comparing different methods of questioning, the different treatments have been applied to distinct samples of respondents. In the case of household expenditures this is an inefficient design because the recall effect has to be detected against the background of between-household variance, almost inevitably a high level of variation.

The Dutta Roy—Mabey experiment in Ghana was based on a proposal by one of the present authors for a design which does not seem to have been used before in research on recall. Scott suggested that all the recall period lengths be applied to each household: a succession of interviews should be conducted at varying intervals with the same household, with recall always based on the period “since my last visit”, and with the different intervals presented in a different, and randomized, order to different households. The starting dates should also be staggered across days of the week and days of the month to eliminate any bias from cyclic variations in household expenditure patterns. Later calculations based on the results of the experiment suggested that, for comparable sensitivity in detecting the main recall effect, a design using independent treatment groups would have required about 10 times as many households in the sample. A further advantage of the proposed design was that, since the recall periods ran consecutively for each household, all recall periods were bounded; that is, the start of each period was signalled by a previous interview and the end by the current interview. There is clear evidence from several studies (for example, Neter & Waksberg, 1965; Plewis, 1972) that unbounded recall can lead to serious telescoping error (mis-dating), with consequent over-reporting.

This design, already tried out in the Ghana 1968 and Botswana 1967-68 experiments (but in the Ghana case without the staggered starting dates), was adopted for the present study. In order to meet the above constraints and achieve a constant interviewer workload each day, a very tightly controlled interview plan had to be implemented. Before detailing the schedule of interviews, we need to give an overview of the experimental design.

The backbone of the study is a series of 9 interviews with each household spaced at intervals of 1 day, 2 days, 3 days and 7 days in such a way that the different interval lengths were presented, twice each, in approximately random order. For example, a given household might be allocated interviews on days 1, 4, 11, 12, 14, 17, 24, 25, 27. The intervals here are, in days: 3, 7, 1, 2, 3, 7, 1, 2. (The second half of the series always repeated the order of the first half.) At each interview except the first, the respondent was asked about expenditures “since my last visit”. The first interview was introduced mainly in order to provide a starting point for the first interval, hence ensuring bounded recall. This series of 9 interviews was called the X-interviews.

An additional objective was to test the various recall approaches used in the World Bank’s LSMS/SDA surveys, namely: a 2-week bounded recall question, a 1-year unbounded recall question, and a synthetic 1-year estimate based on three normative questions. Before presenting these questions, we introduce an important distinction.

Everyday experience suggests that when we are asked to recall repeatable events over a short period we attempt to answer factually, but if the period becomes too long we switch to normative reporting, that is, report our usual experience. A distinction can thus be made between factual and normative response. A similar distinction can be made between questions asked in a factual or normative mode—though it is clear that a factual question may lead to a normative response. Thus one may distinguish between ostensibly factual (OF) questions and explicitly normative (EN) questions, independent of the nature of the response actually obtained.
2-week OF question

The World Bank’s recommended question reads:

“How much has your household spent for ... since my last visit?”

For ease of interviewing, this is preceded by the filter question:

“Since my last visit have the members of your household spent anything on the following items?”

These two questions were transferred almost exactly into our experiment.

12-month OF question

The World Bank’s recommended 12-month OF question reads:

“How much did the household members spend on ... in the past 12 months?”

This question follows the two mentioned above in the World Bank questionnaire. For our experiment we modified this question slightly to give greater emphasis to the recall period, thus:

“I want you to look back 1 year to-day. From that day to this day, how many cedis have you and your household spent on ...?”

It should be mentioned that in the World Bank questionnaire this form of question was used only for purchases of clothing and domestic supplies. For all other items, including all of those chosen for our experiment, only the EN question set below was used for the 1-year response.

EN questions

The World Bank’s recommended set of three normative questions reads:

“During which months in the past 12 months did the members of your household buy ...?” [WRITE each month cited. CODE number of months.]

“During these months, how often did they usually buy ...?” [WRITE number of times and time unit. CODE number of times per month.]

“How much did they usually spend each time?”

These questions were used without modification in our experiment. The annual synthetic estimate is obtained by multiplying the three responses as coded.

As the experiment called for both factual and normative responses for the same items, it was important to avoid contamination: we did not want the respondent, or the interviewer, to use the factual responses as a source of norms on which to base normative long-term estimates. For this reason it was decided to place the long-period (12 months and 2 week) questions before the short period ones. A smaller risk, but still one to be avoided, is that the normative data obtained in the long-period estimate might influence the factual responses. For example, if the interviewer is having difficulty obtaining a response for a 7-day period (perhaps the person who made the purchases is absent and information is coming from another household member) he might refer to previous responses given to the two normative questions: “How often do you buy?” and “How much do you usually spend each time?” and work out a suitable response for 1 week. To prevent this, the best solution seemed to be to have the normative questions asked by a different person and the answers recorded on a questionnaire not made available to the short-term factual questioner.

These considerations led to the following design:

The 1st visit to each household is made by the supervisor, who explains the survey, obtains the household’s firm commitment to cooperate, establishes the list of household members, records some background information about the household, and asks the 12-month OF question for each item. This was called the S1 interview.

The 2nd visit is made by the supervisor again, 2 weeks after the 1st. This time he asks the filter question, the 2-week OF question and the three EN questions. This is the S2 interview.

The 3rd visit comes 1 day later. This is the first of the X-interviews. The supervisor accompanies the interviewer and introduces him to the household. The interviewer then asks about purchases since the supervisor’s visit of the day before. These data are not used in the main analysis, though they are of some interest in investigating start-up bias.

The 4th to 11th visits. Here the interviewer carries out the planned schedule of visits which involve two of each of the intervals: 1 day, 2 days, 3 days, 7 days. At each visit he asks about purchases “since my last visit”.

Thus the household is visited 11 times in all, twice by the supervisor and 9 times by the interviewer. The total duration covered by the factual bounded recall questions in each household comes to 40 days.

Calculations based on the similar Ghana experiment of 1968 suggested that a sample of between 100 and 150 households would suffice to yield significant evidence of a recall trend in the 1–7 day experiment
for the most frequently purchased items. It is clear, however, that the method would become much less efficient for rarer purchases. Moreover there was a need to limit the length of interviews both to reduce the burden on respondents and to keep down overall costs. For these reasons the study was deliberately limited to the 13 most frequently purchased items, based on GLSS results. Only two of these were non-food items. Table 1 shows the frequency of purchase and mean daily expenditure per household in our sample, for each of the selected items, based on the X-interviews. The items are listed in the order of their appearance in the questionnaires. Together these 13 items account for about one third of the total annual cash expenditures by households in Ghana (estimate from Ghana Living Standards Survey).

B. Field work

A sample of 144 households was identified and allocated among 4 field teams, each consisting of 1 supervisor and 3 interviewers. The 12 households for each interviewer were divided into 3 batches of 4 households. Each batch represented a day's interviewing by one interviewer, who would then return to the same batch after an interval of 1, 2, 3 or 7 days. The batches for one interviewer were started off one after another at 1-day intervals and the 3 interviewers of each team started one after another at 3-day intervals. Thus the starting points for the 9 batches allocated to one team were staggered day by day over a 9-day period. However, all 4 teams started this sequence on the same day.

The visits for any one interviewer were interleaved in such a way that the interviewer always had 4 households (1 batch) to interview each day over the whole 27 days of his program, with the exception of two rest days. A specimen schedule is shown in the Appendix.

The success of the experiment depended on very strict adherence to the time table. Interviewers were offered a 100% salary bonus payable at the end of the project if they completed all interviews on schedule. Absence on compassionate or health grounds was allowable only on condition that enough advance warning was given to the supervisor to allow him to arrange for substitution by a reserve interviewer. Four interviewers and one supervisor were trained and held in reserve for this purpose. Interviewers were also required to visit any given household at the same time of day on every visit, to the nearest hour. However, provision for failure to achieve this was made by requiring the interviewer to enter the time and date on each record page and allowing a space for separate recording of any purchases occurring in the over-run period, that is, that part (if any) of the actual interval which fell later than the planned interval by more than one hour. Amounts so recorded were transferred to the proper period prior to data processing. Finally, the interviewer's presence in the household at the right time and on the right day was checked through a clocking-in sheet kept always by the household. All field workers had to enter on this sheet the time and date at every visit to the household and this could be checked by the supervisor. Supervisors were required to be present in their area every day and at least one senior officer visited each area every day.

These very strict control measures were rather successful: only one interviewer-day was missed through field worker error (out of 324). The 4 households affected were dropped from the sample, as were 5 others which went absent or failed to cooperate. In all, 135 households provided the complete required record, a response rate of 93.7%. Over-run intervals were recorded in 4.7% of the sample interviews, but in more than half of these (2.4% of the sample) the over-run was 2 hours or less.

The research method is essentially model based and no attempt was made to provide a representative sample of the population. Two areas were selected purposively, an outer suburb of Accra and a rural area some

<table>
<thead>
<tr>
<th>Item</th>
<th>Frequency of purchase per day per household</th>
<th>Mean expenditure per day per household (cedis)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(1)</strong></td>
<td><strong>(2)</strong></td>
<td><strong>(3)</strong></td>
</tr>
<tr>
<td>Cassava</td>
<td>0.456</td>
<td>51</td>
</tr>
<tr>
<td>Beans</td>
<td>0.237</td>
<td>20</td>
</tr>
<tr>
<td>Garden egg</td>
<td>0.463</td>
<td>16</td>
</tr>
<tr>
<td>Plantains</td>
<td>0.470</td>
<td>31</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>0.848</td>
<td>32</td>
</tr>
<tr>
<td>Kenkey</td>
<td>0.689</td>
<td>73</td>
</tr>
<tr>
<td>Fish</td>
<td>0.933</td>
<td>186</td>
</tr>
<tr>
<td>Palm oil</td>
<td>0.370</td>
<td>26</td>
</tr>
<tr>
<td>Bread</td>
<td>0.656</td>
<td>53</td>
</tr>
<tr>
<td>Soap</td>
<td>0.493</td>
<td>34</td>
</tr>
<tr>
<td>Charcoal/firewood</td>
<td>0.467</td>
<td>33</td>
</tr>
<tr>
<td>Onion</td>
<td>0.733</td>
<td>20</td>
</tr>
<tr>
<td>Maize</td>
<td>0.170</td>
<td>71</td>
</tr>
</tbody>
</table>

*Column (2): Proportion of non-zero responses among the 1-day recall responses. This measure does not take account of the possibility of two purchases of the same item on the same day by one household.

*Column (3): Daily mean expenditure per respondent household over the 26 days covered by the 2nd-9th X-interviews. No adjustments made for possible recall error.*

Garden egg: Eggplant
Kenkey: Crushed maize paste
Maize: Covers maize in grain form only
20 km from Accra. Two teams were allocated to each of these; two separate blocks were chosen in the former and two neighboring villages in the latter. The 36 households required for a given team and location were selected in advance by the supervisor using a determinate procedure designed to spread the sample through the area and to eliminate subjective choice. Selected households were numbered with a chalk mark on the wall, or census numbers were used where they survived. All dates of visits were pre-entered on the appropriate questionnaires with the selected household number at the end of the training session. Each interviewer was also provided with a complete daily schedule of visits and each supervisor was given the complete schedule for all of his team.

Since the X questionnaire consisted essentially of only one question (repeated for 13 items), a single day’s training was sufficient for interviewers. Emphasis was placed on the need for discipline in maintaining the planned schedule. Supervisors were trained over a 2-day period, again with primary emphasis on rigid adherence to the time table.

Interviewers were locally recruited teachers. Supervisors were regular staff of the Ghana Statistical Service. A day’s work for an interviewer would be about one hour; on this basis the interviewers’ pay could be considered lavish. These conditions, together with the very close supervision and the brevity of the interview, must be considered exceptionally favorable to the production of high quality data. It would be reasonable to expect a lower error rate than in a full survey.

C. Data preparation and processing

The first step was the elimination of 9 households which did not supply the complete data set. This left 135 households for analysis.

The next step was to transfer all transactions reported in over-run periods into the interval where they would have fallen had the interview taken place at the scheduled time. This procedure corrects the analytic basis of the data (an interval classified as d days in length maintains an expected total expenditure proportional to d) at the expense of a slight distortion in the recall analysis (a few expenditures are classified with the wrong recall duration).

Some of the questions required coding. This was carried out, followed by data entry.

Preliminary visual examination of the data listings revealed two anomalous observations. Two households had reported a bulk purchase of maize from 50 to 100 times the daily mean. One fell in a 1-day recall interval, the other in a 2-day interval. Their inclusion would dominate the recall effect for maize and distort that for the all-items total. They have been replaced by zero (the modal value) in all the main analyses (they are retained only in Table 1 and in the evaluation of regularity of purchase of maize). No other observations come anywhere near this degree of eccentricity and no other such adjustments have been made.

Computer work was undertaken on the main-frame installation of the Regional Institute for Population Studies at the University of Ghana.

Notes

4. The word “factual” is used above to refer to the semantic intention, not to the accuracy of the response.

5. There is evidence from an analysis of the LSMS survey conducted in Côte d’Ivoire that such contamination occurs in the LSMS interview. (Grootaert & Scott, 1986.)

6. The inadequate randomization over days of the week noted in Section 8 below could have been overcome most easily had this latter feature been modified, for example by staggering the teams’ starting dates at 1-day intervals.
4. Analysis of Results

A. Main results

Analysis of the main recall effect is based on the simple principle that, for any expenditure item, in the absence of recall error, average daily expenditure should be independent of the length of the recall period—or equivalently, expenditure over a recall period of length \( d \) should be proportional to \( d \). Thus, recall error will be indicated by the observation of a trend in average daily expenditure with increasing \( d \).

Common experience leads us to expect that, where recall error occurs, this trend will be downwards, and the error will be least for smallest \( d \). However it is not our intention to take the latter expectation for granted and the question of other interpretations will be discussed in a later section. Nevertheless, for convenience of presentation we shall take \( d = 1 \) as a baseline, corresponding to a recall index 100.

The principle cited above assumes that the observations have been properly randomized over all naturally occurring trends or cycles. We examine the validity of this assumption in later sections. Meanwhile we remind the reader that each of the 135 responding households contributes the same amount of data, namely 2 values for each of the recall periods 1 day, 2 days, 3 days and 7 days, plus 1 value for each of the other estimates: 14 days \( O_F \), 12 months \( O_F \) and 12 months \( E_N \)-composite estimate. One of the 13 expenditure items, plantain, was omitted from the questionnaire S2 through a printing error. This affects some of the comparisons possible at the all-items level. Table 2 shows the basic findings.

For a variety of reasons these results should not be taken at their face value. In the following sections we discuss these reasons and present adjusted results.

Before doing so we examine the statistical significance of some of the differences observed in Table 2.

After converting the raw data into daily mean expenditures we carried out an analysis of variance between the 4 recall periods 1 day, 2 days, 3 days, 7 days, regarded as 4 treatments, with 2 replicates for each observation. (The 1st and 2nd replicates differ in over-

Table 2. Mean Reported Daily Expenditure per Household for Various Recall Durations, Indexed to 1 day = 100

<table>
<thead>
<tr>
<th>Item</th>
<th>1 day</th>
<th>2 days</th>
<th>3 days</th>
<th>7 days</th>
<th>14 days</th>
<th>1 year</th>
<th>1 year</th>
<th>Base 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( O_F )</td>
<td>( E_N )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassava</td>
<td>100</td>
<td>120.00</td>
<td>116.00</td>
<td>107.00</td>
<td>91.00</td>
<td>145.00</td>
<td>95.00</td>
<td>50.52</td>
</tr>
<tr>
<td>Beans</td>
<td>100</td>
<td>94.00</td>
<td>100.00</td>
<td>98.00</td>
<td>68.00</td>
<td>148.00</td>
<td>108.00</td>
<td>18.96</td>
</tr>
<tr>
<td>Garden egg</td>
<td>100</td>
<td>108.00</td>
<td>95.00</td>
<td>86.00</td>
<td>80.00</td>
<td>118.00</td>
<td>98.00</td>
<td>15.52</td>
</tr>
<tr>
<td>Plantains</td>
<td>100</td>
<td>91.00</td>
<td>77.00</td>
<td>91.00</td>
<td>—</td>
<td>138.00</td>
<td>—</td>
<td>30.59</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>100</td>
<td>93.00</td>
<td>81.00</td>
<td>78.00</td>
<td>74.00</td>
<td>93.00</td>
<td>84.00</td>
<td>31.67</td>
</tr>
<tr>
<td>Kenkey</td>
<td>100</td>
<td>95.00</td>
<td>84.00</td>
<td>85.00</td>
<td>81.00</td>
<td>113.00</td>
<td>80.00</td>
<td>72.96</td>
</tr>
<tr>
<td>Fish</td>
<td>100</td>
<td>92.00</td>
<td>81.00</td>
<td>85.00</td>
<td>81.00</td>
<td>101.00</td>
<td>95.00</td>
<td>183.87</td>
</tr>
<tr>
<td>Palm oil</td>
<td>100</td>
<td>74.00</td>
<td>86.00</td>
<td>76.00</td>
<td>81.00</td>
<td>98.00</td>
<td>75.00</td>
<td>26.44</td>
</tr>
<tr>
<td>Bread</td>
<td>100</td>
<td>90.00</td>
<td>83.00</td>
<td>72.00</td>
<td>80.00</td>
<td>108.00</td>
<td>80.00</td>
<td>53.44</td>
</tr>
<tr>
<td>Soap</td>
<td>100</td>
<td>139.00</td>
<td>115.00</td>
<td>104.00</td>
<td>96.00</td>
<td>147.00</td>
<td>95.00</td>
<td>33.39</td>
</tr>
<tr>
<td>Charcoal/firewood</td>
<td>100</td>
<td>103.00</td>
<td>80.00</td>
<td>85.00</td>
<td>98.00</td>
<td>115.00</td>
<td>98.00</td>
<td>33.07</td>
</tr>
<tr>
<td>Onion</td>
<td>100</td>
<td>87.00</td>
<td>82.00</td>
<td>75.00</td>
<td>70.00</td>
<td>74.00</td>
<td>76.00</td>
<td>20.24</td>
</tr>
<tr>
<td>Maize</td>
<td>100</td>
<td>108.00</td>
<td>106.00</td>
<td>96.00</td>
<td>83.00</td>
<td>123.00</td>
<td>112.00</td>
<td>42.43</td>
</tr>
<tr>
<td>Total all items</td>
<td>100</td>
<td>98.21</td>
<td>89.20</td>
<td>87.10</td>
<td>—</td>
<td>113.76</td>
<td>—</td>
<td>613.10</td>
</tr>
<tr>
<td>Total without plantain</td>
<td>100</td>
<td>98.58</td>
<td>89.81</td>
<td>86.89</td>
<td>82.05</td>
<td>112.51</td>
<td>91.36</td>
<td>582.51</td>
</tr>
</tbody>
</table>

\( O_F \): Ostensibly factual answer form
\( E_N \): Explicitly normative question form, composite estimate
all mean by only 0.2%). Using the between-replicates estimate of variance in the denominator we tested the between-treatments variance with an F-ratio for each of the 13 items and for all items pooled. Table 3 gives the results.

The variances within treatments differ substantially. We therefore repeated the analysis at the all-item level after introducing weights inversely proportional to the standard deviation within treatments, as estimated from the 2 replications. The weights were:

<table>
<thead>
<tr>
<th>Duration</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>1.242</td>
</tr>
<tr>
<td>3</td>
<td>1.407</td>
</tr>
<tr>
<td>7</td>
<td>2.007</td>
</tr>
</tbody>
</table>

Re-running the ANOVA we obtained an increased value of F: 7.882 compared with the unweighted value 7.184.

These results leave no reasonable doubt that a recall effect is present, affecting most if not all of the items. The overall trend is clearly downwards, with increasing recall loss at increasing recall duration, at least up to 2 weeks.

B. Weekly and monthly cycles

Marked variations in expenditure are found according to day of the week. These show up most clearly, of course, in the 1-day recall data. Table 4 shows these data, separating out the 1-day recall dummy interview which preceded the main experimental series. A simple way of testing the significance of the day-of-week variation is to observe the correlation between the two columns of the table. This yields $r = 0.886$, which is significant at the level $P < 0.005$. Thus a consistent day-of-week variation is present, with high expenditures on Wednesdays and the weekend.

The authors had expected that their rough randomization of the interviewers' work schedules together with the staggered starting dates would suffice to produce a reasonably even distribution of days of the week between the 4 treatment categories. In the event however the randomization proved excessively lumpy: Table 5 shows the distribution of achieved interviews by day of week.

The most serious distortion here is the under-representation of Wednesday and Saturday (two high-spending days) among the 1-day recall interviews.

Note that the day of the week classification in Table 5 refers to the day of interview, not the day of expenditure. If we consider the days of expenditure which the respondent is trying to recall, obviously all 7 days appear with equal frequency in the 7-day column, while in the 1-day column they occur with widely varying frequencies. This effect vitiates comparability between the recall results for the four durations. To put all four durations on an equal footing we need to re-weight the data so that the 7 days of the week are

### Table 4. Variation in Expenditure by Day of Week (Data from 1-day Recall Period Only)

<table>
<thead>
<tr>
<th>Day</th>
<th>Dummy interview</th>
<th>Main survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>472</td>
<td>571</td>
</tr>
<tr>
<td>Tuesday</td>
<td>578</td>
<td>558</td>
</tr>
<tr>
<td>Wednesday</td>
<td>815</td>
<td>788</td>
</tr>
<tr>
<td>Thursday</td>
<td>526</td>
<td>557</td>
</tr>
<tr>
<td>Friday</td>
<td>559</td>
<td>528</td>
</tr>
<tr>
<td>Saturday</td>
<td>860</td>
<td>702</td>
</tr>
<tr>
<td>Sunday</td>
<td>711</td>
<td>697</td>
</tr>
<tr>
<td>Mean</td>
<td>675</td>
<td>613</td>
</tr>
</tbody>
</table>

### Table 3. Significance Test of Differences Between First 4 Columns of Table 2. Variance Ratio F.

<table>
<thead>
<tr>
<th>Item</th>
<th>$F(3; 942)$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassava</td>
<td>1.24</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>Beans</td>
<td>0.07</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>Garden egg</td>
<td>2.66</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Plantains</td>
<td>2.72</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>3.56</td>
<td>&lt; 0.025</td>
</tr>
<tr>
<td>Kenkey</td>
<td>4.44</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Fish</td>
<td>7.81</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Palm oil</td>
<td>3.07</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Bread</td>
<td>9.17</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Soap</td>
<td>6.78</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Charcoal/firewood</td>
<td>2.17</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>Onion</td>
<td>4.78</td>
<td>&lt; 0.005</td>
</tr>
<tr>
<td>Maize</td>
<td>0.42</td>
<td>&gt; 0.1</td>
</tr>
<tr>
<td>All</td>
<td>7.18</td>
<td>0.001</td>
</tr>
</tbody>
</table>

### Table 5. Distribution of Achieved Interviews by Day of Week and Recall Duration

<table>
<thead>
<tr>
<th>Day</th>
<th>1 day</th>
<th>2 days</th>
<th>3 days</th>
<th>7 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>59</td>
<td>30</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Tuesday</td>
<td>45</td>
<td>30</td>
<td>46</td>
<td>61</td>
</tr>
<tr>
<td>Wednesday</td>
<td>30</td>
<td>46</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>Thursday</td>
<td>47</td>
<td>45</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>Friday</td>
<td>31</td>
<td>30</td>
<td>45</td>
<td>58</td>
</tr>
<tr>
<td>Saturday</td>
<td>15</td>
<td>45</td>
<td>46</td>
<td>61</td>
</tr>
<tr>
<td>Sunday</td>
<td>43</td>
<td>44</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>All</td>
<td>270</td>
<td>270</td>
<td>270</td>
<td>270</td>
</tr>
</tbody>
</table>
Table 6. Mean Daily Expenditure per Household Adjusted for Within-week and Seasonal Variations, by Length of Recall Period Indexed to 1 day = 100

<table>
<thead>
<tr>
<th>Item</th>
<th>1 day</th>
<th>2 days</th>
<th>3 days</th>
<th>7 days</th>
<th>14 days OF</th>
<th>1 year OF</th>
<th>1 year EN</th>
<th>Base 100 in cedis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassava</td>
<td>100</td>
<td>110.00</td>
<td>105.00</td>
<td>91.00</td>
<td>83.00</td>
<td>112.00</td>
<td>74.00</td>
<td>55.75</td>
</tr>
<tr>
<td>Beans</td>
<td>100</td>
<td>81.00</td>
<td>92.00</td>
<td>89.00</td>
<td>60.00</td>
<td>117.00</td>
<td>85.00</td>
<td>21.44</td>
</tr>
<tr>
<td>Garden egg</td>
<td>100</td>
<td>104.00</td>
<td>90.00</td>
<td>80.00</td>
<td>76.00</td>
<td>88.00</td>
<td>73.00</td>
<td>16.33</td>
</tr>
<tr>
<td>Plantains</td>
<td>100</td>
<td>85.00</td>
<td>75.00</td>
<td>84.00</td>
<td>115.00</td>
<td>85.00</td>
<td>32.44</td>
<td></td>
</tr>
<tr>
<td>Tomatoes</td>
<td>100</td>
<td>88.00</td>
<td>78.00</td>
<td>73.00</td>
<td>117.00</td>
<td>74.00</td>
<td>59.00</td>
<td>33.11</td>
</tr>
<tr>
<td>Kenkey</td>
<td>100</td>
<td>96.00</td>
<td>82.00</td>
<td>84.00</td>
<td>76.00</td>
<td>94.00</td>
<td>66.00</td>
<td>72.60</td>
</tr>
<tr>
<td>Fish</td>
<td>100</td>
<td>86.00</td>
<td>75.00</td>
<td>77.00</td>
<td>76.00</td>
<td>85.00</td>
<td>32.44</td>
<td></td>
</tr>
<tr>
<td>Palm oil</td>
<td>100</td>
<td>70.00</td>
<td>84.00</td>
<td>74.00</td>
<td>77.00</td>
<td>83.00</td>
<td>64.00</td>
<td>27.94</td>
</tr>
<tr>
<td>Bread</td>
<td>100</td>
<td>89.00</td>
<td>84.00</td>
<td>70.00</td>
<td>80.00</td>
<td>108.00</td>
<td>80.00</td>
<td>53.50</td>
</tr>
<tr>
<td>Soap</td>
<td>100</td>
<td>137.00</td>
<td>111.00</td>
<td>99.00</td>
<td>94.00</td>
<td>151.00</td>
<td>97.00</td>
<td>34.16</td>
</tr>
<tr>
<td>Charcoal/firewood</td>
<td>100</td>
<td>104.00</td>
<td>80.00</td>
<td>80.00</td>
<td>96.00</td>
<td>103.00</td>
<td>89.00</td>
<td>33.74</td>
</tr>
<tr>
<td>Onion</td>
<td>100</td>
<td>87.00</td>
<td>80.00</td>
<td>72.00</td>
<td>69.00</td>
<td>69.00</td>
<td>71.00</td>
<td>20.49</td>
</tr>
<tr>
<td>Maize</td>
<td>100</td>
<td>104.00</td>
<td>103.00</td>
<td>91.00</td>
<td>77.00</td>
<td>77.00</td>
<td>69.00</td>
<td>46.12</td>
</tr>
<tr>
<td>Total all items</td>
<td>100</td>
<td>93.97</td>
<td>85.16</td>
<td>81.02</td>
<td>—</td>
<td>96.19</td>
<td>—</td>
<td>643.99</td>
</tr>
<tr>
<td>Total without plantain</td>
<td>100</td>
<td>94.44</td>
<td>85.72</td>
<td>80.85</td>
<td>78.16</td>
<td>95.19</td>
<td>77.29</td>
<td>611.55</td>
</tr>
</tbody>
</table>

Notes: The first 4 columns are re-weighted to equalize the frequency of days of the week of interview. All recall periods are shown indexed to the new re-weighted 1-day figure, taken as 100 (actual values shown at right). The two 1-year columns are adjusted for seasonal variation (see Chapter 4, Section C).

represented equally among the days of interview. Such adjusted data are shown in Table 6. The main effect of the re-weighting is to increase the 1-day recall figures; since the other periods are indexed to the 1-day period, they are in most cases reduced, leading to a somewhat steeper recall trend.

A strong monthly cycle, related to pay-day, was reported by Scott (1971) for Yaoundé in Cameroon. In our present data, the 1-day recall interviews are too thinly spread to show this clearly. In Figure 1 we have added together the 1-day and 2-day subsamples, dividing the reported 2-day expenditures equally between the two days covered. This procedure no doubt smooths down the peak somewhat but, in any case, it is not nearly as marked as in the Yaoundé case cited above. (Just under half of the household heads in our sample appear to be wage-earners.) The peak occurs on the weekend of 5–6 March, presumably because it was the first weekend after payday. Expenditures on these two days were 30% above the average for the rest of the period, or 24% above the average for the other weekends. There is a slight deficit of 1-day interviews covering these 2 days, although the other recall durations are satisfactorily represented. If we correct for this deficit, the 1-day mean rises by 0.6%. Since we are indexing to 1-day = 100, this means once again that the other figures fall slightly. Thus each of the figures in Table 6, excluding the extreme left and right hand columns, should be reduced by 0.6%. The extreme right hand column should be increased by 0.6%. We have not incorporated these small—and arguable—adjustments into the table.

C. Seasonal variation

The data for recall durations 1–7 days relate to the period from 20 February to 24 March 1988. The 14-day data relate to fortnights starting from 5 to 13 February. The 1-year data relate to the year preceding the start-up interviews, i.e. roughly mid-February 1987 to mid-February 1988.

Strictly, only the 1–7 day periods are comparable. The 14-day data can be compared with these without risk of serious error since the overall impact of the pay-day peak is very small. But the 1-year results are definitely not comparable because one cannot assume expenditure over the last year to be at the same rate as over any specific month.

It is possible to make some allowance for seasonal variation by reference to the data of the Ghana Living Standards Survey (GLSS). We have examined the first year's results of this survey, running from October 1987 to August 1988, for expenditures on a list of items as nearly as possible the same as the 13 items in our experiment. We compared mean daily expenditure (based on a 2-week recall question) for interviews done within the period of our experiment against the whole sample of interviews, covering October 1987–August 1988. This provides us with a set of seasonal correction factors to adjust the annual figures in Table 2 (both OF and EN) so as to make them comparable with the February-March reference of the shorter term figures (first 4 columns). We have the choice of using such a correction set based on the whole GLSS sample or based on the urban sample only.
We tried out both sets. Neither of them contributed anything towards reducing the wide variation between items seen in the "1 year OF" column. The variance of the figures in this column is marginally increased (by 10%) with the correction based on the GLSS urban sample only, and by 138% with correction based on the whole sample. This suggests (a) that the observed differences between items in the 1-year OF mode are not due to seasonal variation, and (b) that the urban GLSS sample is closer to our experimental sample than would be the whole GLSS sample. In Table 6 we have made the corrections based on the urban sample, for both of the 1-year estimates.

Discrepancy between the "current" data (1-7 days recall) and the 1-year retrospective results could arise from other sources than seasonal variation. For example, the means might differ because of changes in prices or changes in purchasing power between the periods concerned. In our view such discrepancies in the all-item means should not be taken very seriously: our experiment was not designed to investigate them. The main effect of interest in the 1-year data is the wide range of variation between items and the complete failure of seasonal adjustment to remove, or even reduce, this. At the very least, we have to conclude that questioning of this kind leads to very erratic results, at any rate for the frequently purchased items selected for this experiment.

Can it be argued against this that the OF and EN between-item variations agree rather closely? (The correlation between the two columns is as high as 0.806.) Doesn't this suggest that the variations are real? Unfortunately this argument carries little conviction. Observation of interviewers in the field suggests that the two questions are in practice asked in a rather similar manner. The OF question as initially put ("How much have you spent on ... in the past 12 months?") invariably meets with a baffled silence or "don't know". The interviewer is forced to rephrase it for a shorter period and multiply up. The main difference is that in the EN version this process is guided, formalized and more thorough. That the two methods vary in correlation with one another merely indicates that respondents have a fairly consistent perception of their own normal purchasing behavior for any given item, even if this does not agree with the behavior which they report in 1-day recall.

Note

7. We thank the Government Statistician for permission to use these data.

Figure 1. Time Variation in Mean Expenditures: Calendar Dates

Note: Data based on 1-day and 2-day recall only. 1-day recall allocated to day of interview. 2-day recall allocated half to day of interview, half to preceding day.
5. Factors Affecting Recall

A. Time trends within interviewers and within respondents

The data in this chapter are needed as evidence for the discussion on interpretation which follows.

Each of the 12 interviewers worked over a period of 29 days. Four of the interviewers started this sequence on Day 1, four on Day 4 and four on Day 7. In Figure 2, the results are lined up in relation to each interviewer’s work schedule so as to eliminate these staggered starting dates. Thus we look at any interviewer’s work from his 1st day to his 29th day, pooling the 12 interviewers. Where the recall period is d days we have divided the reported expenditure by d to get daily averages, then inflated this to correct for recall loss using the data from the penultimate line of Table 2.

The resulting graph answers the question whether interviewers themselves introduce any trend into the data. For example, if they report less expenditure each day due to increasing boredom with their task we will see a steadily falling graph.

The first 3 days represent the dummy interviews, discussed below. With these eliminated there is no sign of any systematic trend over time. (Rank correlation between day order and expenditure: \( r = -0.130 \). Not significant.)

We can ask a similar question about the respondent’s task. Each respondent household undergoes 9 interviews, covering a period of 26 days. Is there any systematic trend in the responses over this period, within respondents? A simple test of this, which avoids confounding with the effect of recall duration, is to compare the 1st and 2nd replicates for each recall duration within respondents. We have already mentioned that this difference averages only 0.2% of the daily mean. This is quite negligible and of course insignificant.

Up to this point we have ignored the dummy interviews. It will be remembered that these come just before the start of the main recall experiment. They constitute the first interview of each interviewer’s program and the question relates to purchases since the supervisor’s visit 24 hours earlier. Their results have been given in Table 4. They average 10% above the mean of the 1-day recall results from the main survey. However, after re-weighting to equalize the day-of-week frequencies this difference falls to 3%. Whether weighted or not the effect is not significant. Of course the dummy interviews do not constitute the start-up from the respondent’s point of view: they were in fact the 3rd interview with the respondent, following 1 day after the 2nd of 2 interviews by supervisors which were conducted at a 2-week interval. Thus it is perhaps not surprising that the dummies do not show a clear start-up effect. It is, however, possible that the 2-week data include a start-up effect.

A positive start-up bias is a common finding in household expenditure surveys. Scott (1971) gives detailed evidence from a survey in Yaoundé and mentions that discarding of the first few days’ results had been routine practice in many anglophone African surveys, following informal observations suggesting such a bias. Often the survey design makes it impossible to distinguish start-up bias from bias due to unbounded recall, because the first interview typically uses unbounded recall. Both are commonly believed to be positive. In order to measure start-up bias unambiguously one would have to start with an interview not using (and not foreshadowing the use of) expenditure/consumption questions and follow this with at least two expenditure/consumption interviews based on the same length of recall. If the first gave a systematically higher figure than the second we could attribute the difference to start-up bias unambiguously. In Scott (1971) these conditions are approximated in the Yaoundé survey analysis and there is enough evidence to show clearly the existence of a start-up effect. The effect, analyzed in detail in the reference, was found to vary between food and clothing, and between wage-earners and non-wage-earners.

One further observation relating to the start-up effect is discussed in Chapter 5, Section C.
B. Alternatives to the recall lapse hypothesis

As mentioned above, within interviewers there is no detectable time trend from the 2nd X-interview onwards. This finding, together with the fact that the order of the 4 durations (1, 2, 3 and 7 days) is randomized in the experiment, seems to rule out the counter-hypothesis that the apparent recall effect is actually a start-up effect or a fatigue effect. Moreover since all the periods run consecutively, the end of one being the start of the next, all recall periods are bounded. This would make it difficult to sustain the theory that there is over-reporting in the short periods: any such over-reporting would presumably imply double-counting rather than mere date shifts and there seems no reason why this should occur, or why it should be more prominent in shorter periods.

The possibility of over-reporting to impress the interviewer hardly seems to fit the circumstances of the present experiment, which is strictly limited to the 13 most commonly purchased items. These are not the kinds of items whose purchase brings prestige. Nor does it seem tenable to suppose that respondents became embarrassed by the number of "noes" they were saying and put in some imaginary purchases to please the interviewer. As we see from Table 1, among the 13 items, 10 show reported frequencies of purchase above 40% of all household days. Finally, if over-reporting were being stimulated by the interview situation one might reasonably expect this effect to fall off as the series of interviews continues. Yet no such trend is observed once the initial (dummy) interview is over.

C. Change of respondent

There is however a rather more subtle effect which might be invoked to account for the results and which requires more careful attention. While in principle the same respondent was interviewed 9 times over the X-interviews, it occasionally happened that this respondent was temporarily absent and another member of the household had to be used as informant. When such a change occurs the recall period for the new respondent is unbounded, if that start of the period is not signalled by the preceding interview with that respondent. This situation vitiates the claim that all recall periods are bounded. Could over-reporting in these interviews explain the apparent recall effect? Before examining this let us probe a little more deeply. When such a change of respondent appears in the record, the new name usually disappears one or two interviews later when the original respondent takes over again. Thus we observe two types of change of respondent: cases in which the change is to a new
respondent (i.e. one who has not responded before) and cases which change to an old respondent (one whose name has appeared earlier in the record as a respondent for the same household). For new respondents there should be not only an unbounded recall effect but also a start-up effect, while for the old respondents there should be an unbounded recall effect only.

Fortunately this problem was foreseen. Thus at the start of each X-interview the interviewer was required to write in the name and to code whether this was the same as, or different from, the respondent for the preceding interview. All cases coded as different were examined in the office and coded as new or old according to whether the same respondent had appeared before in the booklet of X-interviews for the household.

After re-weighting to equalize days of the week of interview this led to the results shown in Tables 7a and 7b.

Comparing the first line of Table 7a with the total line (which comes from Table 6) we see that the differences are small. Clearly the effect of changes of respondent accounts at most for a small part of the apparent recall effect. Table 7b is of interest parenthetically in that it shows clear evidence of an unbounded recall upward bias, of about 6%. On the other hand, there is no sign of a start-up bias (which would tend to make the second line higher than the third).

These various considerations suggest that the observed trend with recall duration is a genuine recall loss effect and that the 1-day recall duration is, as common sense suggests, the most accurate.

Referring to the last line of Table 6, the figures of the first 5 columns seem to suggest that, as we approach 7 days, the downward trend levels off. It is plausible to suppose that beyond 3 days or so, an increasing number of respondents turn to a normative response mode. Once this happens the length of the interval becomes irrelevant.

**D. Expenditure characteristics**

Does the different behavior of the 13 expenditure items throw any light on the underlying processes? For each of these items we have a number of characteristics which could potentially influence recall. These are:

1. Position in the questionnaire
2. Average amount spent on the item per household per day
3. Proportion of households purchasing item in a 1-day period
4. Mean expenditure per purchase (2 divided by 3)
5. Regularity of purchase, as measured by the ratio of the variances of the 7-day to the 1-day reports.

To relate these to recall error we need a measure of the recall effect. A simple and obvious measure would be the slope $b$ of the regression of the mean daily household purchase on recall duration. This is shown in Table 8 for the 13 items, based on data weighted to equalize the days of the week and covering the X-in-

### Table 7a. Mean daily expenditure adjusted for day of week of interview by length of recall period and type of respondent, indexed to 1 day = 100

<table>
<thead>
<tr>
<th>Respondent relative to preceding interview</th>
<th>Recall duration</th>
<th>Base 100 cedis</th>
<th>No. of interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 day</td>
<td>2 days</td>
<td>3 days</td>
</tr>
<tr>
<td>Same</td>
<td>100</td>
<td>96.32</td>
<td>86.59</td>
</tr>
<tr>
<td>Different, new</td>
<td>100</td>
<td>101.30</td>
<td>88.41</td>
</tr>
<tr>
<td>Different, not new</td>
<td>100</td>
<td>83.46</td>
<td>90.88</td>
</tr>
<tr>
<td>All</td>
<td>100</td>
<td>93.97</td>
<td>85.16</td>
</tr>
</tbody>
</table>

### Table 7b. Mean daily expenditure adjusted for day of week of interview: average across the 4 recall durations, by type of respondent

<table>
<thead>
<tr>
<th>Respondent relative to preceding interview</th>
<th>Average: days weighted equally</th>
<th>Average: interviews weighted equally</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cedis index</td>
<td>cedis index</td>
</tr>
<tr>
<td>Same</td>
<td>544.73 100.00</td>
<td>571.00 100.00</td>
</tr>
<tr>
<td>Different, new</td>
<td>577.03 105.93</td>
<td>605.67 106.07</td>
</tr>
<tr>
<td>Different, not new</td>
<td>590.06 108.32</td>
<td>606.68 106.25</td>
</tr>
</tbody>
</table>
Table 8. Recall Error Index: Percent Loss per Day (Adjusted for Within-week Variation)

<table>
<thead>
<tr>
<th>Item</th>
<th>Recall Error Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cassava</td>
<td>2.35</td>
</tr>
<tr>
<td>Beans</td>
<td>0.72</td>
</tr>
<tr>
<td>Garden egg</td>
<td>3.68</td>
</tr>
<tr>
<td>Plantains</td>
<td>1.64</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>3.95</td>
</tr>
<tr>
<td>Kenkey</td>
<td>2.51</td>
</tr>
<tr>
<td>Fish</td>
<td>3.01</td>
</tr>
<tr>
<td>Palm oil</td>
<td>2.78</td>
</tr>
<tr>
<td>Bread</td>
<td>4.52</td>
</tr>
<tr>
<td>Soap</td>
<td>2.50</td>
</tr>
<tr>
<td>Charcoal/firewood</td>
<td>3.56</td>
</tr>
<tr>
<td>Onion</td>
<td>3.96</td>
</tr>
<tr>
<td>Maize</td>
<td>1.97</td>
</tr>
<tr>
<td>All Items</td>
<td>2.89</td>
</tr>
</tbody>
</table>

Note: For each item the index \( b \) is based on the 4 observed daily means for 1 day, 2 days, 3 days, and 7 days, and represents their regression on 1, 2, 3, 7.

A plausible model for the findings of the present experiment would be to suppose that the recall effect is dominated by the proportion of people giving a normative response, that people tend to switch to normative response when the absolute number of events in the interval exceeds a certain threshold (perhaps 3 or 4?) and that normative responses tend to be underestimated because they overlook the occurrence of exceptionally large expenditures. There is a minor conflict between this hypothesis and the 1-year-OF data of Table 6, which are certainly normative yet do not show any systematic underestimation bias. However the comparability of the 1-year data with the X-interviews is so uncertain that this anomaly need not raise serious doubts.

E. Respondent characteristics

Finally, does an examination of recall variation in relation to respondent characteristics throw any further light on the interpretation? The following characteristics are available from the questionnaires:

1. Total household expenditure on the 13 items together
2. Per capita expenditure on the 13 items together
3. Household size
4. Urban/rural
5. Age of household head
6. Occupation of household head
7. Education of household head

Unfortunately, in seeking to maximize our chances of detecting a recall effect, we deliberately adopted a design which reduced the number of distinct households to a minimum for a given number of interviews. In addition, we were working within a very tight budget and therefore reduced the total sample to the minimum judged likely to yield significant evidence of a recall effect. These two factors imply the smallest possible sample size for detection of respondent differentials. With no more than 135 households in the sample it is not very surprising that no significant differentials appear in any of the above variables in relation to the recall index.

Note

8. Results for the remaining variables: \( r_{b1} = 0.349; r_{b2} = 0.001; r_{b4} = 0.331; r_{b5} = 0.027 \).
6. Conclusions

In summary, testing recall for a selection of frequently purchased items we find clear evidence of a substantial drop in reported expenditure with length of recall period, averaging roughly 3 percentage points for each day added to the period. There is no evidence to suggest that this trend represents anything other than recall lapse: all recall periods were (in principle) bounded and start-up bias was eliminated. It appears most reasonable to accept the common sense interpretation that the most accurate figure comes from the shortest recall period (1 day). On this interpretation the underestimation in using a weekly recall period is of the order of 20% for such frequent purchases. Note moreover that this figure relates to an abnormally well controlled field situation. There was considerable variation between expenditure items, with the largest error found for items purchased most frequently.

We have suggested tentatively the hypothesis that the recall curve is dominated by the proportion of respondents attempting a normative rather than a factual response. We suggest that in normative reporting exceptional expenditures tend to be overlooked; typically, exceptional means exceptionally high, so that this leads to underestimation. We also suggested that respondents tend to switch to normative reporting as the length of the recall period extends beyond a threshold level. The authors plan a fuller analysis of this effect for later publication.

These are largely speculations; confirmation will have to await experiments in which a broader range of items is covered, with more widely varying frequencies of purchase and a more representative sample.

Unless users are content with a highly unpredictable error rate likely to average 20% or more for frequent purchases, these results imply an urgent need for more intensive research to develop an improved methodology for third world household budget surveys.
Appendix

Schedule of field visits

The following schedule relates to Interviewer 1 of each team. Interviewer 2 followed an identical schedule starting 3 days later and Interviewer 3 a further 3 days later. The four teams started together on Day 1. Each X represents 4 households interviewed by one interviewer on the same day. The same 4 households were interviewed repeatedly and each column shows the schedule for one such set of households. The interviewer covered 3 sets of 4 households.

The schedules for the 2nd and 3rd interviewers of the same supervisor are identical in form but start on Day 4 and Day 7 respectively.

\[ S = \text{Supervisor interview (each } S \text{ represents 4 households)} \]

\[ F = \text{Free day.} \]


Distributors of World Bank Publications

ARGENTINA
Carmen Ranch, SRL
Cubierta Guzmán
Florida 165, 4th Floor-025
1330 Buenos Aires

AUSTRALIA, PAPUA NEW GUINEA, P.H. SOMOLINOS ISLANDS, VANUATU, AND WESTERN SAMOA
O.A. Books & Journals
448 Whitehorse Road
Mitcham 3132
Victoria

AUSTRIA
Golden & Co.
Gartenstr A-101 Vienna

BAHRAIN
Bahrain Research and Consultancy Association Ltd.
P.O. Box 22180
Manama 317

BANGLADESH
Micro Industries Development Assistance Society (MEDAS)
House 5, Road 16
Dhaka 1216

BRANCH OFFICE:
135, Par Ashad Sarak
Chittagong 4000

BELGIUM
Publications des Nations Unies
Av de la Rue 222
1000 Bruxelles

BRAZIL
Publicações Tecnológicas Internacionais Ltda.
Rua Ponte de Contum 209
00409 São Paulo SP

BRITISH COLUMBIA
Le Dufferin C.P. 81 1209 Ave. Aigleppe
Beauvoir, Quebec
V0J 1P6

CHINA
China Book & Periodicals Import Exp. Corp.
8, De P6 St Dong Jie
Utrecht

COLOMBIA
Falea Ltda.
Apartado Aeropostal 5270
Bogota D.E.

COTE D'IVOIRE
Centre d'Édition et de Diffusion Africaines (CEDA)
24 P.B. 154
Abidjan Abidjan

CYPRUS
MEMB Information Services
P.O. Box 2289
Nicosia

DENMARK
Samaritans Literature Society
Bjarre All 11
DK-190 Frederiksberg C

DOMINICAN REPUBLIC
Estudios Taller C. por A.
Inst. Nacional de Publicaciones
Riobamba 2100
Santo Domingo

EL SALVADOR
Foundation
Asociación de Enseña Technica Asociante
Avda. 12 de Noviembre, 10
La Union 1241
San Salvador

EGYPT, AAR REPUBLIC OF
Al Ahram
Al Lazat Street
Cairo

The Middle East Observer
4 Chapecot Street
Cairo

FINLAND
Abhandlungen Kriigelapessa
P.O. Box 128
SF-00120 Helsinki 10

FRANCE
World Bank Publications
46, avenue d'Argue
75017 Paris

GERMANY, FEDERAL REPUBLIC OF
UN-Dy-Verlag
Pappelallee 64-66
1000 Berlin 1

GREECE
IKIEME
26, Ippokration Street flats Naziax
Athens 11545

GUATEMALA
Libreria Platera Santa
Centro Cultural de la Plaza Santa
11 calle 6-50 zona 1
Guatemala City

HONG KONG, MACAO
Asia 2000 Ltd.
Mongkok Post Office
from Street No. 37
Mongkok, Kowloon
Hong Kong

HUNGARY
Kutaba
P.O. Box 199
1389 Budapest 62

INDIA
Allied Publishers Private Ltd.
751 Motcomb Road
Mumbai - 400 002

FRANCH OFFICE:
125 N. Hemipole Meg
Baltimore Town 807-038

13/14 And Ali Rd.
New Delhi 110 003

17 Chaundoin Avenue
Calcutta 700 009

Japenese Club Building
36 Mann Road Gangtanggr
Singapor - 54009

3-5-15 Rodolpho Cross Road
Hyderabad - 500 027

Jacona flare, 2nd Floor
Near Thakor Saug, Narsingpur
Ahmedabad - 380 009

Patricia House
46-A Akashnagar
London W2 2JL

INDONESIA
3, Jl. Jatujak Sari
B. Sentul Kelapa Gading
Jakarta 1576

IRELAND
TCD Publications
12 North Frederick Street
Dublin 1

ITALY
Llnea Commercianti Sanremo SPA
Via Serpotta Fortini, 190/10
Casale Pasticcio 522
30120 Moxere

JAPAN
Eastern Book Service
37-3, Shigeo 3-Cho, Bunkyo 531
Tokyo

KENYA
Africa Book Service (E.A.) Ltd.
P.O. Box 4294
Nairobi

KOREA, REPUBLIC OF
Fern Bank Book Corporation
P.O. Box 2, Yongchung-saen
Seoul

KOREA
MEMB Information Services
P.O. Box 5463

MALAYSIA
University of Malaya Compulsive Bookshop
P.O. Box 1177, Aman Pusat Baru
Kota Lumpur

MEXICO
INFOSIC
Apartado Postal 22-860
10000, Tepoztlan, Mexico D.F.

MOROCCO
Societe d'Etudes Marketing Marocain
12 rue Moulay, 6th d'Asfa
Casablanca

NETHERLANDS
Ilse Library and Information Service
P.O. Box 14
7200 BA Leiden

NEW ZEALAND
Hills Library and Information Services
Private Bag
New Market
Auckland

NIGERIA
University Press Limited
Three Crowns Building
Private Mail Bag 5005
Ibadan

NORWAY
Norske Informasjon Center
Bernard Norsenvei 2
P.O. Box 6225 Post
NO-0310 Oslo 6

OMAN
MEMB Information Services
P.O. Box 1512, Seeb Airport
Muscat

PAKISTAN
Mirza Book Agency
65, Mohan-e-Quaid-e-Azam
P.O. Box 799
Lahore 3

PERU
Editorial Desaparej SA
Apartado 950
Lima

PHILIPPINES
National Book Store
771 Real Avenue
P.O. Box 1954
Metro Manila

POLAND
URCA
Tule Kolebny Nivia
54-401 Wroclaw

PORTUGAL
Liberar Portugal
R. Dama Marco 70-74
1601 Lisbon

SAUDI ARABIA
Jami Book Store
P.O. Box 1146
Riyadh 1150

MEMBER INFORMATION SERVICES
Branch office:
Al Ada Street
Al Dawaar City
P.O. Box 718
Riyadh

SAUDI ARABIA
Hajj Ahbu Fatimah Building
King Fahed Street
P.O. Box 9369
Dammam

SOUTH AFRICA, BOTSWANA
For single orders:
Oxford University Press Southern Africa
P.O. Box 941
Cape Town 800

For subscription orders:
International Subscription Service
P.O. Box 6999
Croydon
London CR9 2PD

SPAIN
Mundi-Prensa Libros, S.A.
Casero 37
28007 Madrid

Liberas Internacionales AIDOS
Casero 37, 390
08009 Barcelona

SRI LANKA AND THE MALDIVES
Lake Street Bookshop
P.O. Box 244
50002 Colombo 8

Mauritius
Colombo 2

SWEDEN
For single sales:
Fritske Publikatet Sverige AB
Regentagatan 12, Box 14566
S-103 27 Stockholm

For subscription orders:
Litteraturbygget
Karlberg 37

TANZANIA
Oxford University Press
P.O. Box 5299
Dor en Saliem

THAILAND
Central Department Store
201-211 Siam Road
Bangkok

TRINIDAD & TOBAGO, ANTIGUA
BARBUDA, BARBADOS
DOMINICA, GRENADA, GUYANA,
JAMAICA, MONTserrat, ST.
KRITIKA & NEVIS, ST. LUCIA,
ST. VINCENT & GRENADINES
Systernatics Station Unit
P.O. Box 275
Brisbane

Distributors of World Bank Publications

UNITED ARAB EMIRATES
MEMBER Gulf Co.
P.O. Box 6007
Sharjah

UNITED KINGDOM
Mauritius Ltd.
P.O. Box 3
Alton, Hampshire GU34 2DG

URUGUAY
Instituto Nacional del Libro
Sobre 1116
Montevideo

VENEZUELA
Libraris del Instituto Aprios
Avenida 330
Caracas 100A

YUGOSLAVIA
Jugoslovenska Knjiga
P.O. Box 36
Trg Republike
YU-11000 Belgrade