

*Erling Røed Larsen*

## Estimating Latent Total Consumption in a Household

**Abstract:**

This article presents a new way of estimating latent total consumption in a household that may improve the accuracy of studies into permanent income and consumption inequality. While the frequently used total purchase expenditure in a household is an unbiased estimator of latent total household consumption, it is inoptimal since total purchase expenditure is an un-weighted sum of expenditures that contain measurement errors. We derive a competing estimator, unbiased and variance minimizing, based on a latent variable model. From estimates of error term variance among consumption indicators, we give accurate indicators more weight, and align weights to minimize variance. An advantage of the suggested estimator is that it allows both expenditure and non-expenditure indicators of latent total consumption. We demonstrate empirically how the minimum-variance estimator reduces variance, and find that on Norwegian expenditure data from 1993 the reduction is 44 per cent.

**Keywords:**

**JEL classification:** C51, C81, D12

**Acknowledgement:** This article describes the model employed in the author's unpublished Ph.D. dissertation submitted at the Dept. of Economics, University of California, Berkeley. The Norwegian Research Council and Statistics Norway financed the project; project no. 120636/730. The following lent encouragement and expertise: Rolf Aaberge, Jørgen Aasness, George Akerlof, Erik Biørn, Clair Brown, Brad DeLong, Michael Hanemann, Christina Romer, David Romer, Dag Einar Sommervoll, Kenneth Train, Thor Olav Thoresen, and Knut Reidar Wangen; in addition to seminar participants at Statistics Norway and University of California, Berkeley. Thanks belong to all, shortcomings only to the author.

**Address:** Erling Røed Larsen, Statistics Norway, Research Department.  
E-mail: [erling.roed.larsen@ssb.no](mailto:erling.roed.larsen@ssb.no)

---

**Discussion Papers**

comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of  
Discussion Papers are available on the Internet: <http://www.ssb.no>

For printed Discussion Papers contact:

Statistics Norway  
Sales- and subscription service  
N-2225 Kongsvinger

Telephone: +47 62 88 55 00  
Telefax: +47 62 88 55 95  
E-mail: [Salg-abonnement@ssb.no](mailto:Salg-abonnement@ssb.no)

## 1. Introduction

Total household consumption has so far been measured inaccurately. Mostly, studies have used total purchase expenditure or transformations of it to estimate latent total consumption in a household. It is unfortunately because household consumption levels play an important role in studies of permanent income and life-cycle theories, consumption inequality, and tax evasion. Kay, Keen, and Morris (1984) suggested a way of utilizing consumption patterns to derive a more precise estimation method of household consumption. This article improves upon that method by allowing many expenditure and non-expenditure indicators. In addition, by employing a latent variable method we avoid having to rely on identical purchasing probabilities.

Consumption is a topic much studied in economics so the paucity of attempts at improving the accuracy in estimation methods of latent total consumption is surprising. The interest in consumption, however, is most natural since the consumption of goods is a main source of utility. Moreover, consumption patterns reflect preferences and material standards of living. Dispersion of consumption opportunities is a core element in the distribution of welfare. Moreover, consumption desires lead to aggregate demand, a core determinant of capacity utilization in macro. Thus, the magnitude of household consumption is an interesting variable in many areas of economics. While consumption is unobservable (latent), expenditure is not. It is intuitively appealing to use total purchase expenditure as a measure of total consumption.

The problem, however, with using purchase expenditure as an indicator of consumption is the wedge between purchase and consumption. Food is purchased, stored, and not in all consumed immediately. A car may be bought at one occasion but it still allows extraction of transportation services for a long period. Holiday trips are purchased infrequently but its enjoyment may exceed the time of the purchase. Stock build up, seasonality, and durable goods pose well-known problems for estimating latent total consumption. However, despite the measurement errors, total purchase expenditure in a household is an unbiased estimator for latent total consumption in a household as long as the errors are zero-mean random variables. Statistical agencies and other students of consumption often use this average of manifest purchase expenditure as an estimator of latent consumption. This article shows how to use a weighted sum of both expenditure and non-expenditure indicators of latent consumption in order to minimize the variance that stems from measurement errors. The idea is simple. Dentist expenditures, medical care, and other big item outlays are done infrequently, and have correspondingly large error variances. Thus, an observer should put less emphasis on such

expenditures when she estimates latent consumption in a household. In stead, expenditures that come with smaller error variances should be given more weight. We demonstrate how to weigh the different expenditures before summing them. In addition, we introduce and show how to incorporate informative non-expenditure indicators that may shed light on the magnitude of latent total consumption in a household.

Consumption is an old topic, so studies of it abound. A review of the literature shall not be attempted. Nevertheless, let us make a few remarks. An important literature on consumption, and with it the start of the resolve of consumption puzzles, originated with Modigliani and Brumberg (1954) and Friedman (1957). They put forward novel lifecycle and permanent-income theories of consumption. One major achievement was that the insights allowed investigators to distinguish between types of consumption and exploit the distinction in understanding how households consume. One lesson, which this article utilizes, was that households incorporate their view of their economic position when they consume. The profession utilized the results to refine and sophisticate the existing theory of consumption, and it is now an impressive body of knowledge. Today, there are many branches of microeconomics engaged in understanding consumption.

Even if many theoretical studies have probed deep, there remain some primary empirical problems. Before we go on to discuss one problem, let us state that the economic profession does have a substantial amount of knowledge about consumption. For example, there has been a massive effort into topics such as the response of consumption to income, see e.g. Campbell and Mankiw (1991) and Baxter and Jermann (1999) for recent contributions. There has been a mapping of the relation between consumption and savings and business cycles; see Caballero (1990), Attanasio and Browning (1995), and Baxter (1996). We have acquired insights of demand patterns and Engel relations, see e.g. Blundell et al. (1993). Measurement errors and simultaneity issues in parameter estimation are well understood. However, the profession still sees contributions to measurement error problems, see e.g. a recent study by Lewbel (1996). Yet there are some features of consumption estimation that are insufficiently covered. One of these is how to estimate the magnitude of total consumption in an individual household. Let us identify some contributions to solving the problem.

An elegant attempt at singling out and estimating latent total consumption in a household came with Kay, Keen, and Morris (1984). They presented a procedure for estimating a household's underlying level of total consumption from observed expenditures. Although an attractive way of measuring consumption, the method does rely on the assumption of common purchasing probabilities among

consumers over goods and does not allow for the utilization of non-expenditure indicators of consumption. Interestingly, Saunders (1980) had earlier explored measures of total consumption. However, he was interested including estimates of collective consumption into a combination of private and collective consumption, not in the precise estimation of private consumption itself. Few attempts at modeling latent total consumption have been made since Kay et al. In the empirical literature, inequality inspections have required estimates of latent total consumption in a household. Examples are Cutler and Katz (1992), Pendakur (1998), Theil and Moss (1999), and Sabelhaus and Groen (2000). A common theme in these studies is the reliance on total purchase expenditure. This article seeks to start filling the gap between current practice and potential in estimating household consumption levels.

The model we formulate contains total household consumption as a latent variable that generates manifestations in the form of expenditures. Moreover, since there is a relation between observable income measures and latent total consumption, we include income as a manifest indicator of latent consumption. There is a large literature on latent variable models, and consequently much is known about latent structures. A review cannot be offered here. Let it suffice to point to the early contributions by Goldberger (1972a,b) and Jöreskog (1978) and to the recent semiparametric estimator Lewbel (1998) develops on the basis of a latent variable model. An overview of the current status and an introduction to the use of latent variables can be found in Wansbeek and Meijer (2000). This article takes the apparatus to consumption data, develops a weighted minimum variance estimator for latent total household consumption, and demonstrates empirically the potential gains in acquiring accurate estimates of individual consumption levels.

The novelty of our approach compared to Kay et al. lies in the inclusion of non-expenditure indicators and the general latent variable approach that avoids using common purchasing probabilities. This makes the model open to many expansions and sophistications and applicable to current consumer expenditure data generated by the different sampling techniques employed by statistical agencies around the world.

An empirical pattern emerges from the data. Employing our weighted estimation method on Norwegian Consumer Expenditure Survey (CES) data from 1993, we find that some expenditure categories are accurate and some are inaccurate indicators of latent total consumption. In our final estimate of latent total consumption in a household we multiply expenditures in the categories Clothing and Footwear, Furniture and Household Equipment, and Other Goods and Services by 1.25,

1.43, and 1.35, respectively. They are accurate indicators of latent consumption and the expenditures are thus given large weights. The categories Medical Care, Transportation, and Recreation and Education are inaccurate indicators of latent total household consumption and are given small weights. Their multiplicative factors are 0.29, 0.28, and 0.47, respectively. Again, the inaccuracy originates in large relative variances of measurement errors. Moreover, we use additional income measures to estimate consumption, and we control for demographic composition of the household by adding correctives for number of children and number of adults. As a result, we find that the weighted estimator reduces variance compared to the un-weighted estimator with 44 per cent.

Let us say in advance where we are headed. The next section presents the model of consumption. Section 3 derives the optimum weights. In section 4 we give an empirical example, in which we present estimates on each weight and compare total purchase expenditures with estimates of latent total consumption for four fictitious households. Section 5 discusses improvement potential and guidelines for future research. Section 6 concludes. In the appendix, we include a description of the Norwegian Consumer Expenditure (CES) data we use; describe issues concerning identification, estimation, and the optimization technique; explain the derivation of optimum weights in detail; and present the estimates of error variances.

## 2. The Model of Latent Consumption

Let us describe the consumption model. A household's consumption of goods is related to the household's latent total consumption and demographic attributes of the household. The idea is that a household first considers the problem of computing an optimum division of time between work and leisure time, then how much to spend and save of the income. Given the resulting total consumption, preferences, and household characteristics the household solves for the latent consumption,  $\eta_i$ , of goods of type  $i$ . The consumption of good  $i$  is a function of latent total consumption and household composition. The relationship may be written as:

$$(1) \quad \eta_{ih} = f_i(\xi_h, z_h), \quad s.t. \quad \sum_i \eta_{ih} = \xi_h, \quad i \in I, h \in H,$$

in which  $f$  is an unknown function, latent total consumption of household  $h$  is denoted  $\xi_h$  and  $z_h$  is the composition of household  $h$ .  $I$  is the set of all good categories and  $H$  is the set of all households. The relation given in equation (1) is an Engel relation, much studied in microeconomics. Further, consumption of good  $i$  in household  $h$ ,  $\eta_{ih}$ , is latent and separated from the manifest purchase

expenditure on good  $i$  by household  $h$ ,  $y_{ih}$ , by a measurement error  $u_{ih}$  since expenditures are reported in accounting books only kept for a short while or through interviews. We write:

$$(2) \quad y_{ih} = \eta_{ih} + u_{ih}, \quad i \in I, \quad h \in H.$$

Combining the relationships in equation (1) and (2) we obtain:

$$(3) \quad y_{ih} = f_i(\xi_h, z_h) + u_{ih}, \quad i \in I, \quad h \in H.$$

Equation (3) relates the observable purchase expenditure on good  $i$  by household  $h$ ,  $y_{ih}$ , to unobservable total consumption,  $\xi_h$ , and household characteristics,  $z_h$  in the unknown form  $f$ . Let us for the time being suppress the shape of the relation in order to pause for a moment on the simple consumer relations. The model implies that the household makes purchasing decisions after having contemplated the total consumption of the period in question, after having decided how much to consume of certain goods, and after having considered the relevant household characteristics such as number of adults and number of children. If this is so, realized purchase decisions are observable manifestations ex post of the ex ante consumption decisions, and thus of latent total consumption. Notice that this article shall revolve around how to use the observable purchase decisions as consumption indicators and how to deal with the measurement error  $u_{ih}$  introduced in equation (2). It is the presence of such errors that poses problems of estimating latent consumption. But our idea is that the corresponding error variance varies over categories and that systematic variation can be exploited. Ultimately, our exploitation shall lead to estimation improvement.

We may use equation (3) and go from the observable left side to make inferences about the unobservable part of the right side. Conventionally, statistical agencies around the world make that inference by summing on both sides, obtaining for a household  $h$  that

$\sum_i y_{ih} = y_h = \sum_i f_i(\xi_h, z_h) + \sum_i u_{ih}$ , in which the first equation sign represents the definition of *total purchase expenditure* and the summing is over goods  $i$ . Thus, when error terms have conditional expectation zero, total expenditure is an unbiased estimator for *latent total consumption*, defined by  $\sum_i f_i(\xi_h, z_h) = \sum_i \eta_{ih} = \xi_h$ . When Engel relations are linear<sup>1</sup>, we obtain a simple relation between total purchase expenditure and latent consumption, household composition, and errors:

$y_h = \sum_i (\beta_i \xi_h + \gamma_i z_h) + \sum_i u_{ih} = \xi_h \sum_i \beta_i + z_h \sum_i \gamma_i + \sum_i u_{ih}$ , in which the parameters  $\beta_i$  sum to unity and the parameter vectors  $\gamma_i$  each sum to zero. Again, the conditional expectation of total purchase

---

<sup>1</sup> An affine structure is obtained by including a constant term in the vector of household characteristics,  $z$ .

expenditure,  $y_h$ , is equal to total consumption,  $\xi_h$ . In other words, the un-weighted sum of a household's purchase expenditure is an unbiased estimator of the latent total consumption in that household. Although unbiased, we shall see below that the problem with using total expenditures as an estimator for latent total consumption lies in the uniform treatment of different error terms. Different error variances invite the use of different weights since uniform treatment entails estimator inefficiency. Let us first formulate our linear stochastic model consisting of equations (4)-(6).

$$(4) \quad y_{ih} = \beta_i \xi_h + \gamma_i z_h + u_{ih}, \quad i \in I, \quad h \in H,$$

$$(5) \quad E(u_{ih} | \xi_h, z_h) = 0, \quad i \in I, \quad h \in H,$$

and

$$(6) \quad E(u_{ih} u_{jk} | \xi_h, \xi_k, z_h, z_k) = \sigma_i^2 \quad \text{when } i = j, \quad h = k; = 0 \quad \text{otherwise.}$$

In the model (4)-(6), purchase expenditures,  $y_i$ 's, and household characteristics contained in the  $z$ 's are observable. The vector of household characteristics  $z_h$  may include composition and demographic variables as well as a unit entry that will include a corresponding constant term in one column of the  $\gamma$  matrix. We may solve for latent total household expenditure, and obtain  $\xi_h = (y_{ih} - \gamma_i z_h) / \beta_i + u_{ih} / \beta_i$ . Thus, if the true latent consumption structure is linear, we can observe a proxy for it by observing expenditures. Thus modeled, expenditures mirror consumption except for the interference from the error term. If we had known the Engel coefficients,  $\beta$ , and the demographic coefficients,  $\gamma$ , we could compute an estimate of latent total consumption. But even if latent total expenditure  $\xi_h$  is unobservable, it is possible to estimate the corresponding coefficient,  $\beta$ , by using the covariance structure in manifest variables; see the Appendix for details on identification and estimation in latent variable models. Assume for now that we know the coefficients. Then, one obvious indicator of latent consumption consists of simply using the observable variables. Let us write this indicator:

$$(7) \quad \hat{\xi}_h^i = \frac{y_{ih} - \gamma_i z_h}{\beta_i}, \quad i \in I, \quad h \in H.$$

Equation (7) indicates that there exist at least as many indicators of latent total consumption as there are categories of goods. Each indicator is conditionally unbiased. Combinations of such indicators extend the list of possible indicators. One such combination is the uniformly weighted sum of indicators from equation (7). It can be written simply as a mean,  $(1/n) \sum_i (y_{ih} - \gamma_i z_h) / \beta_i$ , in which  $n$



is the number of good categories in the set  $I$  and it is a demographics-adjusted version of the total purchase expenditure  $\sum_i y_{ih}$ . Its conditional variance is

$\text{var}\left(\frac{1}{n}\sum_i (y_{ih} - \gamma_i z_h) / \beta_i \mid \xi_h, z_h\right) = \text{var}\left(\frac{1}{n}\sum_i (u_{ih} / \beta_i) \mid \xi_h, z_h\right) = \frac{1}{n^2} \sum_i \sigma_i^2 / \beta_i$ , given our error distribution assumptions. This variance is not the conditional minimum variance. By assigning larger weights to some indicators we may reduce the variance yet retain unbiasedness. Moreover, by including non-expenditure indicators we may further reduce variance.

### 3. The estimator of latent consumption

Consider first a combined, weighted estimator using only expenditures as indicators of latent total consumption. This estimator may be written as:

$$(8) \quad \hat{\xi}_h = \sum_i \omega_i \hat{\xi}_h^i = \sum_i \omega_i \frac{y_{ih} - \gamma_i z_h}{\beta_i} = \xi_h + \sum_i \omega_i \frac{u_{ih}}{\beta_i}, \quad i \in I, \quad h \in H, \quad \text{s.t.} \quad \sum_i \omega_i = 1,$$

in which weights are denoted by  $\omega$ . We see that given equations (4)-(6) the weighted estimator in equation (8) will be unbiased regardless of what weights we use, as long as they sum to unity. The conditional variance of this estimator, assuming equations (4)-(6), is given by:

$$(9) \quad \text{var}\left(\hat{\xi}_h \mid \xi_h, z_h\right) = \text{var}\left(\sum_i \omega_i \frac{u_{ih}}{\beta_i} \mid \xi_h, z_h\right) = \sum_i \omega_i^2 \sigma_i^2 / \beta_i^2, \quad i \in I, \quad h \in H.$$

Our program is to choose weights that minimize the conditional variance given in equation (9). The program can be written compactly as:

$$(10) \quad \min_{\omega} \sum_i \omega_i^2 \sigma_i^2 / \beta_i^2 \quad \text{s.t.} \quad \sum_i \omega_i = 1, \quad i \in I.$$

The program in equation (10) yields the solution given in equation (11). The derivation is tedious, but straightforward, and the result is:

$$(11) \quad \omega_i = \frac{\beta_i^2 / \sigma_i^2}{\sum_i \beta_i^2 / \sigma_i^2}, \quad i \in I.$$

The variance-minimizing weights given in equation (11) are intuitively appealing. Given an Engel coefficient,  $\beta_i$ , of a good category  $i$  the weight will be larger the smaller the conditional variance of the associated error term,  $\sigma_i^2$ . The intuition is that the more frequently, or the larger the magnitude of, the discrepancy between the purchase expenditure and the consumption of a good  $i$ , the larger the

conditional error variance of good  $i$ , thus the smaller the information content of the purchase expenditure on that good category as an indicator of latent total consumption. Smaller weights ensure that the good category  $i$  will be given less emphasis when estimating latent total consumption. Moreover, given a conditional variance,  $\sigma_i^2$ , the larger the Engel coefficient  $\beta_i$  the larger the weight. The intuition is that a good may have a large associated conditional variance in its measurement error, in absolute money terms, simply because it accounts for much of total consumption and thus total purchase expenditure. It then follows that if the Engel coefficient is large we expect a large conditional variance because the household undertake many purchases and/or large purchases of this item. For two goods, with identical magnitudes on associated conditional variances, the good with the larger Engel coefficient will be given a larger weight because the conditional variance is relatively smaller.

The resulting weights in equation (11) resemble the weights derived by Kay, Keen, and Morris (1984, p. 175) in their equation (14) and include much the same interpretation. One benefit of the model suggested in this article is that it does not rely on identical purchase probabilities for all households. In stead, in the latent model suggested above we focus attention on the relationship between purchase expenditure and consumption, regardless of purchasing frequency. Potentially, then, our model is relevant not only to consumer expenditure systems that rely on accounting books kept in short periods, but also to systems that rely on other ways of observing purchases. However, the main benefit of the model suggested here is that it allows for *non-expenditure indicators* of latent total household consumption such as income and wealth. Equation (4) may easily and plausibly be generalized to a relationship between variables important in the household decision of how much to consume and latent total consumption. This opens for a vast array of non-expenditure indicators that may be found in such variables as income and wealth. We extend the basic model by writing the relationship between non-expenditure indicators and latent consumption as:

$$(12) \quad x_{rh} = \beta_r \xi_h + \gamma_r z_h + u_{rh}, \quad r \in R, \quad h \in H,$$

in which the non-expenditure indicator is denoted by  $x$ , the subscript  $r$  refers to indicator type, and  $r$  belongs to a set  $R$  of non-expenditure indicators. Notice how the indicators in equation (12) may imply complicated covariance structures. Aasness (1990) and Aasness, Biørn, and Skjerpen (1993) have argued that we may classify consumption into categories between which it is plausible to argue that the stochastic structure contains zero covariance terms between error terms of different categories. It is less plausible to argue that indicators of the type in equation (12) have a non-zero covariance structure between indicators. This follows from the idea that the error terms  $u_{rh}$  may include common elements among non-expenditure indicators. We extend the stochastic model consisting of equations (4)-(6) to

include equations (12)-(15), specifying the zero covariance between error terms of expenditure indicators and error terms of non-expenditure indicators and allowing for non-zero covariance between the non-expenditure indicators:

$$(13) \quad E(u_{rh} | \xi_h, z_h) = 0, \quad r \in R, \quad h \in H,$$

and

$$(14) \quad E(u_{rh} u_{qk} | \xi_h, \xi_k, z_h, z_k) = \sigma_r^2, \quad r = q, \quad h = k; \quad r, q \in R; \quad = \sigma_{rq}, \quad r \neq q, \quad h = k; \quad r, q \in R,$$

in addition to the assumption that conditional expectations of the product of error terms is equal to zero when household  $h$  is different from household  $k$ . In other words, we assume that error terms associated with non-expenditure indicators for household  $h$  do not covary with error terms of household  $k$ , a plausible assumption when households perform purchases and consumption decisions individually. Moreover, we assume that the conditional expectations of the products of error terms is equal to zero when one indicator belongs to the set  $I$  and the other to the set  $R$ , as written in equation (15):

$$(15) \quad E(u_{rh} u_{ik} | \xi_h, \xi_k, z_h, z_k) = 0 \quad \text{when} \quad r \in R, \quad i \in I; \quad h, k \in H.$$

Similar to the indicator in equation (7), we may solve for latent consumption in equation (12) and obtain an explicit formulation of the non-expenditure indicator:

$$(16) \quad \hat{\xi}_h^r = \frac{x_{rh} - \gamma_r z_h}{\beta_r}, \quad r \in R, \quad h \in H.$$

The estimator in equation (16) is conditionally unbiased when errors are conditionally zero-mean since the estimator is a sum of latent consumption and an error term with conditional expectation equal to zero,  $\xi_h + u_{rh} / \beta_r$ . The conditional variance of the non-expenditure estimator in equation (16) is given by  $\sigma_r^2 / \beta_r^2$ , but combinations of non-expenditure estimators have conditional variances including covariance terms.

We may utilize both expenditure indicators of equation (7) type and non-expenditure indicators of equation (16) type when estimating latent total consumption in a household. The program is to combine them in a conditional variance minimizing way. Since each individual estimator is conditionally unbiased, the resulting combination is also conditionally unbiased. In equation (17) we

present an optimum estimator for latent total consumption in a household that employs both expenditure and non-expenditure indicators:

$$(17) \quad \begin{aligned} \hat{\xi}_h &= \sum_i \omega_i \hat{\xi}_h^i + \sum_r \omega_r \hat{\xi}_h^r = \xi_h + \sum_i \omega_i \frac{u_{ih}}{\beta_i} + \sum_r \omega_r \frac{u_{rh}}{\beta_r}, \\ \text{s.t. } \sum_i \omega_i + \sum_r \omega_r &= 1, \quad i \in I, \quad r \in R, \quad h \in H. \end{aligned}$$

The conditional variance will be minimized when we choose appropriate weights. Let us inspect, without loss of generality, the program for the special, but illuminating, case when we have only two non-expenditure indicators. We do this in order to keep the algebra tractable and transparent. The model may easily be expanded to include any number of indicators. To derive first order conditions, we form the Lagrangian and write it in equation (18), after having derived the conditional variance from equation (17).

$$(18) \quad L = \sum_i \omega_i^2 \frac{\sigma_i^2}{\beta_i^2} + \sum_{r,q} \omega_r^2 \frac{\sigma_r^2}{\beta_r^2} + 2\omega_r \omega_q \frac{\sigma_{rq}}{\beta_r \beta_q} + \lambda \left( 1 - \sum_i \omega_i - \sum_{r,q} \omega_r \right), \quad i \in I; \quad r, q \in R.$$

The Lagrangian becomes more involved as more covariance terms emerge together with more non-expenditure indicators or with different modeling assumptions concerning the non-zero covariance between error terms of expenditure indicators. Inspecting equation (18) we realize that using a combination of several indicators, even inaccurate ones, improves upon using only one indicator, e.g. the most accurate indicator. To see why, note that taking variances of weighted sums entails squaring the weights so that the variance of a weighted sum of indicators with large variances may have a smaller variance than the single, small-variance indicator alone.<sup>2</sup> Put differently, using a large-variance indicator improves upon not using it since it some contains information, however little. This points to the problem of choosing the number of indicators, to which we return below. The first order conditions of the program in equation (18) are:

$$(19) \quad \frac{\partial L}{\partial \omega_i} = 2\omega_i \frac{\sigma_i^2}{\beta_i} - \lambda = 0, \quad i \in I,$$

$$(20) \quad \frac{\partial L}{\partial \omega_r} = 2\omega_r \frac{\sigma_r^2}{\beta_r^2} + 2\omega_q \frac{\sigma_{rq}}{\beta_r \beta_q} - \lambda = 0, \quad r \neq q; \quad r, q \in R,$$

---

<sup>2</sup> To see this, consider the case of two indicators: an accurate  $y_1$  and an inaccurate  $y_2$ . Although the variance of the single indicator  $y_1$  is  $\sigma_1^2 / \beta_1^2$  and that of the weighted combination is  $\omega_1^2 \sigma_1^2 / \beta_1^2 + \omega_2^2 \sigma_2^2 / \beta_2^2$ , the latter can be made smaller than the former by appropriate choices of weights.

$$(21) \quad \sum_i \omega_i + \sum_r \omega_r = 1, \quad i \in I, \quad r \in R,$$

in which the subscripts  $r$  and  $q$  refer to the two (in this case) non-expenditure indicators. In the appendix we outline the derivation of the optimum weights given in equations (22)-(24).

$$(22) \quad \left( \omega_i = \frac{\beta_i^2}{\sigma_i^2} \left( \sum_i \frac{\beta_i^2}{\sigma_i^2} + \frac{\sigma_q^2 \beta_r^2 - 2\sigma_{rq} \beta_r \beta_q + \sigma_r^2 \beta_q^2}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \right)^{-1} \right), \quad i \in I; \quad r \neq q; \quad r, q \in R,$$

$$(23) \quad \omega_r = \frac{\sigma_q^2 \beta_r^2 - \sigma_{rq} \beta_r \beta_q}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \left( \sum_i \frac{\beta_i^2}{\sigma_i^2} + \frac{\sigma_q^2 \beta_r^2 - 2\sigma_{rq} + \sigma_r^2 \beta_q^2}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \right)^{-1}, \quad i \in I; \quad r \neq q; \quad r, q \in R,$$

$$(24) \quad \omega_q = \frac{\sigma_r^2 \beta_q^2 - \sigma_{rq} \beta_r \beta_q}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \left( \sum_i \frac{\beta_i^2}{\sigma_i^2} + \frac{\sigma_q^2 \beta_r^2 - 2\sigma_{rq} + \sigma_r^2 \beta_q^2}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \right)^{-1}, \quad i \in I; \quad r \neq q; \quad r, q \in R,$$

in which the weights for the illustrative case with two non-expenditure indicators are derived. The interpretation of the expenditure weights of equation (22) is given above. The interpretation of the non-expenditure indicators of equations (23) and (24) shall be sketched briefly. A non-expenditure indicator is given a large weight when its slope coefficient,  $\beta$ , is large and when the conditional variance of the other non-expenditure indicator is large. The intuition is that a large slope coefficient reveals that the indicator in question is closely related to latent total consumption and that a large conditional variance of the other indicator that the other indicator is inaccurate. The other terms are common to both indicators.

The gains in accuracy may be substantial. Given the model, both total expenditure and our weighted estimator based on the latent variable model combining expenditures and non-expenditure indicators are unbiased for latent total household consumption. However, total expenditure has a conditional variance of  $\sum_i \sigma_i^2$  and the weighted estimator a conditional variance of

$\sum_i \omega_i^2 (\sigma_i^2 / \beta_i^2) + \sum_r \omega_r^2 (\sigma_r^2 / \beta_r^2) + 2\omega_r \omega_q (\sigma_{rq} / \beta_r \beta_q)$ . Choosing zero weights for the non-expenditure indicators and expenditure weights equal to the Engel coefficient will simply replicate the variance of the total expenditure estimator. Thus, the weighted estimator cannot have larger conditional variance than total expenditure. It may, however, have substantially lower conditional variance when weights are chosen optimally. Below, we shall see that this is indeed the case, and we compute an empirical estimate of the gain in accuracy.

## 4. An Empirical Example

Empirically, we observe through estimation (see the appendix for details) that some goods contain error terms with relatively smaller variance than others. These goods are suited for uncovering the magnitude of latent total consumption in a household. Such categories are Clothing and Footwear and Furniture and Household Equipment. Other goods are imprecise indicators of latent total consumption. Examples are Medical Care and Transportation. The pair gross and net income, taken conjointly, unsurprisingly constitutes an important indicator of latent total consumption, in accordance with the life-cycle theories of Modigliani and Brumberg and Friedman. We find, and demonstrate below, that the weighted estimator reduces estimator variance considerably compared to total purchase expenditure.

Let us now look in more detail at the evidence. First, let us inspect the expression of our weighted estimator given in equation (25):

$$(25) \quad \hat{\xi}_h = \sum_i \frac{\omega_i}{\beta_i} y_{ih} + \sum_r \frac{\omega_r}{\beta_r} x_{ih} - \left( \sum_i \frac{\omega_i \gamma_{ci}}{\beta_i} + \sum_r \frac{\omega_r \gamma_{ar}}{\beta_r} \right) c_h - \left( \sum_i \frac{\omega_i \gamma_{ai}}{\beta_i} + \sum_r \frac{\omega_r \gamma_{ar}}{\beta_r} \right) a_h - \sum_i \frac{\omega_i \alpha_i}{\beta_i} - \sum_r \frac{\omega_r \alpha_r}{\beta_r}, \quad i \in I, r \in R.$$

The subscripts  $c$  and  $a$  are short notation for children and adults and refer to the corresponding variables  $c_h$  and  $a_h$  for the number of children in household  $h$  and the number of adults in household  $h$ . In each indicator relation above, we included constant terms  $\alpha_i$  and  $\alpha_r$ , there described generally as parts of the coefficient vector  $\gamma$ . The inclusion of the variables children and adults is only natural since purchase expenditure patterns vary with household type. A large purchase expenditure on food will have a different interpretation if observed in a single-person household than if observed in a family of four. Consequently, the two observations of large food expenditures will entail different estimates of latent total consumption. Notice in equation (25) that each expenditure category and income type is associated with a preceding factor. The factor denotes what multiple of the indicator we use in the weighted optimum estimator. Put differently, information-rich indicators will have a factor above unity, information-poor indicators below unity. We estimate the parameters in equation (25) from the 1993 Norwegian CES data by using a maximum likelihood method and Levenberg-Marquardt optimization technique described in the appendix. In equation (26) we show the estimates of the resulting 1993-estimator of latent total consumption in a household.

$$(26) \quad \hat{\xi}_h = 0.77y_{1h} + 1.25y_{sh} + 0.56y_{3h} + 1.43y_{4h} + 0.29y_{5h} + 0.28y_{6h} + 0.47y_{7h} + 1.35y_{8h} \\ + 0.01x_{1h} + 0.19x_{2h} - (4615 + 1949)c_h - (-2116 + 12916)a_h - (-17971) - (-20465), \quad h \in H.$$

Equation (26) highlights the main points of this study. Notice first that the subscript numbers of variables  $y$  and  $x$  denote indicator type, explicitly explained in Table 1 below. We observe that some goods, by empirical scrutiny, are found to be better indicators of latent total consumption in a household than others. From equation (26) we see that Clothing and Footwear, Furniture and Household Equipment, and Other Goods and Services are given weights above unity. This follows as a consequence of the fact that they contain errors with small variance compared to the size of the outlay (and thus the associated Engel coefficient). Unsurprisingly, Medical Care and Transportation are found to be especially inaccurate indicators of latent total household consumption, and are thus given smaller weights. This is in accordance with expectations since the categories include big one-time outlays such as dentist services, large medical expenses, and purchases of durable transportation goods as automobiles, motorcycles, and bicycles.

**Table 1: Estimating Consumption Parameters and Household Consumption, 1993**

Good Category or Income Type	Estimated Engel/income coefficient	Estimated Weight	Fictitious Household Annualized Expenditures (14-day Account Book) and Incomes			
			Household A: single- person	Household B: couple	Household C: family with 2 children	Household D: family with 2 children
Food, Beverages, and Tobacco	0.0850	0.0654	21000	43000	65000	72000
Clothing and Footwear	0.1128	0.1411	8000	13000	28000	36000
Rent, Fuel, and Power	0.1702	0.0946	39000	49000	65000	71000
Furniture and Household Equipment	0.1283	0.1832	10000	20000	26000	40000
Medical Care	0.0222	0.0065	4000	8500	35000	9000
Transportation	0.1985	0.0551	35000	65000	89000	100000
Recreation and Education	0.1489	0.0698	18000	27000	42000	45000
Other Goods and Services	0.1342	0.1809	16000	23000	35000	55000
Sum	1.0001	0.7966				
Net Income	0.7664	0.0097	120000	200000	390000	430000
Gross Income	1.0081	0.1936	155000	300000	600000	680000
Sum		0.9999				
		Total purchase expenditure	151000	248500	385000	428000
		Estimated latent consumption	161959	246366	384125	452461

Notes: Parameters are estimated using the Norwegian Consumer Expenditure Survey 1993. In the estimation, we employed the SAS CALIS procedure, LINEEQS option. We used the Maximum Likelihood method assuming a multivariate normal distribution of error terms and the Levenberg-Marquardt optimization technique. See the appendix for details.

Table 1 depicts the main empirical findings of estimation performed on the 1993 cross-section data from the Norwegian Consumer Expenditure Survey. In the first column, we list the ten indicators we employed, eight expenditure indicators and two non-expenditure indicators. The second column tabulates the estimated Engel coefficient associated with the indicator. In the third column, we



compute the weights assigned to each indicator on the basis of estimated error variance and Engel coefficient. The last four columns show hypothetical (for anonymity reasons), but typical, purchase expenditures of four types of households: one single-person household, a couple, and two different families comprising two adults and two children. In the second to last row, we sum purchase expenditures to total purchase expenditure. In the last row, we demonstrate our estimation procedure by computing the resulting estimated latent consumption for each household on the basis of the actual empirical parameter estimates combined with the fictitious household expenditures and incomes.

First, we observe in Table 1 that the estimated Engel coefficients sum to unity (1.0001) and that estimated weights also sum to unity (0.9999). The large weights of Clothing and Footwear, Furniture and Household Equipment, Other Goods and Services are striking. Among the ten indicators, they are given weights of 0.1411, 0.1832 and 0.1809, a total of 50 per cent for three indicators out of 10. This large fraction indicates that some types of purchase expenditure contain much more information on latent total consumption than others and support our argument that the suggested method is important. In consequence, every money unit observed spent on furniture is thus counted 1.43 times when we sum all indicators in the weighted estimator. The factor is as large as 1.43 because given the model the estimated error variance was small relative to the estimated Engel coefficient for the indicator, as indicated in the appendix. The ratio of error variance to Engel coefficient shows that the consumption pattern of households is such that outlays on furniture and household equipment mirror the underlying latent total consumption quite well. In other words, given our model, expenditures in this category uncover reliable information about the magnitude of latent total consumption in a household. Notice further that we have combined the categories Food and Beverages and Tobacco in order to preserve the plausibility of our error structure assumptions, an insight from Aasness (1990). The combined category Food, Beverages, and Tobacco has a relatively small weight of 0.0654, and an accompanying factor of only 0.77 in equation (26) above. Thus, expenditures within this category reveal less information about latent total consumption in the household.

At the other end of the scale lie expenditures on medical care. Empirically, medical care expenditures have been found to be inaccurate indicators of latent total consumption. Thus, it has an associated weight of only 0.0065 and a corresponding estimator factor of 0.29. This was expected. The category medical care is typically one in which households most often have zero purchase expenditure but non-zero consumption since households extract consumption streams continuously out of medical services purchased only once. Moreover, when expenditures within this category are non-zero, they are large, reflecting the infrequency of the purchase of medical services and the payment of a service flow over

many periods. Thus, we cannot infer much about the latent total consumption of a household by observing its expenditures on this category. Further, our empirical findings show that gross income is a better indicator of latent consumption than net income. Gross income has a weight of 0.1936. Probably, the large estimated error variance of net income is connected to the intricacies of the tax system and deductibles. We also keep in mind that the obvious correlation between the two income measures leads us to conclude that we should pay more attention to the collected weights of the income measures than the individual ones. After all, our estimates are the result of a complicated maximum likelihood estimation procedure and if two indicators are highly correlated it is the combined result that interests us the most.

In order to illustrate our estimates of latent total consumption in a household we have included in Table 1 fictitious<sup>3</sup> expenditures of four types of household, one single-person, one couple, and two families with two adults and two children. For household A and D, the simulated estimated latent consumption is larger than the total purchase expenditure. For household B and C, it is smaller. For household D, the difference between total purchase expenditure and estimated latent total consumption is particularly large. The difference follows from the fact that the consumption pattern of household D combined with the income measures hints of a large latent consumption. Potentially, such peculiar consumption style could be allowed from financial wealth, real assets such as a valuable home, gifts, or possibly non-reported income from informal markets.

The latter is an essential point, possibly important to as diverse studies as inequality and tax evasion. Exploiting systematic cross-section consumption structures that contain information on error variances combined with non-expenditure indicators in estimating latent consumption for an individual household allows utilization of a more precise estimator of material standards of living. For example, household D has much larger expenditures on Furniture and Equipment and Transportation than household C has, disproportionately more than incomes would imply. This hints at other sources for consumption such as wealth or credit borrowed on an assumption of future earnings from accumulated human capital. In contrast, household D purchases not much more food, beverages, and tobacco than does household C, and household D has smaller purchase expenditures on medical care. Thus, we infer from the way other households usually behave that household D must know something about its economic position that we do not find fully imprinted in total purchase expenditure. Either it knows something about its future economic position that makes its consumption style different and more luxurious, or it can rely on other financial assets. While the difference between households C and D is

---

<sup>3</sup> They are fictitious in order to preserve household anonymity.

43 000 in total purchase expenditures, the difference is as large as 68 336 in latent total consumption. There is a concomitant divergence between a small income differential and a large material standards of living differential between the two households, a finding with particular relevance to inequality studies. Thus, using the weighted optimum estimator approach to estimating latent total consumption entails extracting more knowledge from the rich data set consumer expenditures furnish us with than simply summing the expenditures. Table 1 demonstrates that using total purchase expenditure instead of latent consumption may lead to different impressions of consumption inequality, supporting the argument that income inequality studies should be supplemented with consumption inequality studies.

We recall that household demographics also play a role in how we model and estimate consumption. Observed purchase expenditure on food will not only reflect total consumption, and thus the material standard of living, but also household size. Larger households consume more food, other things equal. To incorporate demographic sources of consumption, we included a term capturing such effects. This article included the two variables that concern household consumption most directly, composition and size or the number of adults and the number of children in the household. When we observe that a small household has large expenditures on food, we infer from our procedure that it might be quite wealthy, and thus has large latent total consumption, since every calorie purchased probably has cost more than average cost per calorie. This observation leads observers to believe that the household enjoys a high material standard of living. In contrast, when a large family has a small purchase expenditure on food, we believe that most of those expenditures have been on necessities, not on luxury qualities. We do so because of the observation that each calorie was bought cheaply. Thus, we infer that such a household is likely to be poorer than the average and have lower latent total consumption. Similarly, by correcting for demographic composition for each indicator we make use of valuable information on the consumption pattern for and eventually obtain an accurate estimator.

How much do we gain by using the suggested weighted estimator instead of the un-weighted sum of expenditures? Combining the estimates in Table 1 above and Table 2 in the appendix, we compute estimates of the estimator variance. Total purchase expenditure has an estimated conditional variance of 7293151786. Our weighted estimator has an estimated variance of 4055974293. In other words, the total purchase expenditure estimate has a variance 1.8 times larger than the variance of our weighted estimator. Thus, weighting the indicators reduces the variance of the estimates of latent total consumption by 44 percent for the Norwegian 1993 data. Put differently, there are big gains to be reaped from utilizing the whole consumption structure of a cross-section to obtain weights that can be employed in the estimates of individual levels of consumption.

## 5. Discussion

Estimating latent total household consumption is a rich field for exploration, and there is much ground waiting to be covered. There exist several sources of improvement to this article, and let us examine some of them below. The advantages of the suggested method, however, are quite substantial and revolve around three facets: i) the assignment of different weights to expenditure indicators of different accuracy, ii) the employment of non-expenditure indicators of latent total household consumption, and iii) the allowance of estimating latent consumption from several types of consumer expenditure surveys.

One obvious source of improvement is the number of indicators. In this study, we have used only the main eight expenditure categories such as food, housing, and health as indicators. In principle, investigators may use fine-grid, disaggregated groups instead of categories. More goods entail more indicators and presumably more information. One problem, however, with the disaggregated levels is the prevalence of non-purchases that arise as a consequence of short observation periods. This problem naturally increases with grid size and the level of group detail and is a function of the infrequency of purchase. Another problem is the covariance structure. It might be hugely complex. Moreover, one cannot increase the number of parameters to estimate for a given number of observations without assuming costs of inaccuracy. Thus, there may be an optimum balance between the benefit of detailed information and the complexity of models with a large number of indicators. It remains unknown what is the optimum balance.

In order to allow identification and estimation we have relied on well-known methods and empirical techniques. Therefore, we used a linear modeling scheme. As many authors point out, linear Engel curves are probably good approximations of the underlying relation at the relevant intervals, but they may be improved upon, see Aasness (1990) for a treatment and Aasness et al. (1993) for modeling issues. Aasness and Røed Larsen (2002) demonstrate the interpretability power and usefulness in policy of linear models. The literature abounds with examples of more complex Engel curves. However, little is known about the properties of using such curves in the latent model structure suggested here in which reliance on the covariance structure is emphasized when employing maximum likelihood procedures. Further, we have included only the most relevant household characteristics in our model, the number of children and the number of adults in a household. Potentially, there are rich arrays of demographic variables that may enhance precision such as age, education, and region of residence. They will also complicate estimation considerably, and the gains must carefully be weighed with costs. In summary, the utilization of quadratic forms, polynomial

Engel curves, demographic variables, semi-or non-parametric techniques may promise great improvements in estimating latent total household consumption.

Furthermore, there are a number of estimation issues that may be studied. We relied on an initial set of parameter estimates that we specified in the estimation in order to ensure convergence in a few iterations, but we do not know to what extent convergence depended upon initial values. We did, however, some sensitivity tests on initial values, convergence criteria, and estimation procedures. For example, robustness checks and manipulations suggest that convergence was quite robust to the choice of initial values, but the question may be explored further. There are many optimization algorithms, for example the Levenberg-Marquardt, Ridge-stabilized, Newton-Raphson, and Quasi-Newton methods. In the econometric community, there is no consensus of which methods are better suited for which problems. We used the former. In utilizing maximum likelihood procedures, the conventional normality assumptions of the error term distribution are used. Even when error terms are non-normal, the normality assumption and use of maximum likelihood procedures may yield approximately correct answers. These matters await further scrutiny. The assumed error term structure is of importance. We build on results in Aasness (1990) and Aasness et al. (1993) for Norwegian purchasing patterns. Ultimately, how measurement problems occur and how stocks are built are empirical questions, and possibly much different over time and space. Allowance for an explicit modeling of preference heterogeneity among households is a natural extension of this article; see Aasness (1990). We have modeled a representative household, implicitly letting error terms include unaccounted-for effects. The work may with some effort be extended to include random coefficient and finite mixture models.

## **6. Concluding Remarks and Policy Implications**

This article demonstrates how it is possible to improve upon the current practice in estimating latent total household consumption. Weighting different consumption indicators optimally reduced variance by more than forty per cent in the 1993 sample compared to the much-employed estimator of latent total household consumption, the manifest total purchase expenditure.

Household consumption is a core variable in both micro and macro studies of economic relations and processes. One example is the study of the consumption inequality, in which equity concerns lead analysts to undertake empirical investigations of the dispersion of material standards of living. Another is the study of life-cycle standards and permanent incomes. A third is investigations into tax evasion. Thus, there are potential rewards to several areas of economics in improving estimates of individual levels of latent consumption. This article shows that one estimator of latent total household

consumption, the sum of purchase expenditure, can be improved upon. This can be done since purchase expenditure and consumption are separated by stock build up and stock depletion, in addition to other measurement errors. Purchases simply need not reflect consumption in a precise manner. Because of this, some purchase expenditures are better indicators of latent total consumption than others. In other words: they reveal consumption more accurately. This article puts forward an estimator of latent total consumption that utilizes the systematic variation in accuracy by weighting different indicators of consumption differently. Moreover, the estimator we suggest allows inclusion of non-expenditure indicators such as income and wealth. We have derived the estimator that minimizes the conditional variance given the model of consumption and showed that it improves estimate precision by utilizing the content in purchase patterns and income data.

Clear patterns emerge from data. Employing our method on Norwegian Consumer Expenditure Survey data of 1993 we find that expenditure indicators such as Clothing and Footwear, Furniture and Household Equipment, and Other Goods and Services in addition to the non-expenditure indicator Gross Income are valuable indicators of latent total consumption in a household. They are given large weights. Expenditure categories such as Medical Care, Transportation, and Recreation and Education have relatively large variances and are assigned correspondingly small weights. Variances are large for error terms connected to these good categories because there are big differences between latent consumption and manifest expenditure. For example, dental services and automobile purchases are done infrequently, but the related services are consumed over a long period. Thus, such expenses are inaccurate in uncovering latent consumption. A fictitious example of household expenditures demonstrated that our method captures important differences in latent consumption financed from non-income sources like wealth, gifts, and transfers. The method manages to deal with such differences because the consumption patterns undertaken by households reveal economic positions other than incomes and total purchase expenditures would.

We mentioned the practical policy implication in the empirical mapping of consumption inequality. Let us briefly explain how. There are at least two benefits to using the framework. First, the benefit of more accurate estimation of latent total household consumption may allow sharper comparisons between time periods. This is much needed since the comparison of summary measures of distributions often require precise statistics. And in turn, sharper comparisons can improve policy evaluation and assessment of how e.g. tax reforms affect the distribution of material standards. The latent model suggested here allows for estimation of consumption in each household and therefore allows many dispersion measures to be applied to the vector of household consumption levels. Second,

the latent model also invites estimation of another parameter relevant to consumption inequality. The covariance structure makes it possible to obtain estimates of the variance of latent total household consumption itself,  $\sigma_{\zeta\zeta}$ . This variance estimate may be particularly useful since it lies at the center of how the distribution of consumption is generated in a population. Again, time series of estimates may help policymakers evaluate effects of tax regimes and implement new and improved policies.

## References

- Aasness, J. (1990): *Consumer Econometrics and Engel Functions*, Doctoral Dissertations in Economics, 8/1990, Oslo: Department of Economics, University of Oslo.
- Aasness, J., E. Biørn, and T. Skjerpen (1993): Engel Functions, Panel Data, and Latent Variables, *Econometrica*, **61**: 6, pp. 1395-1422.
- Aasness, J. and E. Røed Larsen (2002): Distributional and Environmental Effects of Taxes on Transportation, *Discussion Paper* 321, Oslo: Statistics Norway.
- Attanasio, O. P. and M. Browning (1995): Consumption over the Life Cycle and over the Business Cycle, *American Economic Review*, **85**: 5, pp. 1118-1137.
- Baxter, M. (1996): Are Consumer Durables Important for Business Cycles, *Review of Economics and Statistics*, **78**: 1, pp. 147-155.
- Baxter, M. and U. J. Jermann (1999): Household Production and the Excess Sensitivity of Consumption to Current Income, *American Economic Review*, **89**: 4, pp. 902-920.
- Blundell, R., P. Pashardes, and Guglielmo Weber (1993): What Do We Learn About Consumer Demand Patterns from Micro Data, *American Economic Review*, **83**: 3, pp. 570-597.
- Caballero, R. J. (1990): Consumption Puzzles and Precautionary Savings, *Journal of Monetary Economics*, **25**, pp. 113-136.
- Campbell, J. Y. and N. Gregory Mankiw (1991): The Response of Consumption to Income. A Cross-Country Investigation, *European Economic Review*, **35**, pp. 723-767.
- Cutler, D. M. and L. F. Katz (1992): Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's, *American Economic Review*, **82**: 2, pp. 546-551.
- Friedman, M. (1957): *A Theory of the Consumption Function*, Princeton: Princeton University Press.
- Goldberger, A. S. (1972a): Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables, *International Economic Review*, **13**, pp. 1-15.
- Goldberger, A. S. (1972b): Structural Equation Methods in the Social Sciences, *Econometrica*, **40**, pp. 979-1001.
- Halvorsen, B. and K. R. Wangen (1999): Dokumentasjon av utdrag fra skattestatistikken 1974-1994 for kobling mot forbruksundersøkelsen (Documentation of Excerpts from Tax Statistics 1974-1994 for Linking with Consumer Expenditure Surveys), *Notater: 1999/20*, Oslo: Statistics Norway.
- Jöreskog, K. G. (1978): Structural Analysis of Covariance and Correlation Matrices, *Psychometrika*, **43**, pp. 443-477.
- Kay, J. A., M. J. Keen, and C. N. Morris (1984): Estimating Consumption from Expenditure Data, *Journal of Public Economics*, **23**, pp. 169-181.
- Keesling, J. W. (1972): Maximum Likelihood Approaches to Causal Analysis, Ph.D. dissertation, Chicago.



Lewbel, A. (1996): Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side, *Review of Economics and Statistics*, **78**: 4, pp. 718-725.

Lewbel, A. (1998): Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors, *Econometrica*, **66**: 1, pp. 105-121.

Modigliani, F. and R. Brumberg (1954): Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data, in K. K. Kurihara (ed.) *Post-Keynesian Economics*, New Brunswick, N. J.: Rutgers University Press, pp. 388-436.

Pendakur, K. (1998): Changes in Canadian Family Income and Family Consumption Inequality between 1978 and 1992, *Review of Income and Wealth*, **44**: 2, pp. 259-283.

Sabelhaus, J. and J. A. Groen (2000): Can Permanent-Income Theory Explain Cross-Sectional Consumption Patterns?, *Review of Economics and Statistics*, **82**: 3, pp. 431-438.

Saunders, C. (1980): Measures of Total Household Consumption, *The Review of Income and Wealth*, **26**: 4, pp. 351-366.

SAS (1990): User's Guide, Version 6, Fourth edition, Volume 1, Cary, NC: SAS Institute Inc.

Statistics Norway (1990): *The Consumer Expenditure Survey*, Oslo: Statistics Norway.

Theil, H. and C. B. Moss (1999): The Measurement of Inequality by Components of Total Expenditure, *Empirical Economics*, **24**, pp. 559-561.

Wansbeek, T. and E. Meijer (2000): *Measurement Error and Latent Variables in Econometrics*, Amsterdam: Elsevier Science.

Wiley, D. E. (1973): The Identification Problem for Structural Equation Models with Unmeasured Variables, in A. S. Goldberger and O. D. Duncan (eds.): *Structural Equation Models in the Social Sciences*, New York: Academic Press.

## Data Description and Econometric Techniques

### Data

The Norwegian Consumer Expenditure Surveys contain observations of household samples from 1975 to 1995; see e.g. Statistics Norway (1990) for a detailed description of account books, interviews, and classification. The surveys are conducted continuously every year by Statistics Norway, with 1/26 of households reporting each 14-day period of the year. The samples contain expenditure and other information on more than one thousand households<sup>4</sup> each year, and the samples include information on many socio-economic variables of each household. The sampling scheme is a two-stage stratified, random sample with a small sub-sample of a two-year panel. The response rate is typically above 60 percent. The 1993 sample contained 1308 observations out of a population of about 1.5 million Norwegian households. Our income data were obtained from tax records from Norwegian tax registers for the period 1986-1995. The income data contain the variables gross income (pension-earning income before tax) and net income (pension-earning income after tax). Halvorsen and Wangen (1999) document the income sources.

For this study, all nine main categories were employed as expenditure indicators, but the categories Food and Beverages and Tobacco were bundled together into one large category. Aasness (1990) and Aasness et al. (1993) present reasons why such an aggregation is useful in order to avoid unreasonable zero-covariance assumptions needed for identification. In the profession, there exists no unified theory for how to aggregate goods. Classification is, to a certain extent, arbitrary. This is regrettable, but cannot stop us from undertaking empirical scrutiny of consumer expenditures. Aggregation issues remain important, but we did not explore many alternative systems of aggregation. In stead, we used the conventional categories in aggregation.

### Identification, Estimation, and Optimization

The latent variable model we have suggested can be estimated in several ways and by using many statistical packages. There is a large literature on properties of latent variable models, their identification, and optimum estimation. We use the SAS (1990, p.249) CALIS structure summarized

---

<sup>4</sup> There are about 1.8 million households in Norway.

very compactly in equation (27). The structure builds on models; identifiable and estimable; introduced and explored early by Keesling (1972), Wiley (1973), and Jöreskog (1978).

$$(27) \quad \eta = B\eta + T\xi + \zeta, \quad y = \lambda_y \eta + \varepsilon, \quad x = \lambda_x \xi + \delta.$$

In the notation of equation (27),  $B$ ,  $T$ ,  $\lambda$  are coefficient matrices;  $\eta$  and  $\xi$  are vectors of endogenous and exogenous latent variables, respectively;  $x$  and  $y$  are vectors of manifest variables; and  $\zeta$ ,  $\varepsilon$ , and  $\delta$  are errors.

There are several software packages available that handle latent variable models, and model programming varies somewhat in how to incorporate summing-up conditions such that Engel coefficients,  $\beta$ , sum to unity, and demographic coefficients,  $\gamma$ , and constant terms,  $\alpha$ , sum to zero for the consumption part of the chosen indicators. For example, Jöreskog developed the LISREL system to perform such estimations. The SAS CALIS system we used offered several optimization choices and straightforward programming. Our SAS code and details of estimation results are available at request.

Initially, we ran single-equation two-stage-least-square estimations with several instruments prior to the latent variable estimation in order to be able to specify good initial values for the coefficients. The initial values ensured rapid convergence with a minimum number of iterations. The reported estimates were obtained by using the maximum likelihood method and the Marquardt-Levenberg optimization algorithm.

### Derivation of optimum weights

Recall that the first order conditions in the minimization problem are:

$$(28) \quad \frac{\partial L}{\partial \omega_i} = 2\omega_i \frac{\sigma_i^2}{\beta_i} - \lambda = 0, \quad i \in I,$$

$$(29) \quad \frac{\partial L}{\partial \omega_r} = 2\omega_r \frac{\sigma_r^2}{\beta_r^2} + 2\omega_q \frac{\sigma_{rq}}{\beta_r \beta_q} - \lambda = 0, \quad r \neq q; \quad r, q \in R,$$

$$(30) \quad \sum_i \omega_i + \sum_r \omega_r = 1, \quad i \in I, \quad r \in R.$$

Equations (28)-(30) represent 11 linear (in the weights) equations, in which there are ten unknown weights,  $\omega$ , and one unknown Lagrange multiplier,  $\lambda$ . Solving equation (29) for the non-expenditure weights by using Cramer's Rule, we obtain after multiplication by the product of squared coefficients,  $\beta_r^2 \beta_q^2$ :

$$(31) \quad \omega_k = \frac{\lambda}{2} (\sigma_l^2 \beta_k^2 - \sigma_{rq} \beta_r \beta_q) / (\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2), \quad k \neq l; \quad k, l \in R; \quad r, q \in R.$$

Solving, equation (28) for the weights  $\omega$ , inserting the results into equation (30) together with equation (31), and solving for half the Lagrange multiplier yields:

$$(32) \quad \frac{\lambda}{2} = \left( \sum_i \frac{\beta_i^2}{\sigma_i^2} + \frac{\sigma_q^2 \beta_r^2 - 2\sigma_{rq} \beta_r \beta_q + \sigma_r^2 \beta_q^2}{\sigma_r^2 \sigma_q^2 - (\sigma_{rq})^2} \right)^{-1}, \quad i \in I; \quad r, q \in R.$$

Substituting for  $\lambda/2$  for into the preliminary expressions of the non-expenditure weights in equation (31) and in the solution for expenditure weights of equation (28), yield the final expressions of the weights given above by equations (22)-(24).

### Estimation of error variances and covariance

For completeness, we include in Table 2 the estimates of error variances and covariance.

**Table 2. Estimates of variance and covariance of error terms, Maximum Likelihood, 1993**

Variance	Maximum Likelihood Estimate
$\sigma_1^2$	446403957
$\sigma_2^2$	364247416
$\sigma_3^2$	1238047970
$\sigma_4^2$	362996797
$\sigma_5^2$	304232275
$\sigma_6^2$	2890931499
$\sigma_7^2$	1283641658
$\sigma_8^2$	402650214
$\sigma_q^2$	12664522735
$\sigma_r^2$	20204535290
$\sigma_{rq}$	15291201633