Reclassifying Low-Expenditure Consumer Units in the Consumer Expenditure Interview Survey

Steven Bass

One of the primary uses of data from the Consumer Expenditure Survey (CE) is the computation of weights representing the purchases of goods and services in the construction of the Consumer Price Index (CPI), a principal Federal economic indicator. Accurate representation of actual expenditures is thus critically important beyond usual data quality standards. The CE processing system has several screening processes in place to ensure data quality. One such process is the minimal expenditure edit, which screens out consumer units (CUs) with unusually low reported total expenditures for further investigation, to determine whether they should be reclassified as noninterviews—that is, whether they should be treated as if they had refused to take part in the survey. Data from noninterviews are not used in the computation of official expenditure estimates from the CE. A minimal expenditure edit has been implemented for the CE Diary Survey since 2002. In April 2006, a minimal expenditure edit was implemented for the CE Interview Survey, to investigate cases with very low expenditures. This article describes the methodology of the minimal expenditure edit for the Interview Survey, as well as the results attained from its first year of implementation.

Background

Both Interview and Diary Survey data go through a series of edits before publication. Among these edits are consistency checks, outlier review, imputation, and weighting. Minimal expenditure edits for both surveys take place early in the production process, prior to CU weighting and any expenditure or income imputation. Although the minimal expenditure edit process for each survey is essentially the same, a number of differences exist because the Diary Survey is self-administered while the Interview Survey is administered by a field interviewer. For the Diary Survey, respondents record all their expenditures in a diary for two consecutive 1-week periods. It is difficult to ensure that a respondent has completed the Diary form accurately, including all of his or her expenditures, because no one is observing the process. For this reason, the minimal expenditure edit process for the Diary Survey is much more structured than that for the Interview Survey.

The Diary minimal expenditure edit process uses the number of expenditures recorded, the total amount recorded, and CU characteristics (such as the size of the CU) to determine whether a low report for total expenditures by a CU is legitimate. As part of the reclassification algorithm, urban CUs have to meet a higher expenditure threshold than rural CUs, because rural CUs are more likely to do their shop-

---

Steven Bass is an economist formerly working in the Division of Consumer Expenditure Surveys, U.S. Bureau of Labor Statistics.

1 See the glossary in Appendix: Description of the Consumer Expenditure Survey for the definition of a consumer unit.
ping less often. Students and small
CUs also are treated differently be-
cause of their lower expected expendi-
ture levels. In 2006, more than 800 of
the nearly 20,000 eligible diaries were
from CUs that were reclassified as
noninterviews. The process is entirely
automated, and none of the individual
CU reclassifications are manually re-
viewed.

In contrast to the Diary Survey,
which is a self-administered paper sur-
vey, the Interview Survey is a computer-
asisted personal interview (CAPI). Using
a laptop computer, the field inter-
viewer asks the respondent a series
of questions about his or her expendi-
tures and records the responses.2 To
reduce respondent burden, most sets
of related questions (such as those
having to do with telephone expenses)
are preceded by a screener question
(for example, “Have you received any
bills for telephone services?”). If the
respondent has not received any tele-
phone bills, the entire section can be
skipped by answering “No” to the ques-
tion. However, although a “No” answer
can help avoid unnecessary followup
questions, respondents may respond
“No” to screener questions (even
though they have applicable expendi-
tures) in order to skip over questions
and minimize the time it takes to com-
plete the interview. The minimal expend-
titure edit is meant to screen out such
invalid cases of low total expenditures.

The Interview minimal expenditure
edit process was intended to screen for
three separate potential problems:

- Field interviewers not asking all
  of the questions to certain groups
  of people (for example, skipping
  questions to students about
  owned properties).

- The total sum of expenditures is
  less than $100 or

- The total sum of expenditures is
  between $100 and $300, and the
  interview time is less than 15 min-
utes.

In the manual review of the CUs
screened out by the automated proc-
ess, other variables in addition to total
expenditures and interview time are
used to evaluate a case. These addi-
tional variables are related to expendi-
ture reporting characteristics (such as
the number of expenditures recorded
and the number of “Don’t know” or
“Refused” responses), respondent
characteristics (such as the respondent’s
age, the size of the CU, whether the
CU is in an urban or a rural location,
and whether the CU resides in public
housing or student housing), and data
collection characteristics (such as
whether the interview is a telephone
or personal interview and the number
of visits to the CU by the field inter-
viewer). A detailed record of all the
CU’s expenditures and any field inter-
viewer notes are taken into consider-
ation, as is information provided by the
CU in previous interviews.

CUs that are manually reviewed in
the minimal expenditure edit process
are presumptively treated as noninter-
views. The review process consists of
a search for mitigating factors that
would explain the low level of expendi-
tures for the quarter. Elderly respond-
ents, college students, recipients of
food stamps, and occupants of public
housing are almost always treated as
valid low expenditure cases, because
they tend to have lower expenditure
levels than the general population has.
If the respondent has a high number
of “Don’t know” or “Refused” re-
sponses, that is also taken as evidence
of a proper interview, because such re-
pondents still provide information on
the specific items purchased by a CU.
Expenditure amounts for “Don’t know”
and “Refused” responses are imputed
later. Often, field interviewer notes also
will provide valuable information, such
as expenses that have been paid for
by parents or other relatives. If no per-
suasive reason can be found to explain
the low level of expenses for the 3-
month recall period, the CU is reclassi-
fied as a noninterview and is excluded
from the computation of official esti-
mates from the CE.

Methodology

In the minimal expenditure edit process
for the Interview Survey, CUs are se-
lected by an automated procedure and
are manually reviewed on an individual
basis. Two factors—the length of the
interview and the total sum of expendi-
tures reported—are used to determine
whether a CU should be investigated.
In the computation of total expendi-
tures reported by a CU, “Don’t know”
or “Refused” responses are treated as
zeroes. Although some questions ask
for quarterly values while others ask
for monthly values, these different
reference periods are not standardized
to the same period for the purposes of
the minimal expenditure edit.

A CU’s records are manually re-
viewed in greater detail if:

- The total sum of expenditures
  is less than $100 or

- The total sum of expenditures is
  between $100 and $300, and the
  interview time is less than 15 min-
utes.

The Interview Survey minimal expend-
titure edit process was implemented in
2006. For that year, 49 CUs were reclassified as
noninterviews, out of a total of 257
flagged by the edit process. These
reclassified CUs differed significantly
from the general population of CUs.
(See table 1.) Specifically, reclassified
cases have much lower expenditure
totals, interview time, and reported in-
come than the general population has.
Although data can be collected either
in person or over the phone, in-person
interviews generally elicit higher qual-
ity data. Reclassified interviews are
more likely to have taken place over
the phone. Also, respondents are more
likely to have been converted refusals,
a term used for CUs that initially refuse
to participate in the survey. In addition,
the complete absence of any record usage (for example, credit card statements or receipts) shows a low level of diligence on the part of respondents in these CUs.

The differences between the reclassified interviews and those flagged by the process but not reclassified are enlightening. (See table 1.) Although the income level is still low, cases that are flagged but not reclassified have a much higher income level than the reclassified cases. They also have a higher rate of personal interviews and a lower rate of converted refusals. Students and residents of public housing are also among those who are less likely to be reclassified.

Although low annual income would seem to be a reasonable explanation for low expenditure totals, it is likely to be unrepresentative of true income in many cases. The income questions in the Interview Survey are asked at the end of the interview and are part of the Work Experience and Income section. Many respondents choose not to answer these questions fully. In the Interview Survey minimal expenditure edit process, annual income is used mostly as a proxy for respondent diligence. Beginning with 2004 data, the CE implemented an income imputation process to correct for the low response rate of income questions. The Interview Survey minimal expenditure edit is performed at an earlier stage of processing, and only the reported income data are used in the edit.

In addition to reclassifying CUs to noninterview status, the April 2006 minimal expenditure edit revealed that a field interviewer had been systematically falsifying data, and those falsified cases were removed from the database. These cases accounted for 5 of the 49 reclassified CUs. The expenditure data compiled for that edit also have been useful as data quality measures in other analyses. The number of expenditure questions answered, the number of “Don’t know” or “Refused” responses to expenditure questions, and the total reported expenditures before processing are variables that were created specifically for the Interview Survey minimal expenditure edit. However, these variables also serve as indicators of data quality and have been used in other recent research on the Interview Survey (such as comparing the quality of reporting between converted refusers and other respondents and comparing the quality of responses among different treatment groups in an incentive experiment). The data compiled by the edit process can be used to examine other issues as well. CUs with high expenditure totals and short interview times could be interpreted as evidence of inaccurate information provided by the respondent or of fraudulent data entered by the field interviewer. In addition, the data can be used to investigate the correlations between expenditures and other variables, such as participation in public housing or in food stamps programs.

**Conclusion**

Analysis of the effectiveness of the minimal expenditure edit process in the Interview Survey is still ongoing. So far, the low number of reclassified cases has had a negligible effect on the computation of official expenditure estimates from the CE. Further analysis is necessary to determine whether the thresholds used in the edit should be revised to increase the number of cases evaluated.

One consideration for future implementations of the Interview minimal expenditure edit is automation. Currently, cases are selected for manual review on the basis of a fixed set of criteria, but the ultimate decision to reclassify is left to the reviewer. A rule-based approach would increase the consistency of the edit by removing human error, in addition to saving the reviewer time, thereby allowing him or her to consider a larger number of cases. However, this change would come at the expense of flexibility. Experimentation with a more extensive set of criteria, as well as an analysis of the tradeoff between false positives and false negatives, is needed to determine whether the process should be modified to achieve the appropriate balance.
Table 1. Comparison of average expenditures, income, and characteristics, by type of interview, Consumer Expenditure Survey, second quarter 2006 to first quarter 2007

<table>
<thead>
<tr>
<th>Item</th>
<th>Good interviews (n = 32,554)</th>
<th>Reclassified cases (n = 49)</th>
<th>Flagged but not reclassified (n = 208)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure total</td>
<td>$8,542</td>
<td>$44</td>
<td>$42</td>
</tr>
<tr>
<td>Necessities total(^1)</td>
<td>$2,325</td>
<td>$18</td>
<td>$19</td>
</tr>
<tr>
<td>Total time(^2)</td>
<td>60 minutes</td>
<td>29 minutes</td>
<td>32 minutes</td>
</tr>
<tr>
<td>Annual income(^3)</td>
<td>$30,405</td>
<td>$45</td>
<td>$440</td>
</tr>
<tr>
<td>CU size</td>
<td>2.4 persons</td>
<td>1.7 persons</td>
<td>1.5 persons</td>
</tr>
<tr>
<td>Age of respondent</td>
<td>49</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Other statistics (percent):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal interviews(^4)</td>
<td>67.2</td>
<td>55.1</td>
<td>72.6</td>
</tr>
<tr>
<td>Converted refusal(^5)</td>
<td>11.7</td>
<td>38.8</td>
<td>20.7</td>
</tr>
<tr>
<td>Usage of records(^6)</td>
<td>47.4</td>
<td>0</td>
<td>3.4</td>
</tr>
<tr>
<td>Food stamps</td>
<td>4.7</td>
<td>4.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Student housing</td>
<td>1.1</td>
<td>12.2</td>
<td>16.8</td>
</tr>
<tr>
<td>Public housing</td>
<td>2.7</td>
<td>4.1</td>
<td>18.8</td>
</tr>
</tbody>
</table>

\(^1\) The necessities total statistic is the sum of the amounts spent on groceries, utilities, and housing payments—items for which almost all CUs should report expenditures.

\(^2\) Total time is the amount of time spent in the interview process.

\(^3\) Annual income reported here is lower than published income for two reasons. First, respondents who are unwilling to reveal their income levels have the option of selecting an income bracket instead. An income equal to the median of reported incomes inside the bracket selected is then imputed. Second, data are imputed for incomplete income reporters. Income data reported in this table are compiled prior to both bracket and income imputation.

\(^4\) The Interview Survey is conducted either in person or over the phone. In-person interviews are the preferred method, because they generally elicit higher quality data.

\(^5\) The designation converted refusal is selected at the discretion of the field representative if the respondent initially refused the survey, but was eventually convinced to participate.

\(^6\) In the survey assessment section, the field interviewer is asked whether the respondent “Always, almost always, mostly, occasionally, almost never, or never” used records (such as receipts or credit card statements) to answer the questions. The entry “usage of records” is the percentage of CUs classified in a category other than “never.”