Panel effects in consumer research: Statistical models for underreporting
Kaper, E.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
An explanatory model for underreporting in panel surveys

4.1 Introduction

The effects of response burden, estimated in Chapter 2, were rather small, in fact too small to explain the large decrease in reportings as shown in Table 2.1 on page 13. The table shows a large decrease in reported expenditure with time. In the reportings for eggs a peak can be seen in month 4. In 1994, Easter was in this month, and then the Dutch consume considerably more eggs. However, reportings are lower than those in the first month, which is very unlikely. Together with an increasing wave non-response and an increasing reporting of zeroes, the effect on reportings is so large and the observed process is so unlikely that one cannot use such data without correction. Several possibilities are available to make such corrections (Saris, 1996). One possibility is to select respondents who have shown a stable level of responding during a test period of some months, previous to the actual reporting period. Olivier (1987) studied the consequences of this procedure and found an overreporting of cheap products. One explanation for this is that due to the selection the panel is not representative for the population: mainly poor people are left who are most concerned with their expenses. A second possibility for correction is to estimate the effect of non-response, refusals and underreporting on the aggregate reporting levels, and to correct for these different factors on the aggregate level (Van den Oord and Saris, 1994). In order to do so, a lot of assumptions have to be made without detailed knowledge whether these assumptions are correct. Another possibility is to counteract underreporting on the individual level, by correction on the individual level or by interacting with the respondents by stimulation of motivation. The last procedure requires a thorough insight into the processes behind response behaviour.
In the present chapter we want to formulate a model for these processes, which can be applied in the last approach. Firstly, in Section 4.2 a Structural Equations Model for such a process will be formulated. Secondly, in Section 4.3 this model will tested. Finally in Section 4.4 conclusions will be drawn.

4.2 The model

4.2.1 Rationale

Several reasons can be thought of why reported expenditure will differ from true expenditure. Sometimes more members of the household do the shopping, which requires co-ordination. Incidental purchases are easily forgotten, and in some (if not most) cases motivation is not high enough to spend enough time on answering, so there is no time to enter all purchases in the questionnaire. The difference between true and reported expenditure, response error, can be either random or selective. Apart from random typing errors, only the presence of negative selective response errors due to forgetting or demotivation is expected, which is called underreporting. Demotivation is reflected in the fact that the respondent spends less time on answering. The basic idea in the present approach is that underreporting is caused by this reduction in time.

A formal description of the model is as follows: household characteristics like income or family size \((x_i, z_i)\) determine how much a household consumes \((C_{ti}: true consumption; \) subscript \(t\) denotes time-period, subscript \(i\) denotes household\). An intercept \((\alpha_t)\) reflects the collective fluctuations through time. Since a larger consumption leads to a larger number of items to be entered in the questionnaire, the time a households needs for entering the items \((T_{ti}: true time)\) is first of all spent on starting the system \((r_i)\) and increases with consumption with a factor \(\kappa\) for entering time per item.

As mentioned before, the large burden placed upon respondents in a frequently measuring panel using behavioural or expenditure questionnaires leads to increasing demotivation. Due to this demotivation \((m_{ti})\), the time a household wants to spend in answering will become less than needed for answering or, alternatively, the time a household does not use \((V_{ti}: time reduction)\) is more than zero. With this reduction in responding time, the underreporting \((U_{ti})\) increases with a factor \(1/\kappa\). Underreporting is also influenced by household characteristics like number of shops or family size \((y_i\) and \(z_i)\), which causes a household to forget items. It should be noted that some of these household characteristics influence both consumption and underreporting. Therefore a distinction
between these household characteristics will be made, reflected in a third variable vector: \( z_i \) that contains these variables.

Finally, what is observed is reported consumption (\( r_{it} \): actual reportings), which is true consumption minus underreporting. Likewise, the time used for answering (\( f_{it} \): answering time) is the time needed for answering minus time reduction.

This model is described with the following set of equations.

\[
\begin{align*}
    r_{it} &= C_{it} - U_{it} + \delta_{iti} \quad (4.1a) \\
    f_{it} &= T_{it} - V_{it} + \delta_{2iti} \quad (4.1b) \\
    C_{it} &= \alpha_i + \gamma_1 x_i + \gamma_2 z_i + \varepsilon_{1iti} \quad (4.1c) \\
    T_{it} &= \tau_i + \kappa C_{it} + \varepsilon_{2iti} \quad (4.1d) \\
    V_{it} &= \nu m_{it} + \varepsilon_{3iti} \quad (4.1e) \\
    U_{it} &= 1 / \kappa V_{it} + \gamma_3 y_i + \gamma_4 z_i + \varepsilon_{4iti} \quad (4.1f) \\
    m_{it} &= g_i(m_{i,t-1}) \quad (4.1g)
\end{align*}
\]

where \( t \) denotes the period and \( i \) denotes household, \( g_i() \) is some function of demotivation, and \( \gamma_1, ..., \gamma_4 \), and \( \nu \) are parameters to be estimated; furthermore \( \delta_{iti} \) and \( \delta_{2iti} \) are measurement errors and \( \varepsilon_{1iti}, ..., \varepsilon_{4iti} \) are disturbance terms, with an expectation of zero, uncorrelated with \( x_i, y_i, z_i \), and \( m_{it} \). For each period \( t \), \( \delta_{iti}, \delta_{2iti} \) and \( \varepsilon_{1iti}, ..., \varepsilon_{4iti} \) are mutually uncorrelated. Since more variables influence true consumption apart from \( x_i \) and \( z_i \), the errors \( \varepsilon_{iti} \) are assumed to be correlated through time. \( \varepsilon_{2iti} \) is the amount of disturbance in true time. This disturbance is household specific: e.g. some households are more easily disturbed than other households. The disturbance also contains background variables that are omitted. Since the households are the same each period, \( \varepsilon_{2iti} \) is assumed to be correlated through time.

The factors causing true consumption are assumed to be stable through time, and furthermore it is assumed that the disturbance term \( \varepsilon_{iti} \), which accounts for individual fluctuation per period, has constant variance through time.

The model formulated tries to give a description of the process of reporting. The several steps from demotivation to underreporting are specified, giving insight in the effects of the burden placed upon the respondents. If the model can be estimated, it is possible to simultaneously find the means and variances of true consumption and of underreporting, getting a corrected estimate for true consumption.

Several aspects of the model deserve further attention. Firstly, the time needed for entering one item (\( \kappa \)) is supposed to be constant both across households and through time. However, it is not likely that all households need an equal amount of time for entering
each item. It is also not impossible that people fill in answers faster at a later stage, so $\kappa$ becomes smaller. Solutions to these problems are described in Section 4.2.2.

It is obvious that some of the variables in the model, such as true consumption, and underreporting, are not measured. In fact, all variables denoted with capital symbols are latent. Therefore a Structural Equation Model approach will be used for further analysis. Structural Equation Models allow estimation of relations among latent variables together with measurement relations between latent variables and observed variables if the model is identified (more on identification in Section 4.2.4). The complexity of the model, i.e. the number of latent variables with respect to the number of observed variables, is too high to satisfy identification. In Section 4.2.3 a simplification of the model is proposed, while in Section 4.2.4 identification and estimation of the model are discussed.

The process of demotivation, $g(t)$, is not described, except that demotivation is some function of previous demotivation. A further description of the demotivation process will be presented in Section 4.2.5.

### 4.2.2 Time needed for entering: $\kappa$

In the model, the time needed for entering an item, $\kappa$, is assumed to be constant both through time and across households. This is hardly the case. Firstly, some households are quicker in answering than others, so different households have different entering-times. Secondly, this entering time changes through time: some households become quicker through time, while others take more time per item, e.g. because they become more precise. The latter variation is expected to be smaller than the variation across households. The variation in $\kappa$ is reflected in low correlations between $r_t$ and $f_t$ of at most .12 per moment $t$.

In the model one can allow $\kappa$ to vary through time: $\kappa_t$. Since in panel studies the number of households will generally be much larger than the number of periods, it is not possible to model the variation across households likewise, because this would result in a very large number of extra parameters in the model. Therefore, the variation across households has to be taken care of differently.

Several solutions can be thought of: an attractive solution is to treat $\kappa_t$ as a random coefficient (e.g. see Longford, 1993 or Bryk and Raudenbush, 1992). However, multilevel estimation procedures (Bryk et.al., 1996) available do not allow the complexity of models presented here. Another solution is to split up the households into groups with a similar level for $\kappa_t$. Then these groups will have a more or less constant $\kappa$, which enables estimation using a
Structural Equations Model. However, these groups will be either very small, or very heterogeneous with respect to $\kappa_i$.

A third solution is to estimate the individual $\kappa_i$ per household with the average entering time per item for the whole period. $k_i$, the estimator for $\kappa_i$, is then used to create a new variable for answering time, defined in the same unit of measurement as actual reportings. An estimate for $\kappa_i$ can be found using the average answering time per item: $k_i = \Sigma_i f_i / \Sigma_i r_{ii}$. Next, answering time is divided by this $k_i$. Likewise, the latent variables concerned with time are divided by $k_i$. All redefined variables will be marked with an asterisk (*): $r_{it}^*$, $f_{it}^*$, $C_{it}^*$, $T_{it}^*$, $V_{it}^*$, $\delta_{2it}$ and $\tau_{it}^*$. Likewise, parameter $v^*$ is defined as $v/k_i$. When substituting the old time variables by their redefined counterparts in model (4.1), the factor $\kappa_i/k_i$ and its inverse appear. This factor, which will be close to one, will be treated as model parameter $\lambda_i$. It can be tested whether this parameter is constant through time or not.

The equations for the model after substitution become:

\[
\begin{align*}
r_{it}^* &= C_{it}^* - U_{it} + \delta_{it} \quad (4.2a) \\
f_{it}^* &= T_{it}^* - V_{it}^* + \delta_{2it} \quad (4.2b) \\
C_{it}^* &= \alpha_i + \gamma_1^* x_i + \gamma_2^* z_i + \varepsilon_{iti} \quad (4.2c) \\
T_{it}^* &= \tau_i^* + \lambda_i C_{it}^* + \varepsilon_{2it}^* \quad (4.2d) \\
V_{it}^* &= v^* m_{it} + \varepsilon_{3it}^* \quad (4.2e) \\
U_{it}^* &= 1/\lambda_i V_{it}^* + \gamma_3^* y_i + \gamma_4^* z_i + \varepsilon_{4iti} \quad (4.2f) \\
m_{it}^* &= g_i(m_{t-1}) \quad (4.2g)
\end{align*}
\]

The assumptions with respect to the disturbance terms and random errors remain unchanged.

4.2.3 Simplification of the model

The number of variables appearing in the model is high, which was needed to clarify the process of reporting. Now that the rationale of the model is clear, it is useful to simplify it. It is also necessary, for the number of latent variables is high with respect to the number of observed variables. Since $T_{it}^*$ and $C_{it}$ as well as $U_{it}$ and $V_{it}^*$ are closely related, it is attractive to insert the equations for $U_{it}$ and $T_{it}^*$ into the equations for $r_{it}$ and $f_{it}^*$, so that both $r_{it}$ and $f_{it}^*$ depend directly on the two latent variables $C_{it}$ and $V_{it}^*$.

\footnote{This division will generate dependencies between some of the random errors and the observed variables. The effects of these dependencies on the estimates are expected to be minor, and will be ignored.}
The equations for the model now become:

\[
\begin{align*}
    r_{ii} &= C_{ii} - 1/\lambda_i V_{ii} + \gamma_3 y_i - \gamma_4 z_i - \varepsilon_{4ii} + \delta_{1ii} \quad (4.3a) \\
    f_{ii} &= \tau_i + \lambda_i C_{ii} - V_{ii} + \varepsilon_{2ii}^* + \delta_{2ii} \quad (4.3b) \\
    C_{ii} &= \alpha_i + \gamma_1 x_i + \gamma_2 z_i + \varepsilon_{1ii} \\
    V_{ii} &= \gamma m_{ti} + \varepsilon_{3ii}^* \\
    m_{ti} &= g_i(m_{t-i}) \quad (4.3c) \quad (4.3d) \quad (4.3e)
\end{align*}
\]

Since two error terms appear in both equations (4.3a) and (4.3b), only the variance of the sum of the error terms will be identified. Therefore the combined random errors will be studied instead of the separate random errors: \( \zeta_{1ii} = \delta_{1ii} - \varepsilon_{4ii} \) and \( \zeta_{2ii} = \varepsilon_{2ii}^* + \delta_{2ii} \). Due to the correlatedness of \( \varepsilon_{2ii}^* \) through time, \( \zeta_{2ii} \) will now be correlated through time.

### 4.2.4 Identification and estimation

Estimation of Structural Equation Models (SEM) is done by fitting the variance-covariance matrix, implied by the model parameters, to the variance-covariance matrix in the sample. Fitting is done by minimising the (weighted) sum of squared residuals, where a residual is defined as the difference between the implied (co-)variance and the sample (co-)variance (Bollen, 1989). This procedure can be used for models with or without latent variables. This is in contrast to regression models, where the residuals are defined as the differences between the observed endogenous variables and the predicted endogenous variables (e.g. see Greene, 1993), and therefore no unobserved variables apart from the disturbance term in the regression equation are allowed.

The Pseudo Maximum Likelihood (where Maximum Likelihood is used without the restriction of normally distributed data) estimators are consistent. When data are normally distributed these estimators are asymptotically efficient as well. The behaviour of SEM-estimators has been widely studied, so the effects of violation of the standard assumptions (like non-normality) are known (e.g. Bollen, 1989; Satorra, 1992). Standard software for estimation of Structural Equations Models is available, one of which is LISREL (Jöreskog and Sörbom, 1989). This program will be used to estimate the parameters of the model.

An issue to be studied prior to estimation is the assessment of identification of the model. If there is not enough information, the model is underidentified, which implies that there is an infinite number of solutions for the model parameters. Estimation-procedures will arbitrarily produce one of these solutions, which is undesirable. Therefore, it is important to check identification of the model.
Identification can be verified by algebraically solving the set of equations that imply the variances and covariances for the model parameters. However, the number of equations increases with the number of observed variables: if the number of observed variables is \( q \), the system of equations consists of \( q(q + 1)/2 \) equations.

For the simplified model identifiability is not so difficult to determine. The first step is to express all observed endogenous variables in terms of observed exogenous variables and random errors:

\[
\begin{align*}
\tau_i &= \alpha_i + \gamma_1'x_i + \gamma_2'z_i - 1/\lambda_1 \lambda'_m \tau_i - \gamma_3'\eta_i + \gamma_4'z_i + \epsilon_{1ii} + 1/\lambda_i \epsilon_{3ii} + \zeta_{1ii} \\
\lambda_i &= \tau_i + \lambda_i(\alpha_i + \gamma_1'x_i + \gamma_2'z_i) - \lambda_i \lambda'_m \tau_i + \lambda_i \epsilon_{1ii} - \epsilon_{3ii} + \epsilon_{2ii} \\
m_{ii} &= g_i(m_{i-1})
\end{align*}
\]

These equations form a Seemingly Unrelated Regression Model, of which all parameters and error-variances and error-covariances are identified (Greene, 1993). We can find estimates for the variances and covariances of \( (\epsilon_{1ii} + 1/\lambda_i \epsilon_{3ii} + \zeta_{1ii}) \) and \( (\lambda_i \epsilon_{1ii} - \epsilon_{3ii} + \zeta_{2ii}) \) instead of estimates for the separate variances. The issue now is whether the separate variances are identified as well, given that the variances and covariances of the combination are identified. One of the aspects involved in this question is the level and variance of time reduction in the first period. Since no burden is placed upon the respondents before the first period, it is to be expected that time reduction is at its minimum in the first period. Therefore, it is assumed that time reduction is zero in the first period, with zero variance as well. This implies that \( \epsilon_{31ii} \) is zero. This assumption is sufficient to establish identification for all variances of the error terms, as is shown in Appendix 4.A.

These results on identification are achieved under the assumption that demotivation is measured without errors. In the next section it is shown how this assumption can be relaxed.

### 4.2.5 Demotivation as a simplex

One aspect of the model deserves further attention. So far, demotivation has been assumed to be perfectly measured, which is not plausible. Therefore, demotivation has to be considered as a latent variable. To be able to estimate the model with a latent variable for demotivation, it is necessary to have a perfectly measured cause variable for this variable, or at least two measurements for it at each time point, or at least one measure if the extra assumptions of a simplex structure (Wiley and Wiley, 1969) can be made. The last solution is in line with the assumption that demotivation in one period is the same as demotivation in the previous period except for an increase due to the response burden. Then, if three or
more periods are available, only one measurement is sufficient to identify demotivation if, as Wiley and Wiley suggest, the loadings are set equal to 1, and the measurement errors can be assumed to have equal variance.

The complete model becomes:

\[
\begin{align*}
    r_{ti} &= C_{ti} - 1/\lambda_v V_{ti} - \gamma_3 y_i - \gamma_4 z_i + \xi_{1ti} \\
    f_{ti} &= \xi_t + \lambda_v C_{ti} - V_{ti} - \xi_{2ti} \\
    C_{ti} &= \alpha_t + \gamma_1 x_i + \gamma_2 z_i + \varepsilon_{1ti} \\
    V_{ti} &= \nu M_{ti} + \varepsilon_{3ti} \\
    M_{ti} &= \mu_t + \beta \xi M_{-1} + \varepsilon_{3ti} \\
    m_{ti} &= M_{ti} + \delta_{3ti}
\end{align*}
\]

where \( M_{ti} \) is unobserved demotivation. The same assumptions as in the previous models hold for the errors \( \varepsilon_{1ti}, \varepsilon_{2ti}, \varepsilon_{1ti}, \) and \( \varepsilon_{3ti} \). Furthermore, \( \varepsilon_{1ti} \) and \( \delta_{3ti} \) are assumed to be mutually uncorrelated, uncorrelated through time, and uncorrelated with \( x_i, y_i, z_i \) and the errors \( \xi_{1ti}, \xi_{2ti}, \varepsilon_{1ti}, \) and \( \varepsilon_{3ti} \). The variances of \( \delta_{3ti} \) are assumed to be equal across time. For the first period \((t = 1)\) some different assumptions are made: \( V_{1i}, M_{ti} = 0. \) As a consequence, \( \beta_2 = 0. \)

If the parameters of the simplex-model (4.4e) and (4.4f) are identified \( \{\text{Var}[\varepsilon_2], \text{Var}[\delta_3], \) and \( \beta_3\} \), then the full model is also identified, using the proposition that if all parts of the model are identified, then the whole model is identified (see e.g. Bollen, 1989).

The path diagram of the definitive model is depicted in Figure 4.1. In path diagrams, a straight arrow \((-\rightarrow-)\) between two variables means that one variable causes the other variable, while a two-headed curved arrow signifies an unspecified association between two variables and the absence of an arrow denotes the absence of a direct effect between two variables (a coefficient of zero). Furthermore, an observed variable is enclosed in a square, and a latent variable is enclosed in a circle (for more on path diagrams see e.g. Bollen, 1989). To improve readability, the correlated errors \( \varepsilon_{1ti} \) through time have not been drawn, nor the correlated errors \( \varepsilon_{3ti} \). Also, only two \( x \)-variables are included, and no \( y \)-variables nor \( z \)-variables. Intercepts are not visible in the path diagram.
4.3 An empirical illustration

4.3.1 The data

The data described in Chapter 1 are used to provide an illustration of the model presented above. The weekly data are aggregated to monthly data and the data for the first four months of 1994 are used. In this set-up, a month is defined as a period of four weeks. Due to attrition and renewal not all households are in the sample all the time. Only households that participated during that whole period are considered, resulting in a sample of 740 households. The following variables are available: total number of items reported \((r_{i})\); time
used for answering \( f_{i,t} \); family income \( x_{1,t} \); family size \( x_{2,t} \); evaluation with respect to interest \( v_{i,t} \), measured on a 100 point scale.

It has to be mentioned that evaluation can be used as an indicator for motivation, not for demotivation. However, demotivation in period \( t \) is simply the reduction in motivation from time point 0 to \( t \). For that reason, a new variable is created: \( m_{i,t} = v_{0i} - v_{ti} \). Since evaluations are measured afterwards, it cannot act as an indicator for present motivation. Hence the monthly data for evaluation will be aggregated with a shift of one month. More about the indicator for demotivation in Appendix 4.C.

Time used for answering, \( f_{i,t} \), is measured automatically: a clock in the computer starts at the moment the respondent starts answering the questionnaire, and stops at the moment the respondent is finished. This procedure is subject to outliers: for example if a visitor arrives while the respondent is busy answering the questionnaire, the clock in the computer keeps running until the respondent properly finishes. In some cases, over 1 000 minutes seemed to be needed for answering the questionnaire, at a sample average of less than 20 minutes. The outliers resulting from this and other causes were detected using the Mahalanobis distance (see e.g. Rousseeuw and Leroy, 1987). This was done in two directions: per household the Mahalanobis distance was taken through time, and per period the Mahalanobis distance was taken over households. Penny (1996) provides critical values for the Mahalanobis distance.

Table 4.1: Correlation matrix, means and standard deviations for the variables in four months

<table>
<thead>
<tr>
<th></th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( r_3 )</th>
<th>( r_4 )</th>
<th>( f_{1,t} )</th>
<th>( f_{2,t} )</th>
<th>( f_{3,t} )</th>
<th>( f_{4,t} )</th>
<th>( m_i )</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_1 )</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_2 )</td>
<td>0.833</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_3 )</td>
<td>0.782</td>
<td>0.800</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_4 )</td>
<td>0.775</td>
<td>0.790</td>
<td>0.811</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_{1,t} )</td>
<td>0.840</td>
<td>0.758</td>
<td>0.757</td>
<td>0.733</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_{2,t} )</td>
<td>0.668</td>
<td>0.772</td>
<td>0.694</td>
<td>0.693</td>
<td>0.591</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_{3,t} )</td>
<td>0.696</td>
<td>0.719</td>
<td>0.795</td>
<td>0.698</td>
<td>0.593</td>
<td>0.549</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_{4,t} )</td>
<td>0.666</td>
<td>0.674</td>
<td>0.658</td>
<td>0.766</td>
<td>0.543</td>
<td>0.496</td>
<td>0.533</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_{11} )</td>
<td>-0.016</td>
<td>-0.002</td>
<td>-0.007</td>
<td>-0.012</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.017</td>
<td>-0.046</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_{12} )</td>
<td>-0.039</td>
<td>-0.010</td>
<td>-0.001</td>
<td>-0.032</td>
<td>-0.020</td>
<td>-0.035</td>
<td>-0.015</td>
<td>-0.051</td>
<td>0.729</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m_{13} )</td>
<td>-0.073</td>
<td>-0.059</td>
<td>-0.056</td>
<td>-0.038</td>
<td>-0.046</td>
<td>-0.077</td>
<td>-0.056</td>
<td>-0.046</td>
<td>-0.680</td>
<td>-0.692</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_{1} )</td>
<td>0.198</td>
<td>0.169</td>
<td>0.181</td>
<td>0.140</td>
<td>0.182</td>
<td>0.159</td>
<td>0.137</td>
<td>0.097</td>
<td>-0.019</td>
<td>-0.032</td>
<td>-0.035</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_{2} )</td>
<td>0.366</td>
<td>0.334</td>
<td>0.310</td>
<td>0.331</td>
<td>0.314</td>
<td>0.303</td>
<td>0.266</td>
<td>0.260</td>
<td>-0.014</td>
<td>-0.018</td>
<td>-0.078</td>
<td>0.318</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>4.84</td>
<td>4.79</td>
<td>4.42</td>
<td>4.37</td>
<td>5.45</td>
<td>4.61</td>
<td>4.14</td>
<td>4.33</td>
<td>-0.46</td>
<td>0.45</td>
<td>-2.98</td>
<td>3.26</td>
<td>2.29</td>
<td></td>
</tr>
<tr>
<td>st.dev</td>
<td>2.90</td>
<td>3.08</td>
<td>2.95</td>
<td>2.93</td>
<td>3.88</td>
<td>3.47</td>
<td>3.17</td>
<td>3.36</td>
<td>14.99</td>
<td>15.06</td>
<td>16.35</td>
<td>1.43</td>
<td>1.26</td>
<td></td>
</tr>
</tbody>
</table>

After correction for outliers, the variable \( f_{i,t} \) was transformed to the same unit of measurement as \( r_{it} \) by dividing it by the average \( k_i \) as described above: \( f_{i,t}'' \). The sample correlation matrix for these variables is shown in Table 4.1.
Both correction for outliers and reformulating with $k_i$ resulted in an increase in the correlations between $r_{ti}$ and $f_{ti}^*$ from less than .12 to more than .77.

One conclusion to be drawn from this table is that the correlations between $m_{ti}$ (the indicator for demotivation) and $r_{ti}$ and $f_{ti}^*$ are rather small. Furthermore, the mean of $m_{ti}$ does not increase. Although $m_{ti}$ mutually correlate, and show a simplex structure, it is to be expected that $m_{ti}$ is not a very good indicator for demotivation.

4.3.2 Estimation

The model to be estimated, as shown in Figure 4.1 and described above, is formulated in the equations below:

\[ r_{ti} = C_{ti} - 1/\gamma_t V_{ti}^* + \zeta_{1ti} \]
\[ f_{ti}^* = \gamma_t + \lambda_t C_{ti} - V_{ti}^* + \zeta_{2ti} \]
\[ C_{ti} = \alpha_t + \gamma_{1t} x_{1t} + \gamma_{2} x_{2t} + \epsilon_{1ti} \]
\[ V_{ti}^* = \nu^* M_{ti} + \epsilon_{3ti}^* \]
\[ V_{ti}^* = 0 \]
\[ M_{ti} = \mu_t + \beta_t M_{t-i} + \delta_{5ti} \]
\[ M_{ti} = 0 \]
\[ M_{2ti} = \mu_2 + \delta_{2ti} \]
\[ m_{ti} = M_{ti} + \delta_{1ti} \]

with for all $t$, $u$ and $v$: $\text{Cov}[\zeta_{1tu}, \zeta_{1tv}] = \theta_{\delta_{1tv}}$; $\text{Cov}[\zeta_{2tu}, \zeta_{2tv}] = \theta_{\delta_{2tv}}$; $\text{Var}[\delta_{3t}] = \theta_{\delta_{3t}}$; $\text{Cov}[\epsilon_{1tu}, \epsilon_{1tv}] = \theta_{\epsilon_{1tv}}$; $\text{Var}[\epsilon_{1t}] = \theta_{\epsilon_{1t}}$; $\text{Var}[\epsilon_{3t}] = \theta_{\epsilon_{3t}}$; $\text{Var}[\epsilon_{5t}] = \theta_{\epsilon_{5t}}$.

Further assumptions for all $t$, $u$ and $v$: $\mathbb{E}[\zeta_{1tu}] = 0$; $\mathbb{E}[\zeta_{2tu}] = 0$; $\mathbb{E}[^*_{3t}] = 0$; $\mathbb{E}[\delta_{1t}] = 0$; $\mathbb{E}[\epsilon_{1t}] = 0$; $\mathbb{E}[\epsilon_{3t}] = 0$; $\mathbb{E}[\epsilon_{5t}] = 0$; $\zeta_{1tu}$, $\zeta_{2tu}$, $\delta_{1t}$, $\delta_{2ti}$, $\epsilon_{1t}$, $\epsilon_{3t}$, $\epsilon_{5t}$ are mutually uncorrelated and uncorrelated with $x_{1t}$, $x_{2t}$ and $m_{ti}$.

The reason that the random errors in $r_t$ ($\zeta_{1t}$) are assumed to be correlated is due to the fact that in this (simpler) model no exogenous causes for $r_t$ ($y$) are present. Since these exogenous variables concern the same households each period (like number of household members), this causes the random errors to be correlated through time.

The parameters of the model, depicted in Figure 4.1 and equations (4.4) above, have been estimated with LISREL, which is a standard package for estimating Structural Equations Models, using Pseudo Maximum Likelihood (PML) as estimation method. PML estimation has normality as a working assumption. The data for that estimation are taken from Table 4.1.
The results of the estimation are shown in Table 4.2. The LISREL input file is presented in Appendix 4.D.

The results show a good fit of the model. First of all, the χ² goodness-of-fit index has a value of 52.926 with 43 degrees of freedom, which is smaller than the critical value of 59.02 with α = .05. Almost all parameters are significantly different from zero (more than 1.96

Table 4.2: LISREL maximum likelihood estimates of the model parameters, variances and covariances.

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate (s.e.)</th>
<th>stand.</th>
<th>parameter</th>
<th>estimate (s.e.)</th>
<th>stand.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₁ → r₁</td>
<td>1.000</td>
<td>.958</td>
<td>V₂ → r₃</td>
<td>-1.076 (0.044)</td>
<td>.370</td>
</tr>
<tr>
<td>C₁ → f₁ (λ₁)</td>
<td>1.233 (.043)</td>
<td>.876</td>
<td>V₂ → f₂</td>
<td>-1.000</td>
<td>.303</td>
</tr>
<tr>
<td>C₂ → r₂</td>
<td>1.000</td>
<td>.898</td>
<td>V₂ → r₃</td>
<td>-1.103 (0.043)</td>
<td>.296</td>
</tr>
<tr>
<td>C₂ → f₂ (λ₂)</td>
<td>.929 (.038)</td>
<td>.736</td>
<td>V₃ → f₃</td>
<td>-1.000</td>
<td>.251</td>
</tr>
<tr>
<td>C₃ → r₃</td>
<td>1.000</td>
<td>.922</td>
<td>V₃ → r₄</td>
<td>-1.083 (0.045)</td>
<td>.289</td>
</tr>
<tr>
<td>C₃ → f₃ (λ₃)</td>
<td>.907 (.035)</td>
<td>.783</td>
<td>V₄ → f₄</td>
<td>-1.000</td>
<td>.234</td>
</tr>
<tr>
<td>C₄ → r₄</td>
<td>1.000</td>
<td>.950</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C₄ → f₄ (λ₄)</td>
<td>.923 (.038)</td>
<td>.753</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r₁</td>
<td>-.518 (220)</td>
<td></td>
<td>α₁</td>
<td>2.652 (.256)</td>
<td></td>
</tr>
<tr>
<td>r₂</td>
<td>.158 (198)</td>
<td></td>
<td>α₂</td>
<td>2.604 (.259)</td>
<td></td>
</tr>
<tr>
<td>r₃</td>
<td>.125 (170)</td>
<td></td>
<td>α₃</td>
<td>2.229 (.258)</td>
<td></td>
</tr>
<tr>
<td>r₄</td>
<td>.304 (185)</td>
<td></td>
<td>α₄</td>
<td>2.189 (.258)</td>
<td></td>
</tr>
<tr>
<td>x₁ → C₁ (γ₁)</td>
<td>.156 (.068)</td>
<td>.081</td>
<td>μ₁</td>
<td>-.455 (0.552)</td>
<td></td>
</tr>
<tr>
<td>x₂ → C₁ (γ₁)</td>
<td>.736 (.078)</td>
<td>.337</td>
<td>μ₂</td>
<td>.898 (.403)</td>
<td></td>
</tr>
<tr>
<td>M₁ → V₁ (v₁)</td>
<td>-.003 (.005)</td>
<td>.040</td>
<td>μ₃</td>
<td>-.343 (0.457)</td>
<td></td>
</tr>
<tr>
<td>M₃ → V₃ (v₃)</td>
<td>-.007 (.005)</td>
<td>.114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄ → V₄ (v₄)</td>
<td>-.004 (.004)</td>
<td>.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₁ → M₁ (β₁)</td>
<td>.984 (.046)</td>
<td>.979</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₃ → M₃ (β₃)</td>
<td>1.007 (.042)</td>
<td>.904</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄ → M₄ (β₄)</td>
<td>1.000</td>
<td>.863</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄ → m₄</td>
<td>1.000</td>
<td>.864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.887</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var [e₁₂] (θ₁₂)</td>
<td>1.105 (.311)</td>
<td>.998</td>
<td>Var [e₁₃] (θ₁₃)</td>
<td>.632 (.313)</td>
<td>.987</td>
</tr>
<tr>
<td>Var [e₁₄] (θ₁₄)</td>
<td>.616 (.333)</td>
<td>.994</td>
<td>Var [e₂₃] (θ₂₃)</td>
<td>.167 (12.914)</td>
<td>.000</td>
</tr>
<tr>
<td>Var [e₂₄] (θ₂₄)</td>
<td>38.437 (9.475)</td>
<td>.183</td>
<td>Var [e₃₄] (θ₃₄)</td>
<td>.38 (5.473)</td>
<td>.211</td>
</tr>
<tr>
<td></td>
<td>52.926</td>
<td>(P = 0.143)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are in parentheses, the absence of a standard error means that the parameter is fixed at the given value. Estimates which are different from zero at 5% significance level are printed in bold. Completely standardised solution in italics under "stand." (completely standardised means that both latent and observed variables are standardised to have standard deviations equal to 1). Estimates of (co-)variances of random errors ξ₁₁, ξₛₙ and e₃₄ can be found in Appendix 4.B.
times the standard error). Most of the intercepts for $f^*_t$, $\tau^*_t$ are not significantly different from zero, which means that a very short time is used for answering apart from entering items. The loadings $\lambda_t$ for $t > 1$ are close to 1, which was to be expected. If the simultaneous restriction that $\lambda_t = 1$ for $t > 1$ is made, the $\chi^2$ goodness-of-fit statistic has a value of 53.247 with 46 degrees of freedom. This is not a significant difference with the base-model: $53.247 - 52.926 = .321$ is smaller than the critical value at 5% for $\chi^2_3$ of 7.82.

The estimates of $\nu^*_t$ are not significantly different from zero. This is due to the fact that the evaluations $m_t$ are not very good indicators for motivation $M_t$ (see Appendix 4.C). Testing for these parameters implies testing whether the demotivation-simplex is independent of the consumption underreporting model. The hypothesis that $\nu^*_t = 0$ is not rejected: the $\chi^2$ goodness-of-fit statistic with 46 degrees of freedom has a value of 55.165, leading to a difference with the base model of 2.239, which is smaller than the critical value of 7.82.

It can be tested whether the change in consumption from month to month is significant by testing whether the change in $\alpha_t$ is significant, since all other factors influencing consumption are stable. The hypothesis is tested by estimating the restricted model for $\alpha_t = \alpha_0$. The difference in $\chi^2$ is $(109.870 - 52.926) = 56.944$, which is larger than the critical value for $\chi^2_3$ of 7.82 with $\alpha = .05$, so this hypothesis is rejected, indicating that significant changes in the monthly consumption occurred.

4.4 Conclusions

The study of the process of responding in an expenditure survey led to the formulation of a Structural Equations Model. In this model consumption influences the time households need for answering, and demotivation affects the time respondents are willing to spend in answering, causing underreporting.

Since the proportion $\kappa$, by which consumption determines the time a respondent needs for answering, varies across households, individual proportions, $k_t$, were estimated, by means of which variables concerned with time are reformulated. Furthermore, the complex model was simplified to establish identification. Since in a panel more than two periods are present, a quasi-simplex structure can be used to identify demotivation.

As an illustration of the model, data from four months of an expenditure survey about Fast Moving Goods were used for estimation. The estimations based on these data showed that it is possible to estimate the effects of demotivation on the quality of the data.
Although in our data-set an indicator for demotivation itself is not available, the decrease in evaluation with respect to interest in the investigation is used to measure demotivation.

By filling in the parameter estimates and the means of the exogenous variables in model (4.4), the means of the latent variables can be derived. The means of $C_t$ (true consumption per week) and $V_t^*$ (reduction in time for data entrant per week) are presented in Table 4.3, together with the means of $r_t$ (reported consumption per week).

Table 4.3: Estimated means of $r_t$, $V_t^*$ and $C_t$

<table>
<thead>
<tr>
<th></th>
<th>$r_t$</th>
<th>$V_t^*$</th>
<th>$C_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.843</td>
<td>-0.003</td>
<td>4.843</td>
</tr>
<tr>
<td>2</td>
<td>4.793</td>
<td>0.012</td>
<td>4.795</td>
</tr>
<tr>
<td>3</td>
<td>4.423</td>
<td>-0.003</td>
<td>4.420</td>
</tr>
<tr>
<td>4</td>
<td>4.366</td>
<td>0.001</td>
<td>4.379</td>
</tr>
</tbody>
</table>

Only little time reduction is found. Since the estimates of $1/\lambda_t$ are close to 1, the means of $U_t$ are equal to the means of $V_t^*$ (see equation (4.2f)). These small figures can be explained by the very small estimate for $v_t^*$, which in turn is caused by the fact that an indicator for evaluation, not for demotivation, is used.

Since interest is not an indicator for demotivation, effects are estimated that are very small. The results indicate that the model describing the process underneath response behaviour is identified and can provide useful information on mean consumption, corrected for underreporting, if a good indicator for demotivation can be obtained. It is expected that, with a (better) indicator for demotivation, larger effects can be found.

4.A Appendix: Identification of variances

The issue is to assess whether the variances of $\varepsilon_{1t}$, $\varepsilon_{3t}$, $\zeta_{1t}$ and $\zeta_{2t}$ are identified given that the variances and covariances of $(\varepsilon_{1t} + 1/\lambda_t \varepsilon_{3t} + \zeta_{1t})$ and $(\lambda_t \varepsilon_{1t} - \varepsilon_{3t} + \zeta_{2t})$ (to be referred to as $\omega_1$ and $\omega_2$ respectively, and the variances and covariances as $\psi_{11}$, $\psi_{21}$ and $\psi_{22}$) are identified. To see this, it is useful to write the variance-covariance matrix of $\omega_1$ and $\omega_2$.

Since the assumptions regarding the variances and covariances are not equal in the first period ($\omega_1 = (\varepsilon_{11} + \zeta_{11})$ and $\omega_2 = (\lambda_1 + \zeta_{21})$) across time, the matrices for $t = 1$ and $t \neq 1$ are written down separately. Let $\theta_{k1} = \text{Var}[\zeta_{ki}]$, $\theta_{k1} = \text{Var}[\varepsilon_{ki}]$ and $\theta_{k1} = \text{Var}[\varepsilon_{ki}]$, then:
\[
\Sigma'_{\text{exu}} \begin{pmatrix} \psi'_{11} & \psi'_{12} \\ \psi'_{21} & \psi'_{22} \end{pmatrix} = \begin{pmatrix} \theta_{e11} + \theta_{z_{11}} & \lambda \theta_{e11} \\ \lambda \theta_{e11} & \lambda^2 \theta_{e11} + \theta_{e21} \end{pmatrix}
\]

Since in \(\Sigma'_{\text{ex}}\), for period 1, \(\psi'_{11}, \psi'_{12}, \text{and } \psi'_{21}\), this parameter is identified. Once \(\theta_{e11}\) is known to be identified, \(\theta_{z_{11}}\) and \(\theta_{e21}\) are identified as well.

Now, in \(\Sigma'_{\text{ex}}\), for periods \(t \neq 1\), \(\psi'_{11}, \psi'_{12}, \text{and } \psi'_{22}\), \(\lambda_t\) and \(\theta_{zt}\), which is equal to \(\theta_{e11}\), are known to be identified. \(\psi'_{21} = \lambda \theta_{zt} - 1/\lambda \theta_{zt}, \text{so } \theta_{zt} = \lambda(\lambda \theta_{e11} - \psi'_{21}).\) With this, both \(\theta_{z_{11}}\) and \(\theta_{e21}\) are identified, which completes identification of the whole model.

### 4.B Appendix: Covariances of random errors

<table>
<thead>
<tr>
<th>(\text{Cov}[\xi_1])</th>
<th>(\text{(C)})</th>
<th>(\text{Cov}[\xi_2])</th>
<th>(\text{(r)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>6.520</td>
<td>estimate</td>
<td>3.495</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(4.19)</td>
<td>(s.e.)</td>
<td>(3.55)</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.863</td>
<td>compl. stand.</td>
<td>.233</td>
</tr>
<tr>
<td>estimate</td>
<td>6.188</td>
<td>estimate</td>
<td>-.340</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(4.09)</td>
<td>(s.e.)</td>
<td>4.419</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.819</td>
<td>compl. stand.</td>
<td>.256</td>
</tr>
<tr>
<td>estimate</td>
<td>6.530</td>
<td>estimate</td>
<td>-582</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(3.89)</td>
<td>(s.e.)</td>
<td>3.276</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.863</td>
<td>compl. stand.</td>
<td>.047</td>
</tr>
<tr>
<td>estimate</td>
<td>5.831</td>
<td>estimate</td>
<td>-.760</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(3.83)</td>
<td>(s.e.)</td>
<td>-687</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.863</td>
<td>compl. stand.</td>
<td>-.059</td>
</tr>
<tr>
<td>estimate</td>
<td>.722</td>
<td>estimate</td>
<td>.234</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>.205</td>
<td>(s.e.)</td>
<td>.378</td>
</tr>
<tr>
<td>(\text{Cov}[\xi_2])</td>
<td>(\text{(r)})</td>
<td>(\text{Cov}[\xi_2])</td>
<td>(\text{(r)})</td>
</tr>
<tr>
<td>estimate</td>
<td>.686</td>
<td>estimate</td>
<td>-582</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(.205)</td>
<td>(s.e.)</td>
<td>3.276</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.083</td>
<td>compl. stand.</td>
<td>-.047</td>
</tr>
<tr>
<td>estimate</td>
<td>.081</td>
<td>estimate</td>
<td>-.340</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(.192)</td>
<td>(s.e.)</td>
<td>4.419</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>.537</td>
<td>compl. stand.</td>
<td>-.023</td>
</tr>
<tr>
<td>estimate</td>
<td>-.363</td>
<td>estimate</td>
<td>-.582</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(.187)</td>
<td>(s.e.)</td>
<td>3.276</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>-.546</td>
<td>compl. stand.</td>
<td>-.047</td>
</tr>
<tr>
<td>estimate</td>
<td>-.290</td>
<td>estimate</td>
<td>-.760</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(.183)</td>
<td>(s.e.)</td>
<td>-687</td>
</tr>
<tr>
<td>compl. stand.</td>
<td>-.456</td>
<td>compl. stand.</td>
<td>-.059</td>
</tr>
<tr>
<td>estimate</td>
<td>-.034</td>
<td>estimate</td>
<td>.234</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>-.047</td>
<td>(s.e.)</td>
<td>.378</td>
</tr>
</tbody>
</table>
Chapter 4

4.C Appendix: Indicator for demotivation

In the survey of TelePanel, the members of a household got more than one questionnaire, apart from the questionnaire about meat, poultry and eggs. Most of these questionnaires were only sent incidentally. After finishing the answering of a questionnaire, each respondent got five evaluative questions about how interesting, difficult, clear, well-designed and time-consuming the questionnaire was:

1 = very poor - 10 = excellent
if you really don’t know you can type 0 (zero)

Welk cijfer geeft u voor de VRAGENLIJST voor wat betreft:

* BOEIENDHEID van het onderwerp ................. >> __
* HET GEMAK waarmee de vragen zijn te beantwoorden ........ >> __
* DUIDELIJKHEID van de vraagstelling ............. >> __
* VERDELING VAN DE TEKST op de schermen (lay-out)........ >> __
* DUUR van de vragenlijst ................................ >> __

Translated into English:

The only exception to these evaluation questions was the questionnaire on meat, poultry and eggs. Since this questionnaire was a weekly one, the evaluation questions were considered to evoke a negative mood towards this questionnaire.
For the purpose of the questionnaire, therefore, an indicator for evaluation was not directly available. To have an indicator for demotivation, the evaluation with respect to interest was averaged over all questionnaires and all household-members.

Of course, this is not a very good measure for demotivation, but in the questionnaire this was the only indicator available. For future surveys, it is considered very important to have a good measure for (de-)motivation of the respondents.

### 4.D Appendix: LISREL-input

**AN EXPLANATORY MODEL FOR UNDERREPORTING IN PANEL SURVEYS**

**SEE SECTIONS 4.2 AND 4.3 FOR A DESCRIPTION OF THE MODEL**

\[
\begin{align*}
\text{REPT} = r_{t1}, \quad \text{TIME} = f_{*t1}, \quad \text{EVAL} = m_{t1}, \quad \text{INCOME} = x_{1t1}, \quad \text{FAMSIZE} = x_{2t1} \\
\text{CONS} = c_{*t1}, \quad \text{UNDERT} = \gamma_{*t1}, \quad \text{DEMOT} = M_{t1}
\end{align*}
\]

**DATA NI=13 NO=740 MA=CM**

\[
\begin{array}{cccccccc}
1.000 & .840 & 1.000 \\
.833 & .591 & .772 & 1.000 \\
.782 & .757 & .800 & .694 & 1.000 \\
.696 & .593 & .719 & .549 & .795 & 1.000 \\
.775 & .703 & .799 & .691 & .811 & .698 & 1.000 \\
.666 & .543 & .674 & .496 & .658 & .553 & .766 & 1.000 \\
- .016 & - .002 & - .001 & .007 & .017 & - .013 & - .045 & 1.000 \\
- .039 & - .020 & - .030 & - .035 & .001 & .015 & - .032 & - .051 & .729 & 1.000 \\
- 1.198 & - .182 & - .169 & - .159 & .181 & .137 & .140 & .097 & - .019 & - .032 & - .035 & 1.000 \\
.366 & .314 & .334 & .303 & .310 & .266 & .331 & .256 & .014 & .018 & .078 & .318 & 1.000
\end{array}
\]

**SD**


14.9924 15.0619 16.3518 1.4257 1.2581

**ME**


- .4554 .4501 - 2.9833 3.2595 2.2678

**LA**

REP1 TIME1 REP2 TIME2 REP3 TIME3 REP4 TIME4

EVAL2 EVAL3 INCOME FAMSIZE

**SE**

REP1 TIME1 REP2 TIME2 REP3 TIME3 REP4 TIME4 EVAL2 EVAL3 EVAL4 INCOME FAMSIZE

**MODEL NY-11 NM-2 FI NE-10 LY-FU GA-FU, FI TE-SV, FI PS-SV, FI BE-FU TV-FI AL-FI KA-FI LE

CONS1 CONS2 CONS3 CONS4 UNDERT2 UNDERT3 UNDERT4 DEMOT2 DEMOT3 DEMOT4

FR GA(1,1) GA(1,2)

EQ GA(1,1) GA(2,1) GA(3,1) GA(4,1)

EQ GA(1,2) GA(2,2) GA(3,2) GA(4,2)

FR PS(1,1)

FR PS(2,1)

FR PS(3,1) PS(3,2)

FR PS(4,1) PS(4,2) PS(4,3)

EQ PS(1,1) PS(2,2) PS(3,3) PS(4,4)

FR AL(1) AL(2) AL(3) AL(4)

FR KA(1) KA(2)

VA 1 LY(1,1) LY(3,2) LY(5,3) LY(7,4)

FR LY(2,1) LY(4,2) LY(6,3) LY(8,4)

FR TE(1,1)

FR TE(3,1) TE(3,3)

FR TE(5,1) TE(5,3) TE(5,5)

FR TE(7,1) TE(7,3) TE(7,5) TE(7,7)

FR TV(2) TV(4) TV(6) TV(8)

FR LV(3,5) LV(5,6) LV(7,7)

VA -1 LY(4,5) LY(6,6) LY(8,7)
CO LY(3,5) = -1.0*LY(4,2)**-1.0
CO LY(5,6) = -1.0*LY(6,3)**-1.0
CO LY(7,7) = -1.0*LY(8,4)**-1.0
FR TE(2,2)
FR TE(4,2) TE(4,4)
FR TE(6,2) TE(6,4) TE(6,6)
FR TE(8,2) TE(8,4) TE(8,6)
FR PS(5,5) PS(6,6) PS(7,7)
FR BE(5,8) BE(6,9) BE(7,10)
FR PS(8,8) PS(9,9) PS(10,10)
FR BE(9,8) BE(10,9)
VA LY(9,8) LY(10,9) LY(11,10)
FR TE(9,9)
EQ TE(9,9) TE(10,10) TE(11,11)
FR AL(8) AL(9) AL(10)
OUTPUT AD=OFF TV MI SS SC