



**HAL**  
open science

## Increasing the price variation in a repeated cross section

Stefan Hoderlein, Sonya Mihaleva

► **To cite this version:**

Stefan Hoderlein, Sonya Mihaleva. Increasing the price variation in a repeated cross section. *Econometrics*, 2008, 147 (2), pp.316. <10.1016/j.jeconom.2008.09.022>. <hal-00536838>

**HAL Id: hal-00536838**

**<https://hal.science/hal-00536838v1>**

Submitted on 17 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Accepted Manuscript

Increasing the price variation in a repeated cross section

Stefan Hoderlein, Sonya Mihaleva

PII: S0304-4076(08)00146-2

DOI: [10.1016/j.jeconom.2008.09.022](https://doi.org/10.1016/j.jeconom.2008.09.022)

Reference: ECONOM 3102

To appear in: *Journal of Econometrics*



Please cite this article as: Hoderlein, S., Mihaleva, S., Increasing the price variation in a repeated cross section. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.09.022

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Increasing the Price Variation in a Repeated Cross Section

Stefan Hoderlein and Sonya Mihaleva

Mannheim University \*

December 2007

## Abstract

Insufficient price variation seriously hampers many application of consumer demand models. This paper examines the empirical performance of a potential remedy for this problem that is suggested by Lewbel (1989), the construction of individual specific price indices for bundles of goods. These individual specific price indices allow for a population with heterogeneity in preferences for goods within a given bundle of goods. We confine ourselves to heterogeneous Cobb Douglas within bundle preferences, while between bundles we allow for several parametric and even general nonparametric specifications. In a variety of settings we show that such prices produce superior empirical results than the ones obtained through the traditional practise of using aggregate price indices. Our empirical analysis is based on the British Family Expenditure Survey data, and uses several categories of food. Both in parametric as well as nonparametric models, we obtain higher precision of estimates for parameters or functions, as well as economically more plausible results.

**Keywords:** Equivalence Scales, Demand System, Weak Separability, Testing Rationality, Nonparametric.

**JEL Classifications:** D12, C30, C14

---

\*University of Mannheim, Department of Economics, L7, 3-5, 68131 Mannheim, Germany, email: stefan\_hoderlein@yahoo.com, sonya\_mihaleva@yahoo.de.

## 1 Introduction

Overcoming difficulties caused by insufficient and measurement error ridden price variation is one of the main challenges for estimating consumer demand systems in product space in practise. Let us illustrate the scope of the problem by looking at the British Family Expenditure Survey, a data set that has been at the center of many of the applications of consumer demand systems using budget shares as dependent variable. There is individual specific information on budget shares, incomes and household characteristics for several thousand households in every year, for a period of now over 25 years<sup>1</sup>. This means that there is certainly enough variation to estimate the influence of income on consumption accurately, using even the most sophisticated methods for estimating Engel curves (cf. Blundell, Chen, Kristensen (2007)). However, there are only aggregate country wide price indices for any given good in any month, meaning that there are at most 300 different observations in the price dimension. Even ignoring minor problems, like the one that different regions are sampled in different months, this is not a lot of price variation compared to the observed consumption/income variation. This is particularly bad news for more elaborate, but data intensive nonparametric methods.

The problem is aggravated by the following fact: These aggregate prices exhibit extremely strong serial correlation. In fact, many of them are  $I(1)$ , some are even  $I(2)$ , see Hoderlein and Lewbel (2007) for examples. This reduces the effective price variation dramatically. Moreover, the dependent variables in consumer demand systems in product space are usually bundles of goods, like food or clothing. In our application, it will be demands for certain subcategories of food. It is needless to mention that the composition of these goods varies across individuals. And it is equally needless to mention that one aggregate price statistic for food is a very poor measure of the actual price faced by the individual, which may be strongly determined by the place she buys the articles, and the composition of her bundle of goods. As a consequence, price effects are only imprecisely estimated. Indeed, there are a great number of findings, particularly in the older literature (see Deaton and Muellbauer (1980b)), and Lewbel (1999) for an overview), which even indicate that compared to what rationality predicts, price effects have the wrong sign, or they have an implausible order of magnitude, which, as our application shows, may well be caused by this shortcoming.

---

<sup>1</sup>We pick the budget share formulation as it is common in applications.

In this paper, we explore one possible remedy. Following Lewbel (1989), who elaborates on an older idea by Stone, we construct cross section prices, exploring individual specific variation in the composition of the bundles of goods. Specifically, as Lewbel points out, the use of an aggregate price index can be understood as assuming that all individuals have identical Cobb Douglas preferences for all goods within a given bundle of goods. In our application, the bundles of goods will be several categories of food (meat, milk, vegetables, bread and cereals etc), and the individual goods comprising meat will be pork, chicken, veal and so on. Because all individuals have the same within bundle preference, the price of the bundle is simply a linear combination of the individual prices using the same weights for all individuals. We relax this assumption by assuming that all individuals have idiosyncratic Cobb Douglas preferences for all goods within any given bundle of goods. This means that the price of, say, meat is still a linear price index, but the weights vary now across the population. These individual specific prices are now our new cross section prices. After the researchers who proposed them first we call them Stone-Lewbel (SL) cross section prices.

In the second section, we explain in more detail the theory behind the construction of these cross section prices. The emphasis of this paper is, however, not theoretical but applied. Therefore, in a large number of applications, we explore the ability of the SL prices to overcome the problems mentioned before. Our focus is hereby in particular on the validity of the integrability properties, i.e., the core restrictions implied by consumer rationality. In several different parametric and nonparametric models, we compare the results using SL prices with the standard practise of employing aggregate price indices.

We start this discussion in the third section by defining the rationality properties we consider in this paper. In the fourth section we introduce the data we employ. In sections five and six, we consider the effect of SL prices on parametric and nonparametric analysis, respectively. Our result will show that with SL prices we are able to increase the precision of our estimates dramatically. Qualitatively, there are some minor changes depending on which rationality restriction we consider. For instance, in the parametric model symmetry was rejected, while homogeneity is not rejected when we use price indices. But this picture changes diametrically with SL prices: Now it is homogeneity that is rejected. Note, however, that in light of our greater precision more rejections should not come as a surprise, as one is not able to reject any null hypothesis if the standard errors are huge. Indeed, just judged by the simple sum of squared deviations from rationality, SL prices

have improved the picture in the sense that those deviations appear less pronounced. But due to the over-proportionally reduced variance, we reject now in some cases where we did not before. Anticipating our conclusion at the end of the paper, we think of such negative results as healthy since they are based on more precise information.

## 2 Stone-Lewbel Cross Section Prices

The concept of SL prices is already discussed intuitively in the introduction. It has often been argued that as a consequence of theory price indices should be utility dependent, which in a heterogeneous population would result in all individuals having their own complicated price index which would depend on unknown parameters of the utility function. However, as we will show below, indices based on Cobb-Douglas preferences are simply linear in the budget shares of goods within a given group of goods, and no individual specific parameters have to be estimated.

Let  $n$  be the number of groups and  $n_i$  be the number of goods in group  $i$ . Furthermore, we denote the quantity and the price of the  $j$ th good in group  $i$  by  $q_{ij}$  and  $p_{ij}$ , the vectors of  $q_{ij}$  and  $p_{ij}$  for  $j = 1, \dots, n_i$  by  $q_i$  and  $p_i$ , and the vectors of all quantities and prices  $q_{ij}$  and  $p_{ij}$  by  $q$  and  $p$ . Let  $y$  be total expenditures, and  $y_i$  be total expenditures in group  $i$ . Then  $w_i = y_i/y$  is the budget share of group  $i$ , and  $w_{ij} = p_{ij}q_{ij}/y_i$  is the within group budget share of the  $j$ th good in group  $i$ , relative to total expenditures in group  $i$ . Let  $s$  be a vector of observable demographic characteristics.

We assume a weakly separable utility function  $U(u_1(q_1, s), \dots, u_n(q_n, s))$ , where  $U(u_1, \dots, u_n)$  is the between-group utility function, and  $u_i(q_i, s)$  is the within-group sub-utility function for group  $i$ . We assume for simplicity that the group sub-utility functions are Cobb Douglas, but we place no restrictions on the functional form of the between-group model.

The budget share demands of are given by

$$w_i = \varphi_i(\mathbf{p}_1 m_1, \dots, \mathbf{p}_n m_n, y) \quad (1)$$

where  $\mathbf{p}_i$  and  $m_i$ ,  $i = 1, \dots, n$ , are the price index and the equivalence scale for group  $i$ , and  $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_n)'$ . We explain now both concepts: First, since we assume homothetic separability this implies that the price index does not depend on the total expenditures in the group. Instead, the price index depends on the prices in the group and the demographic characteristics of a reference household  $\bar{s}$ . Hence, homothetic separability implies

the existence of functions  $\nu_i$  such that  $\mathbf{p}_i = \nu_i(p_i, \bar{s})$ , and therefore the equivalence scale  $m_i$  can then be written as:

$$m_i = \nu_i(p_i, s) / \nu_i(p_i, \bar{s}). \quad (2)$$

From now on we call  $\pi_i = m_i \mathbf{p}_i = \nu_i(p_i, s)$ ,  $i = 1, \dots, n$ , the Stone-Lewbel (SL) price for the group  $i$ , and the vector  $\pi = (\pi_1, \dots, \pi_n)'$  the vector of SL prices.

The within-group budget share demand, conditional on the expenditures in group  $i$  being  $y_i$  is

$$w_{ij} = h_{ij}(p_i, s, y_i). \quad (3)$$

These budget share equations are obtained by maximizing  $u_i(q_i, s)$  subject to  $p'_i q_i = y_i$ . If we assume, that demands are homothetically separable,  $y_i$  drops out of  $h_{ij}$ , and

$$\log(\nu_i(p_i, s)) = p_{ij} \int h_{ij}(p_i, s) dp_{ij} \quad (4)$$

for  $j = 1, \dots, n_i$ . From the estimation of  $h_{ij}$  we can construct the SL prices  $\pi_i$ . Since  $\pi_i = m_i \mathbf{p}_i$ , we can use  $\pi_i$  directly in place of prices in estimating the between-group demand equations in (1).

The variation in the demographic characteristics causes variation in the equivalence scales and consequently increases the precision with which the between-group demand equations are estimated. We need only enough price variation to identify the within-group demand models and to construct the equivalence scales. Then demographic variation in the equivalence scales replaces the necessary price variation in terms of estimating the between-group demand model.

We consider the special case, where the within-group sub-utility functions are Cobb-Douglas:

$$u_i(q_i, s) = k_i \prod_{j=1}^{n_i} q_{ij}^{w_{ij}}, \quad (5)$$

with a scaling factor  $k_i$ , given by

$$k_i = \prod_{j=1}^{n_i} \bar{w}_{ij}^{-\bar{w}_{ij}} \quad (6)$$

where  $\bar{w}_{ij}$  is the budget share of good  $j$  in group  $i$  of the reference household. Note the dependence of  $u_i(q_i, s)$  through  $w_{ij}$ . We take as reference household the one with average budget shares and construct the SL prices as:

$$\nu_i(p_i, s) = \frac{1}{k_i} \prod_{j=1}^{n_i} \left( \frac{p_{ij}}{w_{ij}} \right)^{w_{ij}} \quad (7)$$

and use it in place of original price data to estimate the between-group budget share equations. We would like to point out that although the within-group model is Cobb-Douglas, the between group model can be arbitrarily complicated. Specifically, in section 5 we use the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980a) and in section 6 a nonparametric model to estimate the between-group demand system.

This defines the model on individual level. From now on, in order to do econometrics we assume to have a heterogeneous population of such individuals, and hence all variables are denoted by capital letters to indicate that they are random variables that vary across time and population. Consequently, to define the relationship of interest we use the following definitions: Let  $P$  define the  $n$ -vector of log-prices, where in the following we will use either price indices (i.e., a realization will be  $\log(\mathbf{p})$ ), or SL-prices (i.e., a realization will be  $\log(\pi)$ ). Moreover, let  $X$  denote log total expenditure, (i.e.,  $x = \log(y)$ ),  $W$  budget shares, and  $S$  and  $A$  observable and unobservable characteristics. Consequently, the Marshallian demand function in the heterogeneous population is defined  $W = \varphi(P, X, S, A)$ , and properties of this objects using both definitions of  $P$  will be at the core of the following analysis.

### 3 Analyzing Rationality Properties using SL-Prices: Models

**Rationality Properties:** Define theoretical demand functions to be rational if they are derived from utility maximization under a linear budget constraint. Define demand function to be homogeneous if they are homogenous of degree zero in the levels of prices and total outlay. This homogeneity condition is also known as the absence of money illusion, and is one of the key implications of rationality. Other conditions required for rationality are the adding up condition, and negative semidefiniteness (short nsd), as well as symmetry of the Slutsky matrix.

To define the latter two, recall that  $P$  defines the  $n$ -vector of log-prices, and that the Marshallian demand function for individual with observables  $S$  and unobservables  $A$  is  $W = \varphi(P, X, S, A)$ , where  $X$  denotes log total expenditure. In the following, let the observables be denoted through  $Z = (P', X, S')'$ . For any demand function  $\varphi$ , the Slutsky matrix  $\mathfrak{S}(Z, A) = (\mathfrak{S}^{ik}(Z, A))_{1 \leq i \leq n-1, 1 \leq k \leq n-1}$  is defined as the Hessian of the cost function with respect to prices. The elements may be expressed in terms of log-price and

log-expenditure as

$$\mathfrak{S}^{ik}(Z, A) = \partial_k \varphi_i(Z, A) + \varphi_k(Z, A) \partial_x \varphi_i(Z, A) + \varphi_i(Z, A) \varphi_k(Z, A) - \delta^i(k) \varphi_k(Z, A), \quad (8)$$

where  $\delta^i(k)$  denotes the Kronecker function to indicate a diagonal element and  $\partial_x = \frac{\partial}{\partial X}$  and  $\partial_k = \frac{\partial}{\partial P_k}$  are used for abbreviation (see Mas-Colell, Whinston and Green 1995).

**Parametric Modelling:** In this paper, we want to analyze the effects of SL-Prices on evaluating these hypotheses both if we use traditional parametric tests, as well as more modern nonparametric, nonseparable models. Our parametric example will be the workhorse model in this literature, the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980a) which we use to estimate the between-group demand system. The parameters of the model are  $n$  vectors  $\alpha$  and  $\beta$ , and a  $n \times n$  matrix  $\gamma$ . The model consists of theoretical demand functions of the form

$$W = m(P, X) + V \quad (9)$$

where  $\mathbb{E}(V \mid P, X) = 0_n$  and

$$m(P, X) = \alpha + \gamma P + \beta [X - g(P)] \quad (10)$$

where

$$g(P) = d + \alpha P + \frac{1}{2} P' \gamma P \quad (11)$$

is the log of a price index that deflates total expenditures.

When estimating the standard AIDS model defined through equations (9) and (10), there are a number of issues to consider. The most important one is endogeneity. In consumer demand, total expenditure is taken as income concept, which is justified by assuming intertemporal separability of preferences. In our case, we take total food expenditure as our income concept. Since the categories of goods considered are broad (e.g., meat, fats) and they frequently constitute a large part of total food expenditure, the latter is believed to be endogenous. As instrument, the demand literature usually employs labor income whose determinants are thought to be exogenous to the unobserved preferences determining, say, meat consumption. To allow for endogeneity, we estimate equation (9) using three stage least squares (specifically, GMM with a weighting matrix that is efficient under homoscedasticity, which for large systems like ours can have good finite

sample properties and fewer numerical convergence problems than general efficient GMM. See e.g., Wooldridge 2002 for details). Estimation is based on the moment conditions

$$\mathbb{E} \left[ \left( W - \alpha - \gamma P - \beta \left[ X - \alpha' P - \frac{1}{2} P' \gamma P \right] \right) R_l \right] = 0_n, \quad l = 1, \dots, L \quad (12)$$

where  $R_1, \dots, R_L$  is the set of instruments.

In this parametric setup, the adding up constraint is that budget shares sum to one and requires  $1'_n \alpha = 1$ ,  $1'_n \beta = 0$ , and  $1'_n \gamma = 0'_n$ . Homogeneity requires  $\gamma 1_n = 0_n$ , and Slutsky symmetry requires symmetry of  $\gamma$ . These restrictions together yield a rational AID system, with indirect utility function given by

$$F [\ln (X - g(P)) - P' \beta] \quad (13)$$

where  $F$  is any function that is strictly monotonically increasing. The AID budget share functions (9) are obtained by applying Roy's identity to equation (13). In this setup, we implicitly assume that unobserved heterogeneity is not a major issue. However, Lewbel (2001) shows that this approach is very restrictive even in the case when the population is of the same parametric form, with parameters that vary across the population.

**Nonparametric Modelling:** A more general alternative is nonparametric modelling. In terms of incorporating observable and unobservable heterogeneity, we use the framework put forward in Lewbel (2001) and Hoderlein (2007). To this end, we introduce the following notation: Recall that  $Z = (P', X, S')'$  and let  $m(Z) = \mathbb{E}[W|Z]$ . Moreover, let  $m_2(Z) = \mathbb{E}[WW'|Z]$  denotes the second moment regression, and  $\overline{B}$  is the symmetrized version of a matrix  $B$  (i.e.,  $\overline{B} = B + B'$ ). The general nonparametric identification results coming out of this approach are (cf. Hoderlein (2007), Proposition 2.3) that in a heterogeneous population:

$$\begin{aligned} \mathfrak{S}(Z, A) \text{ symmetric} &\Rightarrow D_p m(Z) + \partial_x m(Z) m(Z)' \text{ symmetric} \\ &\text{iff } \mathbb{V} [\partial_x \varphi(Z, A), \varphi(Z, A) | Z] \text{ symmetric,} \end{aligned}$$

and

$$\mathfrak{S}(Z, A) \text{ nsd} \Rightarrow \overline{D_p m}(Z) + \partial_x m_2(Z) + 2(m_2(Z) - \delta \{m(Z)\}) \text{ nsd.}$$

where  $\mathbb{V}$  denotes the conditional covariance matrix between two random vectors.

In this paper, we test the respective right hand side hypotheses nonparametrically, using kernel based local polynomial regressions for the demand functions  $m$  and apply bootstrap methods to obtain critical values for hypothesis test statistics as in Hoderlein (2007), Haag and Hoderlein (2006) and Haag, Hoderlein, and Pendakur (2007). Note that nonparametric models are more demanding in terms of their data requirement. Hence we would expect there to be particular gains from applying SL prices compared to the more robust parametric methods.

## 4 The Data: FES

The data we employ consist of observations on food expenditures, demographic composition, and other characteristics drawn from the years 1994 to 1999 of the British Family Expenditure Survey (FES). The sample consist of about 10,000 households. The information is collected partly by interview and partly by records. Records are kept by each household member, and include an itemized list of expenditures during 14 consecutive days. The periods of data collection are evenly spread out over the year. The information is then compiled and provides a repeated series of yearly cross-sections.

The expenditures for food are grouped into four categories. The first category (“carbo”) is related to carbohydrates and consists of the subcategories bread, potatoes, fruits and vegetables. The second category (“meat”) contains expenditures for beef, lamb, pork, bacon and fish. The third category is called “fats” and it contains expenditures for butter, oil and fats, cheese, eggs, milk and milk products. Finally, the last category (“drinks”) consists of tea, coffee and soft drinks. Together all four categories account for 60-65% of food expenditure on average, leaving a sizeable fifth residual category. The corresponding price indices which are components of the Retail Price Index are published at the National Statistics Online web site.

Disposable income, which we take as an instrument for total expenditure is constructed as in the definition of household below average income study (HBAI). It is roughly defined as net income after taxes, but including state transfers. To account for observable demographic heterogeneity, we include the following demographic household characteristics: number of adults, number of male adults, number of retired persons, number of earners, number of children in the household and a variable indicating ownership of a fridge. We provide summary statistics of our data in table 3 in the appendix.

## 5 The Effect of SL-Prices in Parametric Models

### 5.1 Issues in Estimation

By analogy with the common approximate AID model, we first linearly regress  $W$  on a constant,  $P$ , and on  $X - P'W$ . This is not a consistent estimator, but yields reasonable starting values  $\hat{\alpha}$ ,  $\hat{\beta}$ , and  $\hat{\gamma}$  for the calculation of  $g(P)$ . We can obtain new estimates exploiting the conditional linearity of equation (10) given  $g(P)$ . That is, given  $g(P)$ , the system is linear in parameters, and this suggests a natural iterative procedure conditioning on an updated  $g(P)$  at each iteration.

A second remark concerns the possibility of the estimation of  $d$ . When the total price index and the demand system are estimated simultaneously, the magnitude of  $d$  will have an impact on the estimated coefficients via the change of  $g(P)$ . In order to avoid the difficulties of such simultaneous estimation, we choose to estimate  $d$  using the coefficients  $\alpha$ , and  $\gamma$  from the estimation of  $W$ , to calculate the translog price index, and reestimate factor demand equations  $W$ , until convergence is achieved, i.e. until  $\alpha$ , and  $\gamma$  are the same in  $g(P)$  and  $W$ . However, the convergence properties of this procedure in finite samples are very poor. Therefore we adopt the simplification carried out in most of the applied studies and we set the value of  $d$  to zero.

When estimating the system we only impose the adding up constraints by omitting one good from the system. We estimate the model without the homogeneity and Slutsky symmetry restrictions  $\gamma 1_n = 0_n$  and  $\gamma = \gamma'$ , so these restrictions can be tested. We also calculate the Slutsky matrix and test for negative semidefiniteness. To allow for endogeneity we estimate the model a second time using GMM where the set of instruments consists of a constant, log labor income and all elements of  $P$ . We evaluate the Slutsky matrix at the average household.

### 5.2 Empirical Results

As the parametric models under both exogeneity and under endogeneity do not generate substantially different results, we only give the detailed results with respect to endogeneity-corrected estimates. We use labor income as instrument for food expenditure. In table 1.1 we show the uncompensated elasticity estimates of each budget share with respect to prices using the price index, along with standard errors. Table 1.1 also

reports the elasticity estimates of each budget share with respect to food expenditures. Before compensating, the elasticities of the budget shares of meat, fats and drinks with respect to the own price index are negative, but in absolute values smaller than one. The elasticity of the budget share of fats is positive, however, it is not statistically significant. The elasticity estimates of each budget share with respect to own SL-prices are quite different (table 1.2). They are all significant and, except for fats, larger than one in absolute values. Most of the cross price elasticities in the case of price indices exhibit large standard errors and are not statistically significant. However, it is interesting to note that the standard errors of the price elasticity estimates in the case of SL-prices are substantially smaller. The last row of table 1.2 reports positive effects of the change of the SL prices of carbohydrates, meat and fats on the budget share of drinks. Most of the other cross price elasticities are negative. In all cases, we observe elasticities nearly equal to one with respect to food expenditure. As one would have expected, the elasticity of the budget share of meat is larger than one, and those of carbohydrates and drinks are smaller than one. Hence, we can conclude that carbohydrates and drinks are both necessities, and meat is a luxury good. Using price indices, fat is a luxury good, while in the case of SL prices it appears to be a necessity. If we allow for endogeneity, the expenditure elasticities of carbohydrates, meat and drinks exhibit slightly higher standard errors. The use of labor income as instrument for a small fraction of the expenditures, in particular for food, leads to a imprecise estimation of the expenditure elasticities.

Turning to the Slutsky matrix, the compensated own price effects of all goods are negative in the case of SL prices (table 1.4). Furthermore, they are stronger than the compensated own price effect in the case of price indices (table 1.3). Carbohydrates and drinks have compensated own price effects of -0.166 and -0.047 in the case of price indices and -0.291 and -0.103 in the case of SL prices. The compensated own price effects of meat and fats are positive (0.015 and 0.030) in the case of price indices and turn to negative (-0.208 and -0.049) in the case of SL prices. Most of the compensated cross price effects in the case of price indices are not statistically significant. In the case of SL prices most of the compensated cross price effects are negative (except for those of drinks to other goods). As expected, the composite commodity Slutsky matrix in the case of SL prices is negative semidefinite, as all eigenvalues are negative. Since negative semidefiniteness is arguably the core property of rationality (Weak Axiom), this finding is very comforting. By contrast, in the case of price indices we observe one positive and three

negative eigenvalues, and hence the Slutsky matrix is indefinite. Moreover, note that the Slutsky matrix is also more symmetric in the case of SL prices: all but one restriction have at least the same sign, and often their rough order of magnitude coincides.

Table 1.6 presents Wald tests of the homogeneity and symmetry restrictions. In both cases symmetry is rejected. Homogeneity is rejected in the case of SL prices and not rejected in the case of price indices. Table 1.5 lists test statistics and p-values of the homogeneity test separately in each of the four demand equations. In the case of price indices homogeneity is not rejected for any good at the 5 % significance level. In the case of SL prices homogeneity is not rejected only for one good. To analyze further this somewhat unexpected result, we calculate the sum of the price elasticities separately in each equation. We know that the Wald test statistic for homogeneity is a quadratic form of the price elasticity vector weighted by the respective covariance matrix. A careful examination shows that the unweighted squared sum of the price elasticity estimates is in both cases almost equally distant from zero. Hence, the smaller Wald test statistics in each equation in the case of price indices are to a large extent due to the larger standard errors. Far fewer rejections are obtained with individual consumer data in the next section.

## 6 The Effect of SL-Prices on Nonparametric Models

### 6.1 Issues in Estimation

In this subsection we briefly discuss issues that arise when we attempt to estimate the object of interest directly, using a nonparametric approach. Since we are using nonparametric regression, we cannot employ many discrete covariates. There are a number of remedies suggested in the literature: One is to use a partially linear regression with scaling as in Blundell, Chen and Kristensen (2007). Since we are already focusing on a subpopulation (and as such using covariate information), we propose to reduce the information in  $Z$  to an approximately continuously distributed principal component. Specifically, the covariate  $Z$  is constructed as the first principal component of the demographic household characteristics described above. As in the parametric case, endogeneity of total food expenditure is also an issue. Hence we estimate the model a second time allowing for endogeneity of  $X$  by using labor income as an instrument and including corresponding control function residuals in the nonparametric regressions, as in Imbens and Newey

(2007), or Hoderlein (2007).

## 6.2 Empirical Results

In figures 1-4 in the appendix we show the estimated budget shares of the four categories of goods against log food expenditure, together with 90 % bootstrap confidence bands. All functions are plotted at the mean level of all other regressors. The budget share of carbohydrates is constant, the budget share of meat is increasing and the budget shares of fats and drinks are weakly decreasing in log food expenditure. As was to be expected there is little effect of the introduction of SL prices on the estimation of the budget share Engel curves, as these use largely the plentiful income variation in the cross section. For a similar reason, the confidence bands in the case of SL prices do not differ substantially from those in the case of price indices. For low levels of income the confidence bands of the budget shares of carbohydrates and meat are slightly larger in the case of SL prices. In contrast, the effect of SL prices shows very clearly when estimating the compensated own price effects. These are presented in figures 5-8 for the four goods. As predicted by theory, the effects of carbohydrates, meat and drinks are negative, whether we use SL prices or price indices. However, these negative effects are more pronounced in the case of SL prices than in the case of price indices. Moreover, the confidence bands would be much smaller in the case of SL prices, implying that all compensated own price effects are statistically different from zero (and smaller), while with price indices we cannot reject the null of a zero or positive price effect even in cases where the point estimate is negative. Note in particular, that the compensated own price effect of fats is positive across the whole expenditure range, but negative in the case of SL prices. These observations suggest that we may reject negative semidefiniteness with price indices, while we may accept it using the SL prices. Moreover, the confidence bands of the compensated own price effects of all goods are substantially smaller when we apply SL prices instead of price indices, reflecting again the higher precision obtained. From these results it is obvious that SL prices provide a more precise and more reasonable estimate of the Slutsky matrix.

Again we only give the detailed results with respect to endogeneity-corrected estimates. Table 2.1 and 2.2 report the average uncompensated elasticity estimates of each budget share with respect to price indices (table 2.1) or with respect to SL prices (table 2.2) and with respect to food expenditures, along with bootstrapped 95% confidence intervals. The elasticities with respect to the own price show similar patterns to the parametric results.

Table 2.1 presents positive elasticities of meat and fats and negative elasticities, but smaller than one in absolute values, of carbohydrates and drinks with respect to the own price indices. Table 2.2 reports negative and larger than one in absolute values elasticities of the budget shares of carbohydrates, meats and drinks with respect to the own SL prices, and only the budget share of fats exhibits negative but inferior effects. Similarly to the parametric results, the confidence bands around the price elasticities are smaller in the case of SL prices. However, applying SL prices does not affect substantially the elasticity estimates with respect to log total expenditure, confirming results in Hoderlein, Klemelä and Mammen (2007) which show that both effects are approximately independent.

In the nonparametric case the Slutsky matrix differs across observations. We calculate the Slutsky matrix at each observation using the corresponding prices and expenditures. Then we build the average over all Slutsky matrices. Table 2.3 presents the average Slutsky matrix in the case of price indices and table 2.4 in the case of SL prices. In both cases the compensated own price effects are similar to those in the parametric model. The compensated cross price effects in the nonparametric model are estimated with higher precision and are quite different from those in the parametric model. Carbohydrates, meat and drinks have stronger negative own price effects in the case of SL prices (-0.322, -0.221 and -0.091) than in the case of price indices (-0.141, -0.038 and -0.038). Fats has a positive own price effect in the case of price indices and a negative own price effect in the case of SL prices.

Table 2.5 lists results of the tests of rationality restrictions. We evaluate the estimated nonparametric demand functions and their derivatives at each data point, and for each observation we test whether the demand functions at that point satisfy homogeneity, negative semidefiniteness, and Slutsky symmetry. A rejection means that the original condition, e.g., negative semidefiniteness in the underlying heterogeneous population, cannot hold in the neighborhood of the point where we have tested. Although this gives an accurate picture of the behavior on individual level, it generates a flood of results, and we aggregate across the population as a means of condensing the result. For each hypothesis, Table 4.5 lists the percent of observations at which the hypothesis is not rejected at the 0.95 confidence level.

Firstly, in contradiction to the parametric results, homogeneity is well accepted, we cannot reject in 93 % in the case of SL prices and 99 % in the case of price indices. This is very comforting as there is no way that negative semidefiniteness can hold without homo-

geneity of degree zero holding, too. Secondly, we cannot reject negative semidefiniteness in 69 % in the case of SL prices and in 0.6 % in the case of price indices. The inclusion of endogeneity alters the latter result slightly. Negative semidefiniteness is not rejected in 70 % in the case of SL prices and in 0.3 % in the case of price indices. From this result it is obvious that negative semidefiniteness is not rejected only if we incorporate heterogeneity into the prices, i.e. use SL prices. Similarly to the parametric results, symmetry is not well accepted in general. We do not reject in 16 % to 17 % in the case of SL prices and in 21 % to 23 % in the case of price indices. Exogeneity and endogeneity do not generate materially different results. Again, we observe a smaller rejection rate under absence of demographic price variation. Finally we provide the percentage of households satisfying all restrictions. These are only 0.3 % in the case of price indices and 6 % in the case of SL prices. The inclusion of endogeneity diminishes the latter result slightly. We cannot reject all hypotheses in 0.2 % in the case of price indices and 5.7 % in the case of SL prices. Taking all results together, one could conclude that rationality is rejected. However we have to consider that the rationality concepts described above are applicable to individuals, but not implicitly to many-person households. There is mounting empirical evidence that the unitary model is regularly rejected on household data, e.g. yields symmetry violations, see Browning and Chiappori (1998).

## 7 Summary

In this paper we are concerned with the performance Stone Lewbel cross section prices, a device that allows to increase the cross section variability of prices by exploiting the fact that consumption for various goods within a given bundle of goods varies across individuals. Specifically, a theorem by Lewbel (1989) allows to construct individual specific prices that vary across the population. In the typical situation in demand data, where there is a wealth of cross section variation, but relatively little time series variation, this is a crucial advantage.

The focus of this paper is applied. We analyze the performance of SL prices across a variety of scenarios ranging from standard applied parametric setups to recent more theory heavy nonparametric setups. We interpret our evidence as establishing universally the advantages of SL prices in all settings. The regressions using SL prices are not just more plausible in terms of the sign of the coefficients obtained. Also, the precision of

our parameter estimates improves dramatically. In so far as this has as consequence that economic theory is rejected we conclude that SL prices allow a more decisive look at the data. However, most point estimates obtained actually show that economic theory fares better, even in light of the higher precision. In particular, negative semidefiniteness, arguably the core property of consumer rationality, appears to be better supported by the data now. We conclude that, wherever they can be constructed, SL prices may very well be a way out of the problem of insufficient and nonstationary price variation in practise, and recommend their use.

## References

- [1] BLUNDELL, R., CHEN, X. and KRISTENSEN, D. , 2007, Semi-nonparametric IV Estimation of Shape-invariant Engel Curves, Forthcoming in *Econometrica*.
- [2] BROWNING, M., and P.-A. CHIAPPORI, 1998, Efficient Intra-household Allocations: A General Characterization And Empirical Tests, *Econometrica*, 66, 1241-1278.
- [3] DEATON, A., and J. MUELLBAUER, 1980a, An Almost Ideal Demand System, *American Economic Review*, 70, 312-26.
- [4] DEATON, A., and J. MUELLBAUER, 1980b, Economics and Consumer Behavior, *New York: Cambridge University Press*.
- [5] HAAG, B. and HODERLEIN, S., 2006, Testing Economic Hypotheses in Systems of Equations, *Working Paper*, Mannheim University.
- [6] HAAG, B., S. HODERLEIN and K. PENDAKUR, 2007, Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems, *Working Paper*, Mannheim University.
- [7] HODERLEIN, S., 2007, How Many Consumer are Rational? *Working Paper*, Mannheim University.
- [8] HODERLEIN, S., J. KLEMELÄ and E. MAMMEN, 2007, Reconsidering the Random Coefficient Model, *Working Paper*, Mannheim University.
- [9] HODERLEIN, S. and A. LEWBEL, 2007, Price Dimension Reduction in Demand Systems With Many Goods, *Working Paper*, Mannheim University.
- [10] IMBENS, G. and W. NEWEY, 2007, Identification and Estimation of Triangular Simultaneous Equations Models without Additivity, *Working Paper*, MIT.
- [11] LEWBEL, A., 1989, Identification and Estimation of Equivalence Scales under Weak Separability, *Review of Economic Studies* 56, 311-16.
- [12] LEWBEL, A., 1999, Consumer Demand Systems and Household Expenditure, in PESARAN, H. and M. WICKENS (Eds.), *Handbook of Applied Econometrics*, Blackwell Handbooks in economics.

- [13] LEWBEL, A., 2001, Demand Systems With and Without Errors, *American Economic Review*, 611-18.
- [14] MAS-COLELL, A., M. WHINSTON, and J. GREEN, 1995, *Microeconomic Theory*, Oxford University Press.
- [15] WOOLDRIDGE, J., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

## Appendix: Empirical Results

Table 1.1: Estimated Uncompensated Price Elasticities and Budget Elasticities under Endogeneity using Price Indices

	elasticities carbo-price	elasticities meat price	elasticities fats price	elasticities drinks price	elasticities totexp
Carbo	-0.9613 (0.0990)	-0.8449 (0.1701)	0.3894 (0.2206)	0.1787 (0.1180)	0.9793 (0.0430)
Meat	0.2457 (0.1889)	-0.0047 (0.3245)	-0.9466 (0.4209)	-0.2777 (0.2252)	1.0557 (0.0820)
Fats	0.0910 (0.1242)	-0.4906 (0.2134)	0.0561 (0.2768)	-0.6455 (0.1481)	1.0709 (0.0539)
Drinks	-0.2499 (0.1807)	-0.5504 (0.3104)	1.2906 (0.4026)	-0.6357 (0.2154)	0.9835 (0.0785)

Table 1.2: Estimated Uncompensated Price Elasticities and Budget Elasticities under Endogeneity using SL Prices

	elasticities carbo-price	elasticities meat price	elasticities fats price	elasticities drinks price	elasticities totexp
Carbo	-1.4670 (0.0550)	0.0204 (0.0514)	-0.2090 (0.0301)	-0.1791 (0.0612)	0.9835 (0.0448)
Meat	-0.2734 (0.1071)	-1.6828 (0.1001)	-0.3956 (0.0587)	0.3581 (0.1191)	1.0575 (0.0871)
Fats	-0.3953 (0.0704)	-0.2997 (0.0657)	-0.4215 (0.0386)	-0.0537 (0.0782)	0.9562 (0.0572)
Drinks	0.3681 (0.1034)	0.3661 (0.0967)	0.2284 (0.0567)	-1.2678 (0.1150)	0.9861 (0.0842)

Table 1.3: Slutsky matrix (Compensated Effects of Price Indices) under Endogeneity

	carbo	meat	fats	drinks
carbo	-0.1660	-0.1674	0.1368	0.0667
meat	0.0601	0.0146	-0.1051	-0.0260
fats	0.0464	-0.0619	0.0304	-0.0925
drinks	0.0023	-0.0353	0.1305	-0.0472

Table 1.4: Slutsky matrix (Compensated Effect of SL Prices) under Endogeneity

	carbo	meat	fats	drinks
carbo	-0.2911	0.0380	-0.0093	-0.0206
meat	-0.0074	-0.2047	-0.0327	0.0573
fats	-0.0356	-0.0330	-0.0490	0.0017
drinks	0.0576	0.0460	0.0371	-0.1028

Table 1.5: Parametric Model Homogeneity Tests for Each Equation under Endogeneity

	Price Indices		SL Prices	
	test statistic	p-value	test statistic	p-value
carbo	0.9431	0.3315	78.9364	0.0000
meat	0.0013	0.9718	29.5578	0.0000
fats	0.0013	0.9714	2.0102	0.1562
drinks	3.6456	0.0562	15.4915	0.0001

Table 1.6: Parametric Model Systemwide Tests under Endogeneity

	Price Indices		SL Prices	
	test statistic	p-value	test statistic	p-value
homogeneity	4.3914	0.3556	149.7909	0.0000
symmetry	52.7783	0.0000	28.4355	0.0001

Table 2.1: Nonparametrically Estimated Uncompensated Price Elasticities and Budget Elasticities under Endogeneity using Indices

	elasticities carbo-price	elasticities meat price	elasticities fats price	elasticities drinks price	elasticities totexp
Carbo	-0.866 (-1.032; -0.717)	0.058 (-0.324; 0.408)	0.169 (-0.267; 0.578)	0.015 (-0.244; 0.311)	0.977 (0.964; 0.989)
Meat	0.019 (-0.429; 0.421)	0.403 (-1.199; 0.346)	-0.815 (-1.688; 0.027)	-0.920 (-1.423; -0.298)	1.100 (1.071; 1.127)
Fats	-0.026 (-0.374; 0.205)	-0.769 (-1.348; -0.183)	0.368 (-0.352; 0.844)	-0.765 (-1.160; -0.354)	0.852 (0.835; 0.867)
Drinks	-0.110 (-0.43; 0.215)	0.293 (-0.293; 0.968)	1.790 (0.728; 2.488)	-0.525 (-1.047; -0.031)	0.926 (0.905; 0.950)

Numbers in parentheses are lower and upper bounds of the 95% confidence interval

Table 2.2: Nonparametrically Estimated Uncompensated Price Elasticities and Budget Elasticities under Endogeneity using SL Prices

	elasticities carbo-price	elasticities meat price	elasticities fats price	elasticities drinks price	elasticities totexp
Carbo	-1.543 (-1.687; -1.488)	0.028 (-0.138; 0.144)	-0.213 (-0.262; -0.142)	-0.248 (-0.347; -0.141)	0.968 (0.955; 0.982)
Meat	-0.434 (-0.579; -0.262)	-1.967 (-2.203; -1.504)	-0.471 (-0.573; -0.293)	0.549 (0.307; 0.796)	1.098 (1.070; 1.122)
Fats	-0.196 (-0.317; -0.133)	-0.397 (-0.640; -0.226)	-0.461 (-0.540; -0.341)	-0.149 (-0.340; -0.001)	0.815 (0.795; 0.834)
Drinks	0.192 (-0.066; 0.409)	0.280 (-0.105; 0.454)	0.289 (0.111; 0.412)	-1.207 (-1.304; -0.858)	0.930 (0.910; 0.960)

Numbers in parentheses are lower and upper bounds of the 95% confidence interval

Table 2.3: Slutsky Matrix (Compensated Effects of Price Index) under Endogeneity

	carbo	meat	fats	drinks
carbo	-0.1414	0.0820	0.0733	0.0229
meat	0.0391	-0.0372	-0.0699	-0.0971
fats	0.0264	-0.1112	0.0568	-0.1169
drinks	0.0142	0.0405	0.1648	-0.0360

Table 2.4: Slutsky Matrix (Compensated Effects of SL-Prices) under Endogeneity

	carbo	meat	fats	drinks
carbo	-0.3213	0.0453	-0.0153	-0.0334
meat	-0.0229	-0.2214	-0.0376	0.0908
fats	0.0024	-0.0485	-0.0590	-0.0092
drinks	0.0345	0.0347	0.0330	-0.0911

Table 2.5: Nonparametric Model Tests

Percent of observations not rejecting each hypothesis at the 95% confidence level

<i>Hypothesis</i>	<i>% not rejected</i>	
	<i>Indices</i>	<i>SL-Prices</i>
Negative Semidefiniteness under Exogeneity	0.6%	69%
Negative Semidefiniteness under Endogeneity	0.3%	70%
Homogeneity under Exogeneity	99%	93%
Homogeneity under Endogeneity	99%	93%
Symmetry under Exogeneity	21%	17%
Symmetry under Endogeneity	23%	16%
All Hypotheses under Exogeneity	0.3%	6%
All Hypotheses under Endogeneity	0.2%	5.7%

## Appendix: Data

Table 3: Descriptive statistics

	10% quantile	mean	median	90% quantile	stdev
CARBO	0.1493	0.2398	0.2330	0.3380	0.0783
bread	0.0753	0.2134	0.1892	0.3754	0.1350
potatoes	0.0487	0.2393	0.2199	0.4495	0.1570
fruit	0.0151	0.2430	0.2287	0.4667	0.1655
oth.vegs	0.1194	0.3042	0.2970	0.4947	0.1474
MEAT	0.0378	0.1309	0.1149	0.2440	0.0843
beef	0.0000	0.2564	0.1740	0.6811	0.2883
lamb	0.0000	0.0984	0.0000	0.3705	0.1978
pork	0.0000	0.1244	0.0000	0.4134	0.2051
bacon	0.0000	0.2051	0.1239	0.5502	0.2578
fish	0.0000	0.3157	0.2242	1.0000	0.3234
FATS	0.0819	0.1614	0.1535	0.2492	0.0692
butter	0.0000	0.0555	0.0000	0.1803	0.0985
oil.fats	0.0000	0.1037	0.0750	0.2553	0.1176
cheese	0.0000	0.2047	0.1818	0.4362	0.1733
eggs	0.0000	0.0742	0.0539	0.1794	0.0927
milkfres	0.1332	0.3947	0.3799	0.6728	0.2106
milkprod	0.0000	0.1672	0.1298	0.3989	0.1654
DRINKS	0.0310	0.0884	0.0791	0.1549	0.0531
tea	0.0000	0.1957	0.0681	0.6013	0.2736
coffee	0.0000	0.2015	0.0000	0.6133	0.2694
softdrin	0.0214	0.6028	0.6445	1.0000	0.3348

Table 3: Descriptive statistics (cont.)

	10% quantile	mean	median	90% quantile	stdev
SL-price carbo	6.1059	6.2381	6.2244	6.3882	0.1121
SL-price meat	5.9257	6.0144	6.0241	6.0806	0.0707
SL-price fats	6.2067	6.3528	6.3715	6.4800	0.1164
SL-price drinks	6.4149	6.4879	6.4946	6.5583	0.0599
price index carbo	6.1784	6.2343	6.2319	6.2932	0.0417
price index meat	5.9550	6.0107	6.0218	6.0453	0.0331
price index fats	6.3261	6.3572	6.3613	6.3768	0.0206
price index drinks	6.4221	6.4885	6.4918	6.5388	0.0439
log foodexp	2.8386	3.6214	3.6835	4.3141	0.5750
log income	4.6334	5.5821	5.6606	6.4637	0.9403
number adults	1.0000	1.8257	2.0000	3.0000	0.6830
number male	0.0000	0.8531	1.0000	1.0000	0.5536
number retired	0.0000	0.3613	0.0000	1.0000	0.6450
number earners	0.0000	1.0557	1.0000	2.0000	0.9619
number kids	0.0000	0.6978	0.0000	2.0000	1.0561
fridge	1.0000	0.9943	1.0000	1.0000	0.0756

Table 6 presents descriptive statistics of the data. First we list 10 % quantile, mean, median, 90 % quantile and standard deviation of the budget shares. The budget shares of the groups (between budget shares) are denoted by capital letters. The budget shares of the goods in a group relative to total expenditures in the group (within budget shares) are denoted by small letters. Then we report descriptive statistics of the SL prices and priceindices of all groups, log food expenditures, log labor income and the demographic characteristics described in section 4.

## Appendix: Graphs

Figure 1: Mean regression budget share "carbohydrates" on log food expenditure with 90% confidence band width

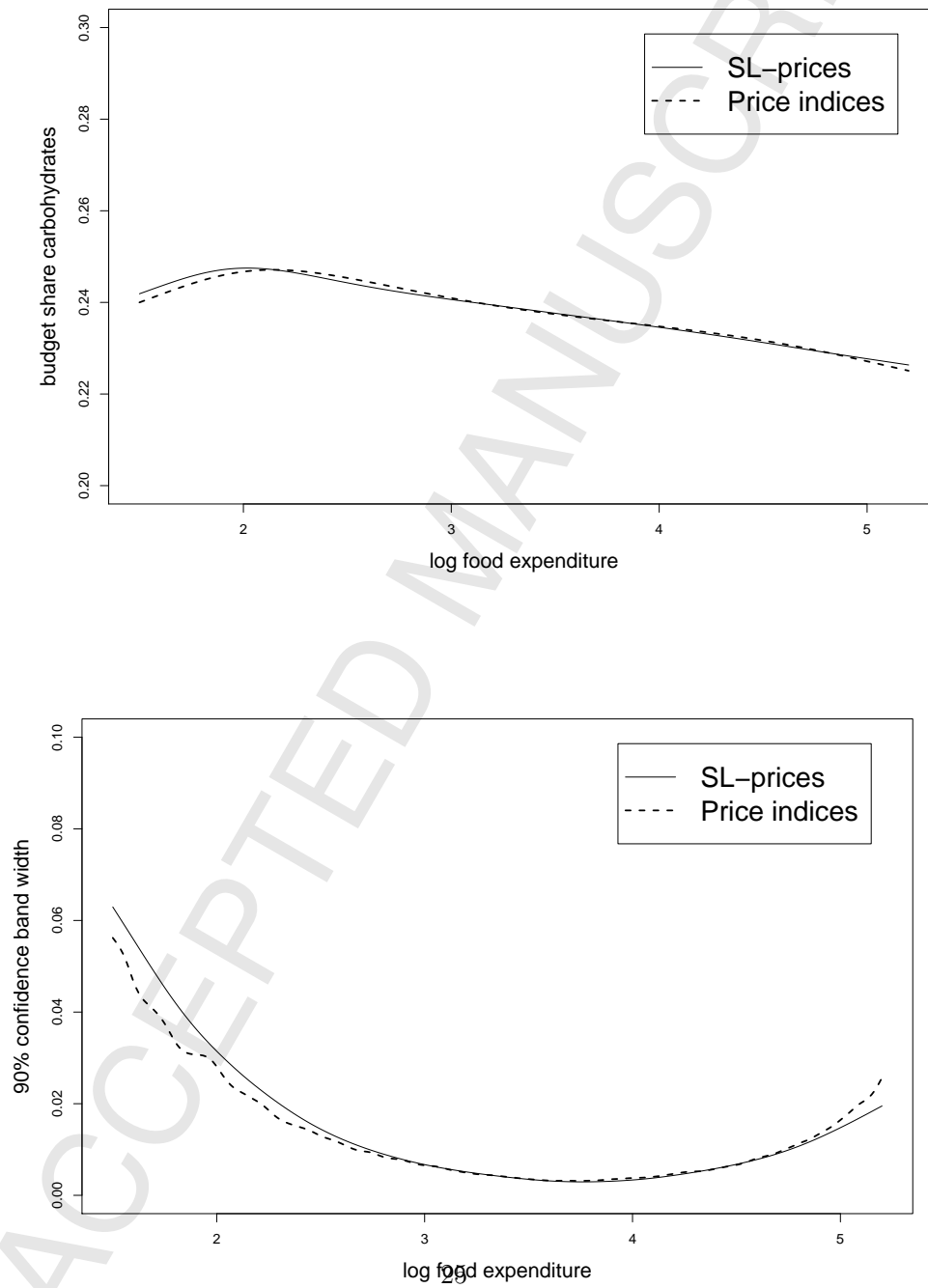


Figure 2: Mean regression budget share "meat" on log food expenditure with 90% confidence band width

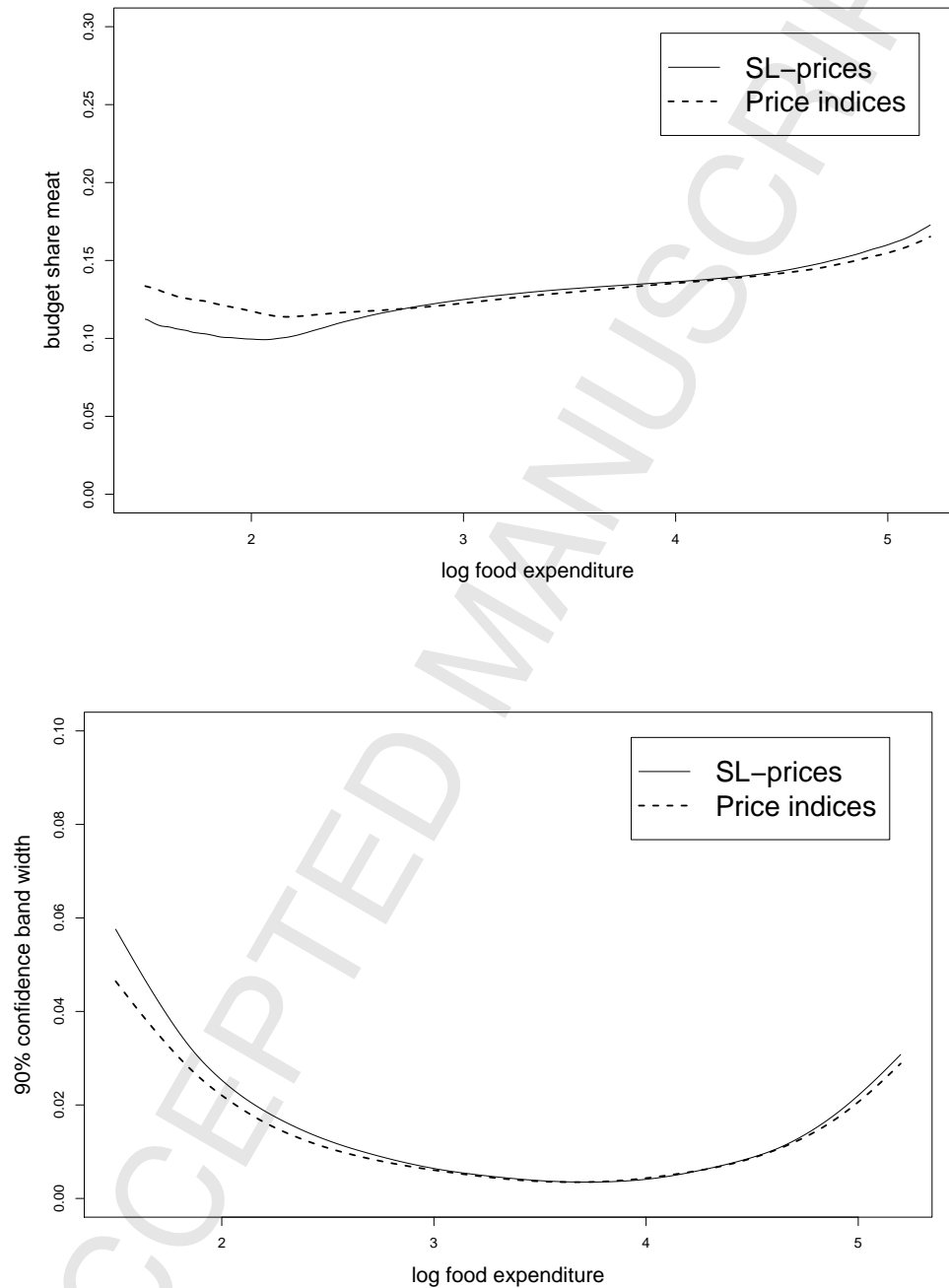


Figure 3: Mean regression budget share "fats" on log food expenditure with 90% confidence band width

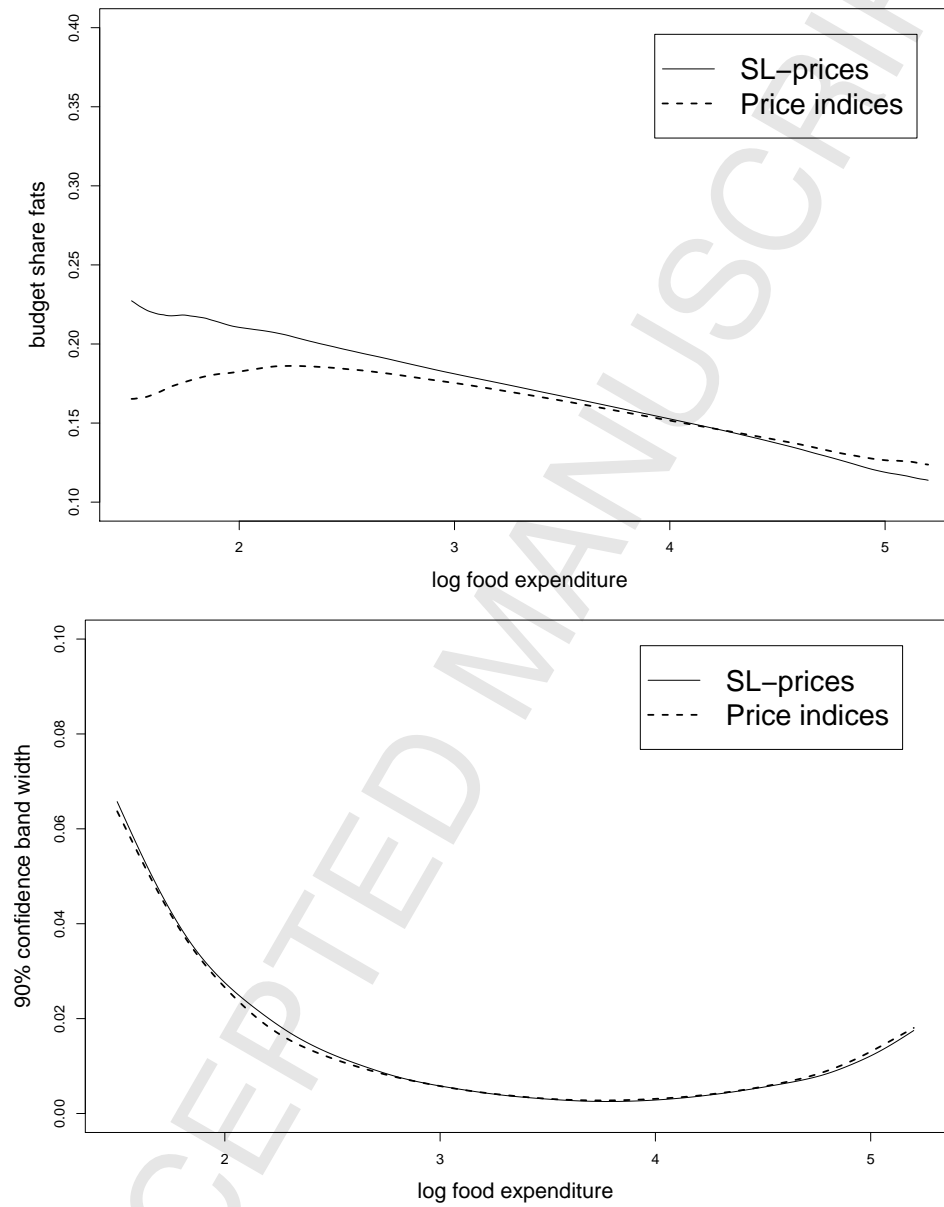


Figure 4: Mean regression budget share "drinks" on log food expenditure with 90% confidence band width

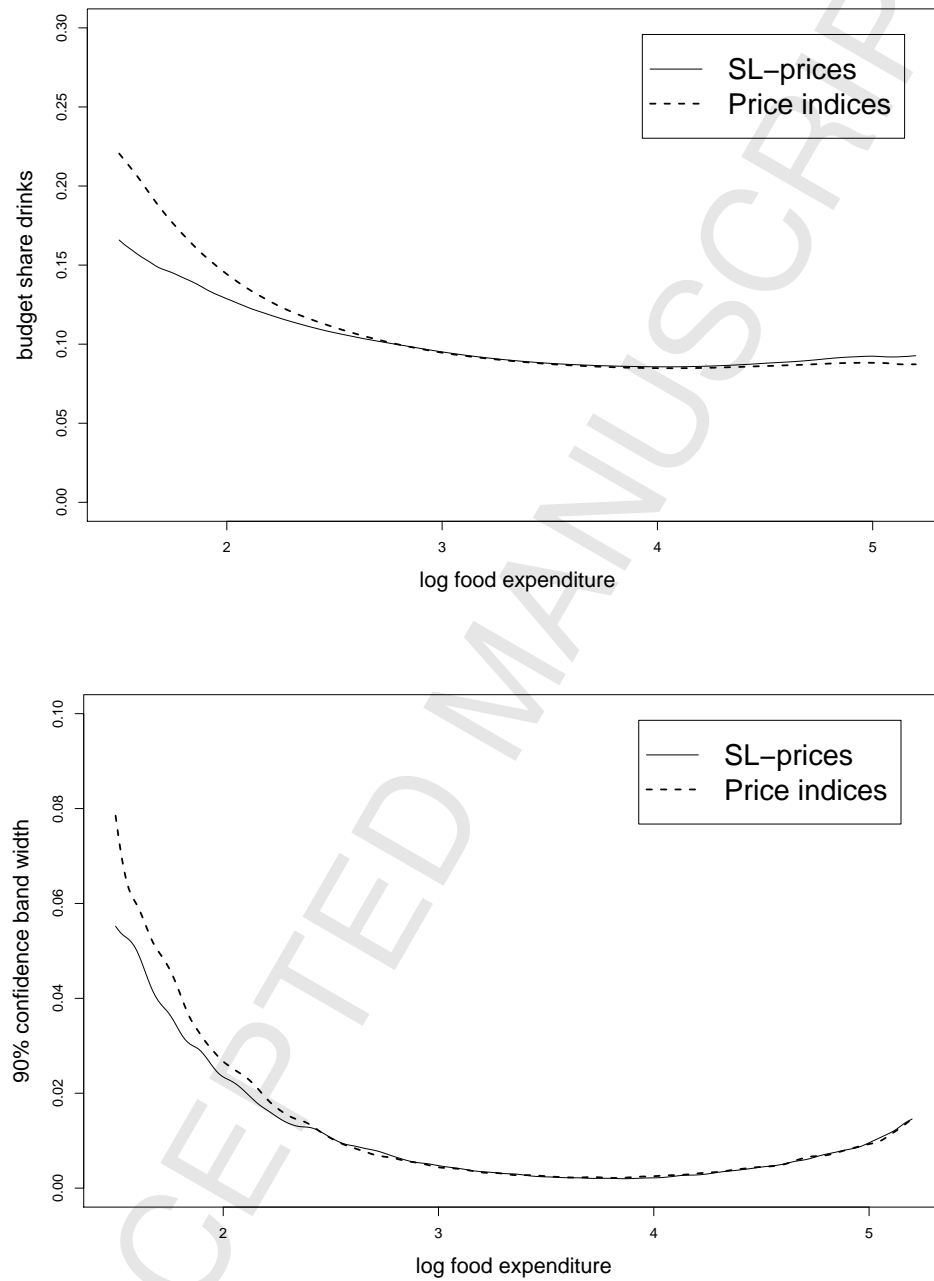


Figure 5: Compensated own price effect "carbohydrates" on log food expenditure with 90% confidence band width

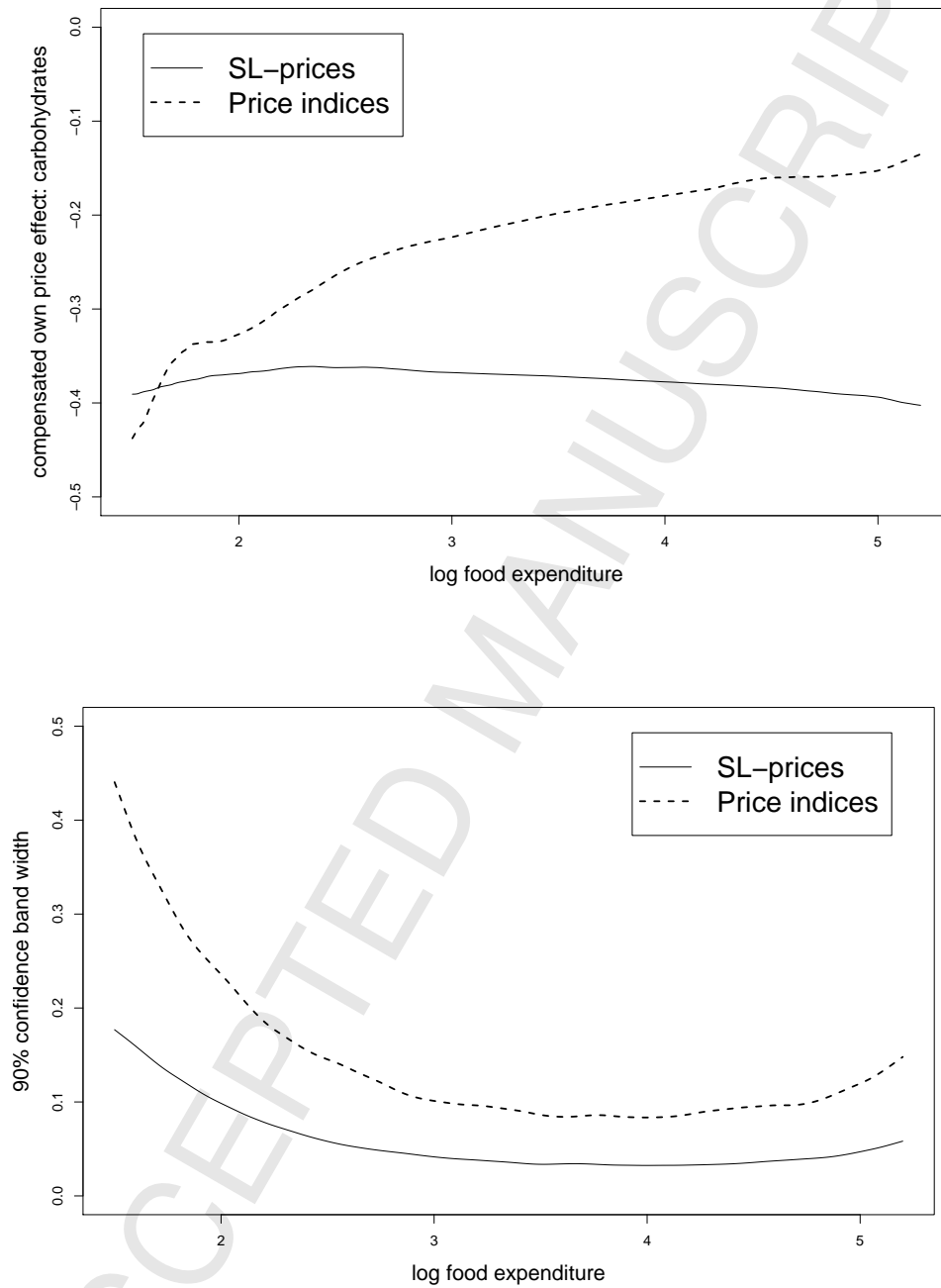


Figure 6: Compensated own price effect "meat" on log food expenditure with 90% confidence band width

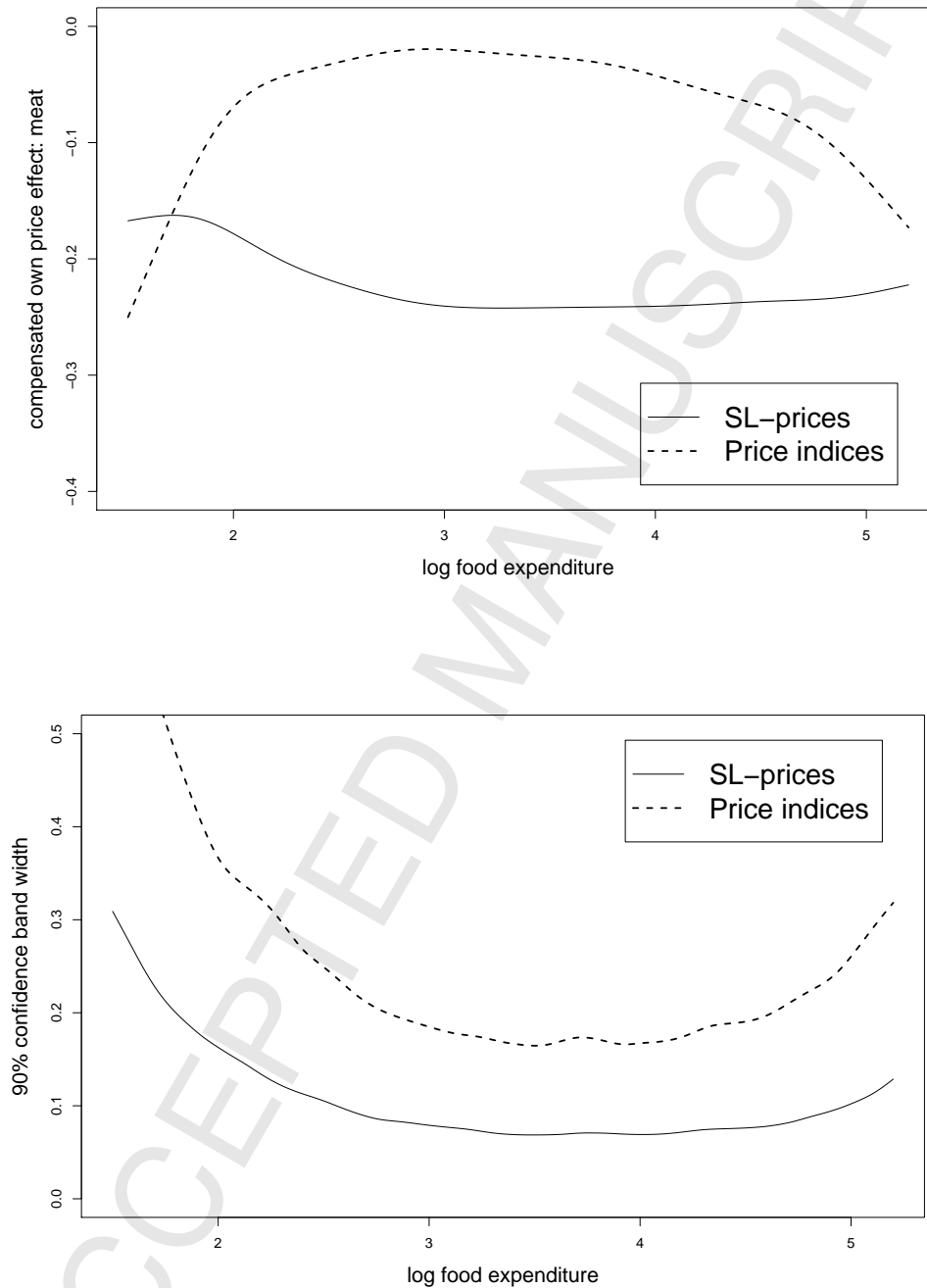


Figure 7: Compensated own price effect "fats" on log food expenditure with 90% confidence band width

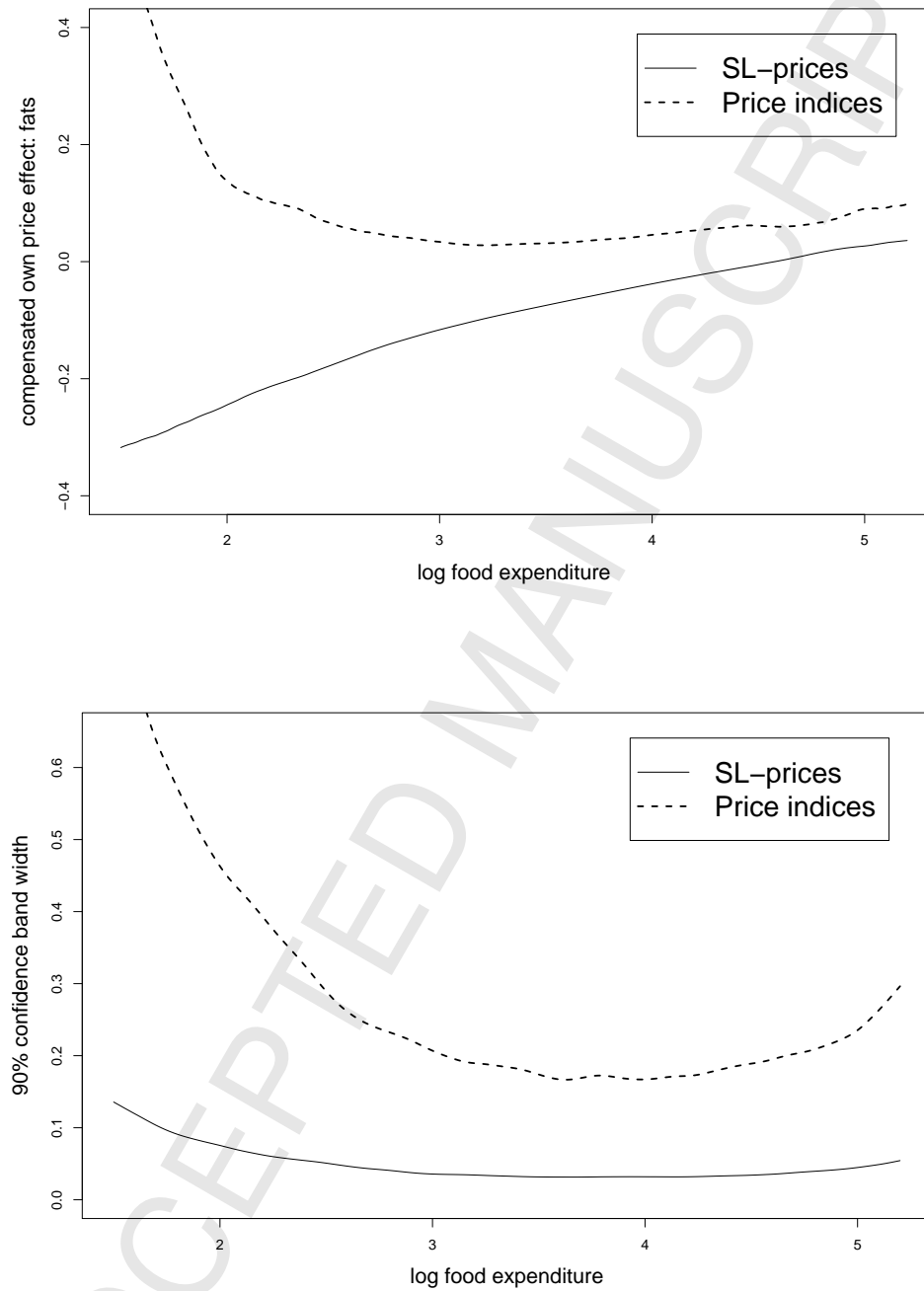


Figure 8: Compensated own price effect "drinks" on log food expenditure with 90% confidence band width

