



Hoderlein, Stefan und Winter, Joachim:
Structural Measurement Errors in Nonseparable
Models

Munich Discussion Paper No. 2009-2

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at <https://doi.org/10.5282/ubm/epub.9192>

Structural Measurement Errors in Nonseparable Models*

Stefan Hoderlein Joachim Winter
Brown University University of Munich

January 29, 2009

Abstract

This paper considers measurement error from a new perspective. In surveys, response errors are often caused by the fact that respondents recall past events and quantities imperfectly. We explore the consequences of recall errors for such key econometric issues as the identification of marginal effects or economic restrictions in structural models. Our identification approach is entirely nonparametric, using Matzkin-type nonseparable models that nest a large class of potential structural models. We establish that measurement errors due to poor recall are generally likely to exhibit nonstandard behavior, in particular be nonclassical and differential, and we provide means to deal with this situation. Moreover, our findings suggest that conventional wisdom about measurement errors may be misleading in many economic applications. For instance, under certain conditions left-hand side recall errors will be problematic even in the linear model, and quantiles will be less robust than means. Finally, we apply the main concepts put forward in this paper to real world data, and find evidence that underscores the importance of focusing on individual response behavior.

Keywords: Measurement Error, Nonparametric, Survey Design, Nonseparable Model, Identification, Zero Homogeneity, Demand.

1 Introduction

Motivation. Consider a question in a household survey about an economic quantity like the following: “How much did you spend on food items in the last month?” We will call the

*Stefan Hoderlein (corresponding author): Department of Economics, Brown University, Department of Economics, Box B, E-Mail: stefan_hoderlein@yahoo.com. Joachim Winter: Department of Economics, Ludwigstr. 28 RG, 80539 Munich, Germany, E-Mail: winter@lmu.de. We have received helpful comments from Norbert Christopeit, Arthur Lewbel, Charles Manski, Jörg Stoye and two anonymous referees. We would like to particularly thank Daniel McFadden for encouraging us to pursue this line of research. Financial support by Landesstiftung Baden-Württemberg “Eliteförörderungsprogramm” and the European Union (grant no. ERBFMRCT960016) is gratefully acknowledged.

target quantity of a survey question Z . Most often, individuals do not know the exact answer and make an informed guess \tilde{Z} , which can be viewed as a “backcast” using the immediately available information. Obviously, this recall \tilde{Z} is only an imperfect measurement of true food consumption Z , and any analysis based on this imperfect measurement may potentially be flawed.

Traditionally, the econometrics literature has focussed on the error in the measurement of Z , i.e., on $Z - \tilde{Z}$, when dealing with such situations. In this paper we reverse the viewpoint and focus on properties of the backcast \tilde{Z} instead. Unlike in parts of the natural sciences where (mis)measurements are given exogenously, in surveys in the social sciences individuals actively decide about their response. From this perspective, measurement errors are only a byproduct of this process. Consequently, we argue in this paper that the survey design and measurement error literatures can gain important new insights by focusing on the properties of the backcast \tilde{Z} rather than concentrating on properties of the response error.

Reversing the viewpoint in this way has several advantages: First, viewing the object from a different perspective allows us to gain new insights. For instance, it is well documented in validation studies that left-hand side measurement errors are not as well behaved as they are in the textbook (classical measurement error) case. In this paper, we give conditions for when left-hand side and right-hand side measurement errors are likely to exhibit “nonstandard” behavior, and provide explanations. Second, armed with a better understanding of the measurement error, it is possible to investigate the design of survey questions formally, and to minimize the distortions arising from response behavior *ex ante* rather than having to deal with them *ex post* (see also McFadden, et al., 2006, and McFadden, 2006). We emphasize the latter point because there is now ample experimental and empirical evidence suggesting important effects of the design of a survey on the actual response behavior of individuals.

Throughout this paper, we concentrate on distorting effects that are caused by the respondent. We neither consider effects caused by (social) interaction between the interviewer and the respondent, nor do we consider the case where Z is elicited accurately but is used only as a proxy for another variable in an econometric model. Specifically, we are concerned with the respondents’ limited ability to acquire, process, and recall information. By focusing on this key type of measurement error, we can allow for heterogeneity of individuals in preferences and mental abilities in a very general fashion.

Related literature. There is a significant amount of work in econometrics on mismeasured variables. Within this body of literature, our approach contributes to the line of work that discusses measurement errors in nonlinear models, and contributes more generally to work that explores nonparametric identification using nonseparable models. Key recent contributions in the nonlinear measurement error literature are, *inter alia*, Hausman, Newey, and Powell (1995) for classical measurement errors, and Hu and Schennnach (2006) for nonclassical measurement errors. In this literature, Matzkin (2008) is most closely related to our work in terms of the nonseparable framework, see also Matzkin (2007) for a survey of these type of models. For additional literature, see references in the cited papers as well as the monograph of Wansbeek

and Meijer (2000), the extensive survey of Bound, Brown, and Mathiowetz (2001), and the conceptual paper by Hausman (2001). Compared with our approach, this strand of the measurement error literature is of fairly reduced form. There is no behavioral model that explains the respondents' error, and consequently the error is a purely statistical object. The advantages of either approach are obvious: While the reduced form approach provides a broader safeguard against all types of measurement errors, our structural approach is more restrictive as it only considers one type of error, although perhaps an important one. However, focusing on only one type of measurement error allows us to describe more accurately its determinants and implications.

Our work shows some overlap with the parametric work of Hyslop and Imbens (2001), who as we argue that the measurement literature in social sciences should rather view measurement errors as a byproduct of a forecasting problem. While we share this similarity in philosophy, our approach is markedly different in several dimensions. First, we consider general nonparametric identification, estimation and testing results, including nonparametric quantile and mean regressions. In contrast, Hyslop and Imbens (2001) consider a restrictive linear setting with normally distributed errors. Moreover, in our more general setting we provide conditions for when recall errors are relatively well behaved, and when they are less well behaved, and we provide bounds as a possible remedy in the case of less well behaved recall errors.

From the general philosophy, our approach is more closely related to work in survey research and psychology (for a review see Tourangeau, Rips, and Rasinski, 2000). Indeed, our work allows to link this literature to econometrics via nonparametric identification. In a companion paper, we provide a behavioral model for the recall process, which helps to clarify conditions under which the error is less well behaved (see Hoderlein and Winter, 2008). There is also a strand of work in economic theory which considers the response behavior under a market perspective, see Philipson (1997, 2001), but McFadden (2006) is the only paper we know of that aims at unifying insights from the psychological literature on survey response behavior and an incentive-based economic analysis of survey response in an econometric setup. Finally, conceptually there is some overlap between our paper and the work of Manski and his collaborators, see Manski (2004) for an overview.

Key issues and contributions. In considering the behavior of respondents in detail and relating it to nonparametric identification, we aim specifically at the following issue: *Under which conditions can we identify marginal effects of the economic relationship of interest?* As a minimal requirement, can we at least identify some average structural marginal effects in a world of heterogeneous agents? Indeed, if these quantities are not identified, not much scope for any type of econometric analysis remains. Also, are conditional means or conditional quantiles better suited to deal with difficulties arising in this situation?

To illustrate the contributions in this paper related to these questions, we illustrate some important issues through the following textbook example of a linear model¹. We assume that

¹Throughout the rest of the paper, many of the ideas will be discussed in an abstract fashion with the intention to show that these ideas apply generically to a large class of models, e.g. to nonlinear models or

the model is defined as follows: Let

$$Y = X'\beta + A, \quad (1.1)$$

where Y is a random scalar, β is a fixed K -vector of coefficients, X is a random K -vector of regressors, and A is a random scalar such that $\mathbb{E}[A|X] = 0$. Due to limitations in the individual's ability to recall, we do not observe Y , but rather \tilde{Y} , which we model as a function of the variables which span the individuals' information set \mathcal{F} .

What are the implications? Since the individual answers \tilde{Y} , instead of (1.1) we use the regression $\tilde{Y} = X'\tilde{\beta} + \tilde{A}$, where \tilde{A} is defined through $\mathbb{E}[\tilde{Y}|X] = X'\tilde{\beta}$. By standard arguments, $\tilde{\beta} = \mathbb{E}[XX']^{-1} \mathbb{E}[X\tilde{Y}]$, provided $\mathbb{E}[XX']$ is nonsingular. Suppose first that when backcasting individuals use the squared error loss, and assume that $\mathcal{F} \supseteq \sigma(X)$. Then, $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}]$, and

$$\tilde{\beta} = \mathbb{E}[XX']^{-1} \mathbb{E}[X\mathbb{E}[Y|\mathcal{F}]] = \beta + \mathbb{E}[XX']^{-1} \mathbb{E}[X\mathbb{E}[A|\mathcal{F}]].$$

But due to $0 = \mathbb{E}[A|X] = \mathbb{E}[\mathbb{E}[A|\mathcal{F}]|X]$ and the law of iterated expectations, the second term vanishes and consequently $\tilde{\beta}$ coincides with β . This case is analogous to the conventional wisdom that left hand side measurement errors do not matter for identification in linear models.

However, turn to the scenario where individuals do not actively use all information available to the econometrician. Formally, consider the case where in order to form a backcast individuals use the squared error loss in connection with $\mathcal{F} = \sigma(X'\gamma, A_1)$, where γ is a nonrandom nonzero vector and A_1 contains additional information not available to the econometrician. Assume for simplicity that $\mathbb{E}[A|\mathcal{F}] = \mathbb{E}[A|A_1] \neq 0$ and that A_1 is independent of X . Then,

$$\tilde{\beta} = \mathbb{E}[XX']^{-1} \mathbb{E}[X\mathbb{E}[Y|\mathcal{F}]] = \left\{ I_K - \mathbb{E}[XX']^{-1} \mathbb{E}[XV'] \right\} \beta,$$

where $V = X - \mathbb{E}[X|X'\gamma]$. Obviously, the second term in curly brackets will not vanish in general. Consequently, even though one may identify and consistently estimate $\tilde{\beta}$, this will only produce a biased version of β .

What is going wrong with conventional wisdom here? The answer is that the measurement error is nonstandard in the sense that $U = Y - \tilde{Y}$ is not independent of Y . More importantly for the linear model, $\mathbb{E}[U|X] \neq 0$, i.e. the regressors become endogenous.

The question of whether individuals actively use all information employed by the econometrician in their backcast seems to be the crucial question more generally. What differs in the general nonparametric case, however, is that only average structural effects are identified, even in the case where individuals use more information than the econometrician. Hence it is impossible to disentangle the heterogeneity in preferences from the heterogeneity in recall abilities. Moreover, it is our opinion that there will always be limitations in the quality of the backcast and that the recall errors are bound to be misbehaved. For instance, it is straightforward to see

nonparametric models defined by quantile restrictions. Hence, we will make largely use of nonseparable models where the dependent variable is a smooth, but otherwise fairly unrestricted function of a K -random vector of regressors X , and a (potentially infinite dimensional) vector of unobservables A .

that in general recall errors on right hand side variables X are differential in the sense of Carroll, Ruppert and Stefanski (1995). The reason is that the additional information individuals use will generally cause dependence between Y and the backcast \tilde{X} , even after controlling for X . Hence, some way of dealing with this problem is called for: We explore the standard econometric remedy of instruments, but argue that this route may be treacherous as individuals have to employ the information in the instruments in their backcast. As an alternative, we propose bounds. Using a simplified version of the bounds provided in theorem 2 and specializing it to the linear model (1.1) with poorly recalled Y (and the added assumption that $A = \sigma(X)P$, where σ is a smooth function and P is independent of X), we obtain for the first coefficient β_1 :

$$|\mathbb{E}[\partial_{x_1} m(X)] - \beta_1| \leq \mathbb{E}[Y^2]^{1/2} \mathbb{E}[Q_{x_1}^2]^{1/2},$$

where m denotes the nonparametric regression of \tilde{Y} on $X = (X_1', X_{-1}')'$, and $\partial_{x_1} g$ denote the partial derivative of a vector valued function g with respect to a scalar x_1 , $f_{X_{-1}|X_1}(x_{-1}, x_1)$ denotes the conditional density, and $Q_{x_1} = \partial_{x_1} \log f_{X_{-1}|X_1}(X_{-1}, X_1)$. Finally, whenever convenient we suppress the arguments of the respective functions.

There are a couple of things worthwhile noting, all of which will be shown below in generality: First, the term $\mathbb{E}[Y^2]$ reflects knowledge of the true population distribution of Y . Hence, we do not require a validation dataset, but we do require a correct aggregate statistic. Second, if we are willing to invoke an assumption about the variance of the error, $\mathbb{E}[U^2]$, then we may replace $\mathbb{E}[Y^2]$ by $\mathbb{E}[\tilde{Y}^2] + \mathbb{E}[U^2]$. Third, we may come up with alternative bounds which replace $\mathbb{E}[Y^2]$ by $\mathbb{E}[\text{Var}[Y^2|\mathcal{F}]]$, where $\text{Var}[Y^2|\mathcal{F}]$ may be obtained by eliciting a measure of the individuals' uncertainty. This line of thought connects our approach to that of Manski (2004). Fourth, the assumption about the structure of the recall information shows up in the second term $\mathbb{E}[Q_{x_1}^2]$, but we do not have to invoke any structural assumption about the measurement error. More specifically, the bounds are free of γ so we do not have to know how individuals exactly reduce the information. Finally, it is the average nonparametric regression and not the marginal effect $\tilde{\beta}_1$ for which we will derive natural bounds. The fact that individuals reduce information introduces a nonlinearity. If bounds on β_1 are to be obtained using $\tilde{\beta}_1$, then we also have to take into account a measure of distance of the nonparametric regression from the linear projection.

Organization of the paper. Throughout the paper, we are mainly interested in identification of the partial derivative of a structural economic model $Y = \phi(X, A)$ with respect to its first component X_1 , where Y, X and A are as above, but A may now be infinite dimensional. Now, due to imperfect recall we do not observe either Y or X_1 correctly. Instead, we observe \tilde{Y} or \tilde{X}_1 , which are the backcasts the individuals form, given a certain loss function and an information set. For instance, if respondents use the squared error loss and the information set $\sigma(X, J)$, where J denotes additional information, their backcast would be $\tilde{Y} = \mathbb{E}[Y|X, J]$. In this specific case for instance, the structural model is given by

$$\begin{aligned} Y &= \phi(X, A), \\ \tilde{Y} &= \mathbb{E}[Y|X, J], \end{aligned} \tag{1.2}$$

where $A \perp X$. In section 2, we consider the implications of such a model for the measurement of the partial derivative ∂_{x_1} . Because it is the most significant innovation, we consider only the case where the dependent variable Y is poorly recalled, and leave the case where the independent variable X_1 is poorly recalled to future work.

Our identification approach is entirely nonparametric. To show the important advantage that this directly allows to study all relationships between variables in a nonparametric fashion we retain an entirely nonparametric structure. More specifically, we briefly introduce nonparametric test statistics (section 3) and show in an empirical application that nonstandard recall errors can explain important negative findings in the demand literature (section 4). In particular, we show that the common rejection of homogeneity of degree zero may be related to recall errors in self reported consumption. Finally, summary and outlook conclude.

2 Identifying Marginal Effects in the Presence of Recall Errors

Throughout this paper, as a result of incomplete recollection of the dependent variable $Y \in \mathbb{R}$, we assume to observe the data $D = (\tilde{Y}, X', S')'$ only. Here, \tilde{Y} denotes either $\mathbb{E}[Y|\mathcal{F}]$ if individuals use the squared error loss, or the median $med_Y(\mathcal{F})$ if they use the absolute error loss, $X \in \mathbb{R}^K$ denotes the set of regressors in the economic relationship of interest, S denote variables that determine the precision of the recall (eg, time used or mental capacity), and we let $W = (X', S')'$. In the following, we will largely be concerned with the derivative with respect to a typical regressor. Hence, let $X = (X_1, X'_{-1})'$ define a partition of X , where X_1 is the variable of interest, for simplicity a scalar. Finally, denote by $A \in \mathcal{A}$ an unobservable possibly infinitely dimensional vector (element of a Borel space \mathcal{A}), and introduce the partitions $W = (X_1, W'_{-1})'$, i.e. $W_{-1} = (X'_{-1}, S')'$, and $A = (A'_1, A'_{-1})'$. Let F_G and $F_{G|H}$ be the cdf and conditional cdf of a random variable G given H . We assume that $F_{G|H}$ is absolutely continuous w.r.t. a measure ν , with $f_{G|H}$ the associated Radon-Nykodim derivative (a conditional density when ν is the Lebesgue measure, or a conditional probability if ν is the counting measure).

The first element of our analysis consists of a causal relationship between variables generated by an economic decision. As econometricians we are interested in estimating this relationship as well as in the validity of restrictions economic theory places upon it. For instance, in the case of demand for a single good, Y would be a log expenditure for the good, while X would be a K -vector of prices, income and individual characteristics. In this example, allowing A to be in a Borel space is very important as it allows for (unobserved) heterogeneity in piecewise continuous utility functions.

With these definitions we can characterize the economic relationship of interest:

Assumption 1. *The relationship of interest is*

$$Y = \phi(X, A),$$

where ϕ as a fixed, scalar valued Borel-measurable function defined on $\mathbb{R}^K \times \mathcal{A}$, continuously differentiable in x_1 , with derivatives $\partial_{x_1}\phi$ that are uniformly bounded.

Assumption 1 is typical for the literature on nonparametric identification and nonseparable functions. Note, however, that we explicitly allow for an infinite dimensional error and we do not impose monotonicity on the function ϕ . Analysis of this model has been considered by Altonji and Matzkin (2006), Imbens and Newey (2005), Hoderlein (2005, 2008) and Hoderlein and Mammen (2007). Like these papers, we will be interested in local average structural derivatives (LASD) with respect to some σ -algebra, e.g., $\mathbb{E}[\partial_{x_1}\phi(X, A)|X = x]$. This type of object reduces often to well known quantities, e.g., to β in the linear model².

The question we are interested is now: what can we learn from the data $D = (\tilde{Y}, X', S)'$ about the derivative with respect to a typical economic variable of interest, i.e. $\partial_{x_1}\phi$? To answer this question, we focus on the mean regression function $\mathbb{E}[\tilde{Y}|W = w] = m(w)$, as well as the conditional α -quantile of \tilde{Y} given $W = w$, denoted $k_\alpha(w)$, $\alpha \in (0, 1)$. And we want to define circumstances under which it is more appropriate to use one or the other.

As always in econometrics we require an identification restriction. Given the generality of our assumptions, it comes in the following form:

Assumption 2. *For the joint distribution F_D , either of the following restrictions hold:*

1. $F_{A|X_1 W_{-1}} = F_{A|W_{-1}}$.
2. $F_{A|XS} = F_{A|S}$.

Basically, the first assumption states that – conditional on recall factors S and other regressors X_{-1} – the regressor of interest X_1 and unobserved heterogeneity are independently distributed. To give an example: let Y be wage, \tilde{Y} reported wage, X_1 be experience in years and suppose we were given data of working individuals including the characteristics “type of work”, “gender” and, as an example for S , “time elapsed since decision”. Consider a specific subgroup of the population, e.g. female workers in the iron industry, for which the survey is elicited 20 days after their economic decision. Suppose there are only two kinds of individuals in this group, those who are in the company for a while (“long” experience), and those who recently joined (“short” experience). Suppose further that there are two types of unobservables, type 1 and 2, where the unobservables might include unobservable ability to recall things and unobservable job related abilities more generally. Then, for individuals with long and

²More precisely, if the underlying model is truly linear, i.e., $Y = X'\beta + \varepsilon$, with fixed K -coefficient β and (mean independent and heteroscedastic) scalar $\varepsilon = \Sigma(X)A$, where Σ is a smooth function and $A \perp X$ as well as $\mathbb{E}[A] = 0$, then $\mathbb{E}[\partial_{x_1}\phi(X, A)|X = x] = \beta_1$. If it is a random coefficient model, i.e., $Y = X'\beta(A)$ with $\mathbb{E}[\beta_1(A)|X] = \bar{\beta}_1$ where $\bar{\beta}_1$ is a constant, then $\mathbb{E}[\partial_{x_1}\phi(X, A)|X = x] = \bar{\beta}_1$. Finally, if the underlying model is an index model, $Y = G(X'\beta, A)$ with $A \perp X$, then $\mathbb{E}[\nabla_x \phi(X, A)|X = x] = \beta$ (here, ∇_x denotes the gradient), up to a scale normalization. Note that from the LASD, average marginal effects $\mathbb{E}[\partial_{x_1}\phi(X, A)]$ are straightforwardly obtained by simply averaging.

short experience *within this subpopulation*, the proportion of type 1 and 2 must be identical. Assumption 2.2 strengthens assumption 2.1 so that all X are independent of unobservables.

Regarding the individual's information set \mathcal{F} , there are two cases that are worthwhile distinguishing, and they are contained in the following assumption.

Assumption 3. *Let all variables be as defined above. The relationship between the individual's σ -algebra \mathcal{F} and the variables at the disposal of the econometrician is given by either of*

1. $\mathcal{F} = \sigma(W, J) \supseteq \sigma(W)$.
2. $\mathcal{F} = \sigma(\varphi(W), J)$ such that $\sigma(\varphi(W), J) \not\supseteq \sigma(W)$ where φ is a Borel function.

Here, $J \in \mathcal{J}$, \mathcal{J} denotes additional information, and \mathcal{J} is a Borel space. The first case is when the individuals actively use more information when backcasting than the econometrician has in the regressors. The second case is when individuals do not use the entire information given to the econometrician, but (potentially) use additional information³. In order to assess all implications of our assumptions, we require some additional conditions, which may be found in the appendix. The one essential restriction is that we confine our analysis to continuously distributed (recalled) dependent variables, the rest are regularity restrictions.

2.1 Individuals Actively use all Information Available to the Econometrician

Consider the scenario defined by assumption 3.1 first. This turns out to be a rather benign scenario, as is seen by the following results on the marginal effects with respect to X_1 and S_1 , proof of which may be found in the appendix. We use the notation $S = (S'_1, S'_{-1})'$, $\theta(\tilde{y}, w) = \left[f_{\tilde{Y}|W}(k_\alpha(w), w) \right]^{-1} \mathbf{1}[\tilde{y} \leq k_\alpha(w)]$ and $Q_{x_1} = \partial_{x_1} \log f_{J|X_1 W_{-1}}(j; w)$. All equalities involving conditional expectations are meant to hold everywhere except possibly on a set of zero measure.

Theorem 1. *(Consequences of Large Information under Squared Error Loss): Let all the variables and functions be as defined above. Let assumptions 1, 2.1, 3.1 and 5 hold. First, suppose that $J = A_1$, i.e., the additional information is conditionally independent of X_1 . Then we obtain for the derivatives of the conditional quantile, for all w :*

1. $\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, \tilde{Y} = k_\alpha(w)]$.
2. *If instead of assumption 2.1, $F_{A|S_1 X S_{-1}} = F_{A|X S_{-1}}$ holds, then $\partial_{s_1} k_\alpha(w) = 0$.*

If assumption 6 holds instead of assumption 5, we obtain for derivatives of the conditional mean, for all w :

³The latter nests the case when the individuals use less information than the econometrician. However, this subcase is not very attractive: As in immediate implication of our assumptions, by measurability $\tilde{Y} = \mathbb{E}[Y | \mathcal{F}_m] = \mathbb{E}[\tilde{Y} | W] = k_\alpha(\tilde{Y} | W)$ holds for all α . But this is in contradiction to the observed residual variation in \tilde{Y} after controlling for regressors in virtually any micro data set.

3. $\partial_{x_1} m(w) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w]$.

4. If instead of assumption 2.1, $F_{A|S_1 X S_{-1}} = F_{A|X S_{-1}}$ holds, $\partial_{s_1} m(w) = 0$.

For general additional information J , $J \not\subseteq A$ and possibly dependent on X_1 conditionally on W_1 , $\partial_{x_1} m(w) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w]$ continues to hold, but

5. $\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, \tilde{Y} = k_\alpha(w)] + \text{Cov}[\theta(\tilde{Y}, W), Q_{x_1} | W = w]$,

under slightly modified regularity conditions.

Discussion of Theorem 1. This theorem analyzes the situation when the backcast is the conditional mean. It states that in this case each individual's empirically obtained marginal effect is the best approximation (in the sense of minimizing distance with respect to L_2 -norm) to the individual's theoretical marginal effect conditioning on either all variables D in the case of the quantile, or just on all regressors W in the case of the mean. We may identify the average structural derivative of all female workers in the iron industry who have been surveyed 20 days after payday and – in the case of the quantile regression – earn high wages, but not the marginal effect of every single individual woman. This quantity is hidden from us by the “double veil” of unobserved heterogeneity both in job related abilities, as well as ability to recall. We may summarize: *If individuals use the squared error loss and more information than the econometrician to backcast Y , then (local) average structural derivatives are identified.* Note that both the derivative of the conditional mean regression and of the quantile regression identify something meaningful, as long as the additional information of the individual is independent of the regressor of interest (Theorem 1.1 and 1.3). In the more plausible scenario where the additional information and the regressor are dependent, essentially due to the law of iterated expectations the mean regression yields an unbiased estimator of a sensible average, while the quantile inherits a bias term. Contrary to conventional wisdom, the quantile is less robust to the presence of this measurement error (Theorem 1.5). As an example, suppose that the underlying population is $Y = X'\beta + A$, with $A \perp W$ and $\mathbb{E}[A] = 0$. Assume in addition that $\mathbb{E}[A | J, W] = \tau(J)$, but J not independent of W . Then, $\partial_{x_1} m(w) = \beta_1$, but $\partial_{x_1} k_\alpha(w) = \beta_1 + \partial_{x_1} k_{\alpha, \tau(J) | W}(w)$. This failure may be attributed to the fact that a law of iterated expectations holds for the conditional mean, but not for the conditional quantile. Finally, if we differentiate with respect to a recall production factor s_1 for which (conditional) independence of unobservables is plausible, then this factor should not have an effect. If we are convinced to possess such exogenous factors then we may use them to test our model specification (Theorem 1.2 and 1.4).

This theorem has a number of interesting extensions. In particular, it is to a surprising degree a knife edge results as the following extensions establish:

Extension 1: Endogenous regressors and instruments. Consider now the case where in addition X_1 is endogenous, i.e., A is not independent of X_1 given W_{-1} , for instance due to simultaneity of Y and X_1 , as arises in supply and demand models. In the mean regression case,

$$\partial_{x_1} m(x_1, w_{-1}) = \mathbb{E}[\partial_{x_1} \phi(X, A) | X_1 = x_1, W_{-1} = w_{-1}] + \text{Cov}[Y, \bar{Q}_{x_1} | X_1 = x_1, W_{-1} = w_{-1}], \quad (2.1)$$

where $\bar{Q}_{x_1} = \partial_{x_1} \log f_{A|X_1 W_{-1}}(a; x_1, w_{-1})$, provided assumption **3.1** still holds, and a similar result holds for the quantile. A standard remedy would be to introduce instruments. At this level of generality, the natural way is a control function approach (see Imbens and Newey (2007), Hoderlein (2005, 2008) or Hoderlein and Mammen (2007)). It amounts to including control function residuals V (i.e., the residuals from a regression of endogenous regressors on instruments), such that $W^* = (W', V')'$. For this modified set of controls W^* , the above assumptions are now required to hold. This means in particular that $\mathcal{F}^* = \sigma(W^*, J) \supseteq \sigma(W^*)$ has to hold, meaning that *the individuals have to use the additional information in the instruments in their recall if the econometrician wants to employ them*. Indeed, by arguments analogous to those following theorem **2**, it is possible to show that otherwise a bias arises.

Extension 2: Median as backcast. When individuals use the absolute error loss to backcast (i.e., their backcast is the conditional median) and strictly more information than the econometrician, it is possible to show that $\partial_{x_1} \mathbb{E} [\tilde{Y} | W = w] \neq \mathbb{E} [\partial_{x_1} \phi(X, A) | W = w]$ in general, see the appendix. Hence, individuals should rather be made to remember the mean, and not the median, provided both differ substantially. The reason for this failure is again that iterated expectations cannot be used. If we assume that $\tilde{Y} = \text{med}_Y(W, A_1) = \psi(W, A_1)$ is strictly monotone in the scalar second component, with assumption **2.1** strengthened to $A_1 | W \sim \mathcal{U}[0, 1]$, then

$$\partial_{x_1} k_\alpha(w) = \mathbb{E} [\partial_{x_1} \phi(X, A) | W = w, A_1 = \alpha, Y = \text{med}_Y(w, \alpha)], \quad (2.2)$$

may be established. As in the literature on triangular models, monotonicity allows for iterative arguments. However, one should emphasize that modelling the information by assuming monotonicity in a scalar A_1 independent of the regressors is not an innocuous set of assumptions.

2.2 Individuals Reduce Information Available to Econometrician

In this scenario, it turns out to be very hard to say something in general. For concreteness, we therefore assume several specifications listed in the following assumption. All of them cover an important scenario. Recall that we are interested in the derivative with respect to x_1 :

Assumption 4. *Let all variables be as defined above. The relationship between the individuals' σ -algebra \mathcal{F} and the variables to the disposal of the econometrician is given by either of*

1. $\mathcal{F}^1 = \sigma(X_{-1}, S, A_1)$
2. $\mathcal{F}^2 = \sigma(X_1, S, A_1)$
3. $\mathcal{F}^3 = \sigma(B, S, A_1)$, where $B = X'\gamma$ and $\gamma_1 > 0$.

These three specifications are designed to capture three different cases. The first is that individuals completely discard the information given in X_1 . The second is that they use the information in X_1 , but discard all other variables of interest that are potentially correlated.

Finally, the last scenario is when all information contained in X is shrunk, but in a proportional fashion and without discarding the variable of interest completely.

The next theorem summarizes the consequences of the variations in information reduction. We will make use of the score $\partial_{x_1} \log f_{X_{-1}|X_1 S}$ (which we denote as Q_{x_1}). Recall that this was defined as the Radon-Nikodym derivative of $F_{X_{-1}|X_1 S}$ with respect to some fixed measure ν . Since this measure does not depend on X_1 and S , this mean that we effectively rule out cases where the support is discrete for some X_1 and continuous for others. Finally, let $\mathcal{S} = \left\{ (p, q) \in \mathbb{R}_+^2 \mid 1/p + 1/q = 1, \mathbb{E} \left[\left| Y - \tilde{Y} \right|^p \right] < \infty, \mathbb{E} [|Q_{x_1}|^q] < \infty \right\}$ and $\mathcal{T} = \left\{ (p, q) \in \mathbb{R}_+^2 \mid 1/p + 1/q = 1, \mathbb{E} [|Y|^p] < \infty, \mathbb{E} [|Q_{x_1}|^q] < \infty \right\}$

Theorem 2. *Let all the variables and functions be as defined above. Let assumptions 1, 2.2, 4.1 and 6 hold. Then follows $\partial_{x_1} m(W) = 0$ (a.s.). If 4.1 is replaced by 4.2, then $\partial_{x_1} m(W) = \mathbb{E}[\partial_{x_1} \phi(X, A) | X_1, S] + \text{Cov} [Y - \tilde{Y}, Q_{x_1} | X_1, S]$. Moreover, for the local average structural derivative,*

$$|\partial_{x_1} m(W) - \mathbb{E}[\partial_{x_1} \phi(X, A) | X_1, S]| \leq \min_{(p, q \in \mathcal{S})} \mathbb{E} \left[\mathbb{E} \left[\left| Y - \tilde{Y} \right|^p \mid \mathcal{F}^2 \right] \mid X_1, S \right]^{1/p} \mathbb{E} [|Q_{x_1}|^q \mid X_1, S]^{1/q},$$

and for the average structural derivative,

$$|\mathbb{E} [\partial_{x_1} m(W)] - \mathbb{E}[\partial_{x_1} \phi(X, A)]| \leq \min_{(p, q \in \mathcal{T})} \mathbb{E} [|Y|^p]^{1/p} \mathbb{E} [|Q_{x_1}|^q]^{1/q}.$$

Finally, if 4.2 is replaced by 4.3, then $\partial_{x_1} m(W) = \mathbb{E}[\partial_{x_1} \phi(X, A) | B, S] + \text{Cov} [U, Q_{x_1} | X_1, S]$, and the bounds remain unchanged. Similar results hold for the conditional quantile, and may be found in the appendix.

Discussion of Theorem 2. The first simple – yet important – result is the following: If individuals do not use any information contained in X_1 at all in **their** mental backcast, then an econometrician may never be able to recover even average marginal effects. Conversely, if we obtain a zero marginal effect it does not mean that the true effect has to be zero. Hence it is imperative that individuals use in particular all variables we are interested in when forming a backcast. However, even that is not sufficient to obtain unbiasedness, as we learn from the second part of theorem 2. Neglect the bias term for the moment. Then we conclude that only $\mathbb{E}[\partial_{x_1} \phi(X, A) | \mathcal{F}_\alpha]$, $\mathcal{F}_\alpha \subseteq \mathcal{F}$ is identified, which in turn allows to identify at least the average marginal effect across the entire population, i.e., $\mathbb{E}[\partial_{x_1} \phi(X, A)]$. But this presupposes that the bias term $\mathbb{E}[Y Q_{x_1} | X_1, S] = \text{Cov} [Y - \tilde{Y}, Q_{x_1} | X_1, S]$ can be handled. The bias depends either on the true variable Y or on the recall error $U = Y - \tilde{Y}$, both of which are unobserved.

However, we may provide bounds for this effect, if we have either a measure of the p -th absolute moment of Y or of U . For instance, if we have a measure for the precision of Y , i.e. $\text{Var} [Y | \mathcal{F}] = \text{Var} [U | \mathcal{F}]$, we may by Cauchy-Schwarz give at least an upper bound for the absolute value of the bias, because the term $\mathbb{E} [Q_{x_1}^2 | X_1, S]$ depends entirely on observables. Observe that the bias term does only depend on aggregate moments if average effects are to be elicited. Note further that the bias and the bounds remain unchanged if the individual uses the index-type information reduction, even if the true value of γ is unknown to the econometrician.

3 Tests in the Presence of Imperfectly Recalled Variables

There are several important novel features our model suggests: 1. Is there evidence for reduction of the information provided to the econometrician? 2. Can we find the quantities necessary to construct the bounds? 3. How do we incorporate the bounds when evaluating economic hypotheses? Evaluating these questions will be the topic of this section. We focus on homogeneity of degree zero in consumer demand, i.e., the property that $g(p/\xi, \xi) = g(p/\xi)$ for p prices and ξ income, or, equivalently, that a partial marginal effect is zero. We discuss first how to test this property in the absence of recall error, then we show how to test for information reduction, and finally we establish how to test the hypothesis of a zero marginal effect in the presence of information reduction and bounds.

1. Testing for homogeneity of degree zero. Our null hypothesis is defined as $H_0: \mathbb{P}\{\mu(X_1, W_{-1}) = \eta(W_{-1})\} = 1$ where μ and η are mean regressions. The alternative is that both regressions differ on a subset of their support of positive measure. Using a nonzero and bounded weighting function a , the null is equivalent to

$$\Gamma_1 = \mathbb{E}\left(\sum_{j=1}^{d_Y} (\mu^j(X_1, W_{-1}) - \mathbb{E}[\eta^j(W_{-1})|W])^2 a(W)\right) = 0. \quad (3.1)$$

We may use sample counterparts to check whether Γ_1 is significantly different from 0. To obtain sample counterparts, one simply has to replace all functions by nonparametric estimates.

2. Testing for information reduction. As we have seen in the previous sections, the question of whether individuals reduce information to make their forecast is of crucial importance for the ability of econometricians to identify something meaningful. Recall that under the assumption of dimension reduction, **3.2**, the optimal predictor constitutes a nonseparable model of the form $\tilde{Y} = \psi(\varphi(W), A_1)$, regardless of the specific loss function assumed. If we add the assumptions that A_1 is independent of X given S and that X enters through an index, this model may be analyzed using Hoderlein and Mammen's (2007) theorem⁴. We may estimate the parameter β (up to scale) in the model $\tilde{Y}_k = \psi(X'\beta, S, A_1)$, where \tilde{Y}_k denote the k -th component, by use of the equations

$$\partial_x k_\alpha(w) = \beta \mathbb{E}[\partial_l \psi_k(X'\beta, S, A_1)|W = w, \tilde{Y}_k = k_\alpha(w)], \quad k = 1, \dots, d, \quad (3.2)$$

where ∂_l denotes the derivative with respect to the index. Imposing a normalization condition, we obtain $\mathbb{E}[\int \partial_x k_\alpha(W) w(\alpha, W) d\alpha] = \beta$ for some weighting function w . This identifies β . For testing the hypothesis $\varphi_k(X, A_1) = \psi_k(X'\beta, A_1)$, $k = 1, \dots, d$, (a.s), we use that equation (3.2)

⁴At this point we want to emphasize that we could devise a test for dimension reduction more generally, but for the purpose of this paper we concentrate on the assumption that dimension reduction takes the form of the index, ie $\mathcal{F} = \sigma(B, S, A_1)$, where $B = X'\beta$.

implies that $\partial_{x_l} k_\alpha(w)/\partial_{x_j} k_\alpha(w) = \beta^l/\beta^j$, $k = 1, \dots, d$, $l, j = 1, \dots, p$. This suggests to form

$$\Gamma_2 = \mathbb{E} \left[\sum_k \sum_{j>l} \int [\partial_{x_l} k_\alpha(W)\beta^j - \partial_{x_j} k_\alpha(W)\beta^l]^2 \gamma(W, \alpha) d\alpha \right], \quad (3.3)$$

for some bounded nonzero weighting function γ .

Intuitively, a test for information reduction is translated into a test for the validity of the null that the model has an index. Of course, failure to reject an index restriction may still mean that the true model has an index. But if we have two measurements of the same object, and we find the index being rejected at one specification while not at the other (e.g., we have a validation data set, or several different imperfect measurements of the same variable), we conclude that the one for which we fail to reject exhibits information reduction. The asymptotic distribution theory for this test is too involved to be useful in practise, hence we use the following bootstrap procedure:

1. Calculate (multivariate) single indices $X'_i \hat{\beta}$, using the estimated coefficients $\hat{\beta}$ from equation (3.2).
2. Randomly draw α_i^* from $U[0, 1]$.
3. Generate the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \dots, n$ by $\tilde{Y}_i^* = \hat{k}_{\alpha_i^*}(X'_i \hat{\beta}, S_i)$, $X_i^* = X_i$ and $S_i^* = S_i$.
4. Calculate $\hat{\Gamma}_2^*$ from the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \dots, n$.
5. Repeat steps 2 to 4 often enough to obtain critical values for $\hat{\Gamma}_2$.

3. Testing for zero marginal effect in the presence of information reduction. If we detect information reduction we might be able to provide bounds on the marginal effects, as outlined in theorem 2. In this section we will consider as examples the bounds introduced in the introduction, i.e., $(\mathbb{E}[\partial_{x_1} m(X)] - \mathbb{E}[\partial_{x_1} \phi(X, A)])^2 \leq \mathbb{E}[Y^2] \mathbb{E}[Q_{x_1}^2]$, where m denotes the nonparametric regression of \tilde{Y} on $X = (X'_1, X'_{-1})'$, $f_{X_{-1}|X_1}(x_{-1}, x_1)$ denotes the conditional density, and $Q_{x_1} = \partial_{x_1} \log f_{X_{-1}|X_1}(X_{-1}, X_1)$. Moreover, let $\beta_1 = \mathbb{E}[\partial_{x_1} \phi(X, A)]$, and assume for the time that $\mathbb{E}[Y^2]$ is consistently estimable. Suppose now that we want to test the hypothesis that $\beta_1 = 0$. We use the fact that this is equivalent to testing the null hypothesis that $\mathbb{E}[\partial_{x_1} m(X)]^2 - \mathbb{E}[Y^2] \mathbb{E}[Q_{x_1}^2] \leq 0$. This suggests to use

$$\Gamma_3 = \mathbb{E}[\partial_{x_1} m(X)]^2 - \mathbb{E}[Y^2] \mathbb{E}[Q_{x_1}^2],$$

and test $H_0 : \Gamma_3 \leq 0$ versus the (one sided) alternative $H_1 : \Gamma_3 > 0$. To see that the bootstrap is consistent, rewrite $\Gamma_3 = g(\mathbb{E}[\partial_{x_1} m(X)], \mathbb{E}[Y^2], \mathbb{E}[Q_{x_1}^2])$, where $g : \mathbb{R}^3 \rightarrow \mathbb{R}$, $g(x, y, z) = x^2 - yz$ is a smooth function of the sample averages. Assume $\mathbb{E}[\partial_{x_1} m(X)] \neq 0$, and $\mathbb{E}[Y^2] > 0, \mathbb{E}[Q_{x_1}^2] > 0$ which rules out the degenerate case when g is not being differentiable at $(\mathbb{E}[\partial_{x_1} m(X)], \mathbb{E}[Y^2], \mathbb{E}[Q_{x_1}^2])$, see Datta (1995) for a discussion. While assuming $\mathbb{E}[Y^2] > 0$

and $\mathbb{E}[Q_{x_1}^2] > 0$ is fairly innocuous, $\mathbb{E}[\partial_{x_1} m(X)] \neq 0$ is less obvious. But note that homogeneity of degree zero implies (only) that $\partial_{x_1} m(X) + \mathbb{E}[YQ_{x_1}|X] = 0$. Since $\mathbb{E}[YQ_{x_1}|X] \neq 0$ on a set of probability one, under the conditions of theorem 2 this implies that $\partial_{x_1} m(X) \neq 0$ and hence $\mathbb{E}[\partial_{x_1} m(X)] \neq 0$ on a set of probability one. This implies that the sample counterpart statistic

$$\hat{\Gamma}_3 = \left[n^{-1} \sum_i \partial_{x_1} \hat{m}(X_i) \right]^2 - \left[n^{-1} \sum_i Y_i^2 \right] \left[n^{-1} \sum_i Q_{x_1,i}^2 \right],$$

is \sqrt{n} asymptotically normally distributed. To see this, apply a Taylor expansion to g :

$$\begin{aligned} \hat{\Gamma}_3 - \Gamma_3 &= 2\mathbb{E}[\partial_{x_1} m(X)] \left\{ n^{-1} \sum_i \partial_{x_1} \hat{m}(X_i) - \mathbb{E}[\partial_{x_1} m(X)] \right\} \\ &\quad + \mathbb{E}[Q_{x_1}^2] \left\{ n^{-1} \sum_i Y_i^2 - \mathbb{E}[Y^2] \right\} + \mathbb{E}[Y^2] \left\{ n^{-1} \sum_i Q_{x_1,i}^2 - \mathbb{E}[Q_{x_1,i}^2] \right\} + o_p(n^{-1/2}). \end{aligned}$$

Note that this expansion only makes sense under the assumption that rule out the degenerate case. Moreover, note that in order to obtain \sqrt{n} -convergence of $n^{-1} \sum_i \partial_{x_1} \hat{m}(X_i)$ by an U -statistic CLT, we require higher order smoothness assumptions on m and have to employ a higher order Kernel. Under these assumptions, an iid bootstrap based estimator of the distribution function of $\sqrt{n}(\hat{\Gamma}_3^* - \hat{\Gamma}_3)$ yields a consistent estimate of the distribution of $\sqrt{n}(\hat{\Gamma}_3 - \Gamma_3)$. To test whether $\Gamma_3 < 0$, we propose to use a hybrid bootstrap based on the one sided confidence interval using $\underline{\Gamma}_3$, where $\underline{\Gamma}_3$ is the lower 0.05 confidence bound. If this value is above zero, then we reject H_0 . To obtain this bound, in the hybrid bootstrap we invert the confidence interval, i.e. if γ_τ^* denotes the 0.95 percentile of $\sqrt{n}(\hat{\Gamma}_3^* - \hat{\Gamma}_3)$ we calculate $\underline{\Gamma}_3 = \hat{\Gamma}_3 - 1/\sqrt{n}\gamma_\tau^* = 2\hat{\Gamma}_3 - k_{.95}(\hat{\Gamma}_3^*)$, where $k_{.95}(\hat{\Gamma}_3^*)$ denotes the .95 percentile of $\hat{\Gamma}_3^*$. To obtain the distribution of $\sqrt{n}(\hat{\Gamma}_3^* - \hat{\Gamma}_3)$, we run the following bootstrap procedure:

1. Generate the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \dots, n$ by i.i.d. resampling from the data.
2. Calculate $\sqrt{n}(\hat{\Gamma}_3^* - \hat{\Gamma}_3)$ from the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \dots, n$.
3. Repeat steps 1-2 often enough to obtain percentiles of $F_{\sqrt{n}(\hat{\Gamma}_3^* - \hat{\Gamma}_3)}$.

Can we do something in the case where we only have $\sum_i \tilde{Y}_i^2$? Note that if we assume individuals to backcast with squared error loss, we obtain that $\mathbb{E}[\tilde{Y}^2] \leq \mathbb{E}[Y^2]$. Hence, with probability approach one, $n^{-1} \sum_i \tilde{Y}_i^2 \leq n^{-1} \sum_i Y_i^2$. Let $\hat{\Gamma}_4 = n^{-2} \left[\left\{ \sum_i \partial_{x_1} m(X_i) \right\}^2 - \sum_i \tilde{Y}_i^2 \sum_i Q_{x_1,i}^2 \right]$. Then, $\mathbb{P}[\hat{\Gamma}_3 \leq c] \geq \mathbb{P}[\hat{\Gamma}_4 \leq c]$ for all c , as n gets large. Consequently, if we compute in the same fashion as above the is the lower 0.05 confidence bound of $\hat{\Gamma}_4^* = n^{-2} \left[\left\{ \sum_i \partial_{x_1} \hat{m}(X_i^*) \right\}^2 - \sum_i \tilde{Y}_i^{*2} \sum_i Q_{x_1,i}^{*2} \right]$, and obtain that it is below zero then we conclude that $\Gamma_3 \leq 0$ is not rejected, because a test based on the bootstrap distribution of $\hat{\Gamma}_3^*$ would squarely not reject.

4 Recall Effects in Real World Data: Empirical Evidence from the HRS

In this section, we are concerned with answering the following question: “What can be learned about an economic property from a recall error ridden data set?” More specifically, we are concerned with the impact of recall effects on the evaluation of zero homogeneity of food consumption, a key economic property commonly known as “absence of money illusion”. It is a nonparametric property which arises when maximizing any regular utility function subject to a linear budget constraint, and translates into a zero derivative restriction as mentioned above. We emphasize the nonparametric character at this point, as it shows the strength of our approach which allows to analyze zero homogeneity, taking the effect of recall errors into account without assuming a functional relationship between all variables or specifying the distribution of the errors.

Distortions in response behavior in questions on consumption items are of great interest in applied economic research (see Battistin, Miniaci, and Weber, 2003, and Browning, Crossley, and Weber, 2003, for reviews). We focus on food consumption data, because they are usually considered to be relatively accurately measured. However, we find huge recall errors alongside with severe implications for economic analysis, and we therefore feel that it is a particular fruitful area for the application of our theoretical analysis.

Data description. The data we use comes from the Health and Retirement Study (HRS) and a supplemental survey, the Consumption and Activities Mail Survey (CAMS). We use data from the 2000, 2002, and 2004 waves of the HRS main survey and from the 2001 and 2003 waves of CAMS. The HRS is a biennial panel of older Americans; the target population of the HRS was the cohorts born in 1931–1941 (see Juster and Suzman, 1995). In 2000 the HRS interviewed about 20,000 subjects in 13,100 households. For the first wave of CAMS, a random sample of 5,000 HRS households was interviewed, see Hurd and Rohwedder (2005).

The dependent variable in our analysis is food expenditure. Specifically, we use the item “Food and beverages: food and drinks, including alcoholic, that you buy in grocery or other stores”. The covariates are mostly taken from the RAND version of the HRS 2000, 2002, and 2004 data, and matched to CAMS where necessary. Self-rated memory and change of self-rated memory are not contained in RAND’s HRS distribution; these variables are taken directly from the HRS raw data. A more detailed documentation of the covariate data is available on request (some descriptive statistics can be found in tables A1-A3 in the Appendix). Price data are taken from seasonally adjusted U.S. CPI time series provided by the Bureau of Labor Statistics (BLS), separately for “all items less food” and “food and beverages”. Following the demand literature we assume additive separability of demand from labor supply and the intertemporal allocation decision. Thus, we use primarily total expenditure defined as sum over all expenditure categories as income concept, but we also use total household income.

Evidence on recall measurement error. To answer this question, we exploit the fact

that in the CAMS respondents can decide about the reporting horizon for consumption items themselves. The respondents were provided with the following options: “We have included three time periods so that you can estimate your spending in the way that is easiest for you for each category. For example, if it is easiest for you to think about what you spent on food and beverages last week, then please enter the amount in the first column.” The four options given for each item are: “Amount spent last week / Amount spent last month / Amount spent in last 12 months / No money spent on this in last 12 months.”

Since food is not subject to infrequent purchases, economic theory predicts that there should not be a difference in response behavior between the three measures. However, psychological research tells us that memory deteriorates over time, and that frequently occurring events are “estimated” and not memorized in detail (see Tourangeau, Rips, and Rasinski (2000)). Hence we would expect to find some differences between any two mean regressions that use alternative time horizons. Moreover, our theory predicts that the association between income and food consumption should be the weaker the more time elapsed since the decision is made. For instance, assume that there is heterogeneity in σ -algebras in the sense that a fraction of the population uses more information than the econometrician when backcasting while another fraction discards the information in the regressor of interest completely (as in the first part of theorem 2). Then we would expect the marginal effects of the mean regression (a local average of the two subpopulations!) to attenuate.

Figures 1 and 2 show the levels and derivatives of the nonparametric regression of the food budget share on total expenditure, using different time horizons as well as different definitions of the dependent variable.

These pictures illustrate that left-hand side recall error matters dramatically for the economic issue we want to consider. Moreover, we obtain the predicted attenuation effect. The average derivatives are: weekly -0.065026 , monthly -0.044245 , yearly -0.033798 . The picture changes slightly if we control for endogeneity of total expenditure. In this case, we obtain the ADEs: weekly -0.070767 , monthly -0.029048 , yearly -0.031103 . This difference is statistically significant, as is corroborated by a test available from the author’s upon request. We conclude that there is a persistent difference between the remembered weekly expenditures, and the “estimated” monthly and yearly ones. Note that the effect of total expenditure is more than twice as strong with the weekly than with the monthly or yearly response category, demonstrating that left hand side recall effects can be enormous.

Before we jump hastily to the conclusion that everything behaves as we predicted, we should consider first alternative explanations, in particular possible self selection into the different response categories. A possible explanation could for instance be that individuals in institutions (e.g., retirement homes – recall that our sample consists largely of individuals around retirement age) are used to think in annual rates, while individuals outside these institutions are not. If we compare the various groups in our data using the descriptive statistics in table A2 in the appendix, we find little difference between the groups. Nevertheless, we correct for selection by including the propensity score of choosing a certain type of answer as control function. As

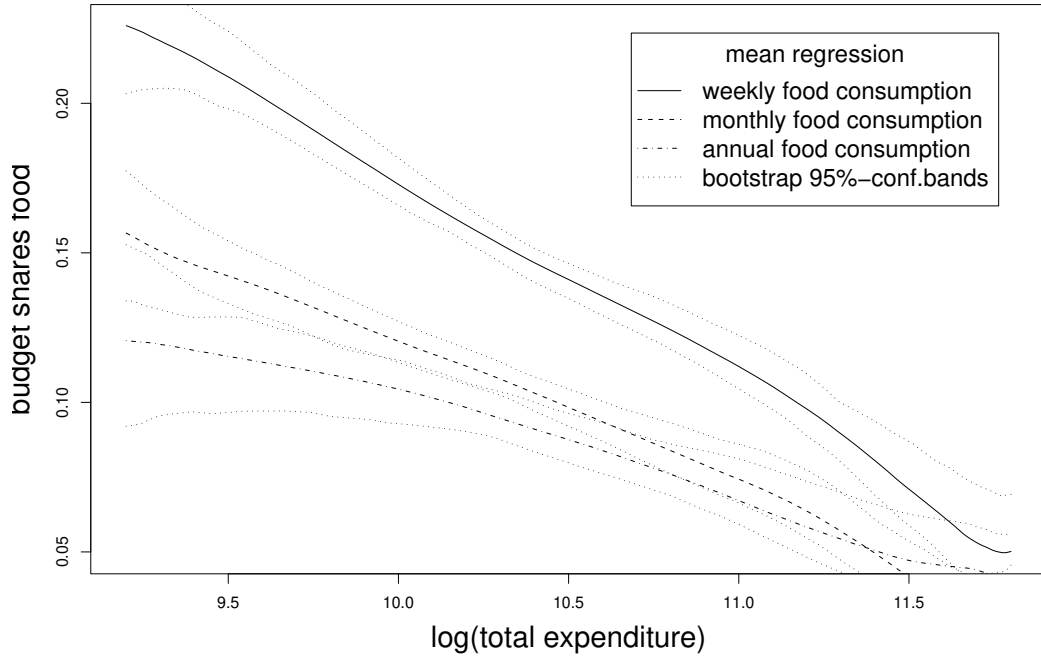


Figure 1: Regressions of Food Consumption vs Total Expenditure based on Weekly, Monthly and Yearly Budget Share Data.

Table 1: Comparison Weekly, Monthly, Yearly

	Index Test for Information Reduction		
	weekly	monthly	annual
$mc = low$	0	0.21	0.10
$mc = medium$	0	0.09	0.24
$mc = high$	0	0	0.18

P-values of the test for information reduction for the three groups of reported food consumption, and for three levels of mental capacity (mc).

exclusion restriction we use the fact that mental capacity and time used for answering may influence whether one is in an institution, but not the preferences for food. Including these propensity scores, the results remain materially unchanged. Together with the fact that the average derivatives do not change much when selection is accounted for (weekly -0.065808 , monthly -0.043622 , yearly -0.033902 , with correction for endogeneity, weekly -0.078327 , monthly -0.029762 , yearly -0.026364) we can rule out selection effects as driving force.

An alternative explanation for our results is that our measure of food consumption (essentially “food at home”) generates selection. For instance, it is imaginable that those who answer the weekly category are the ones who most often eat at home, while those who eat out frequently also buy food less frequent in stores and hence select themselves into the yearly

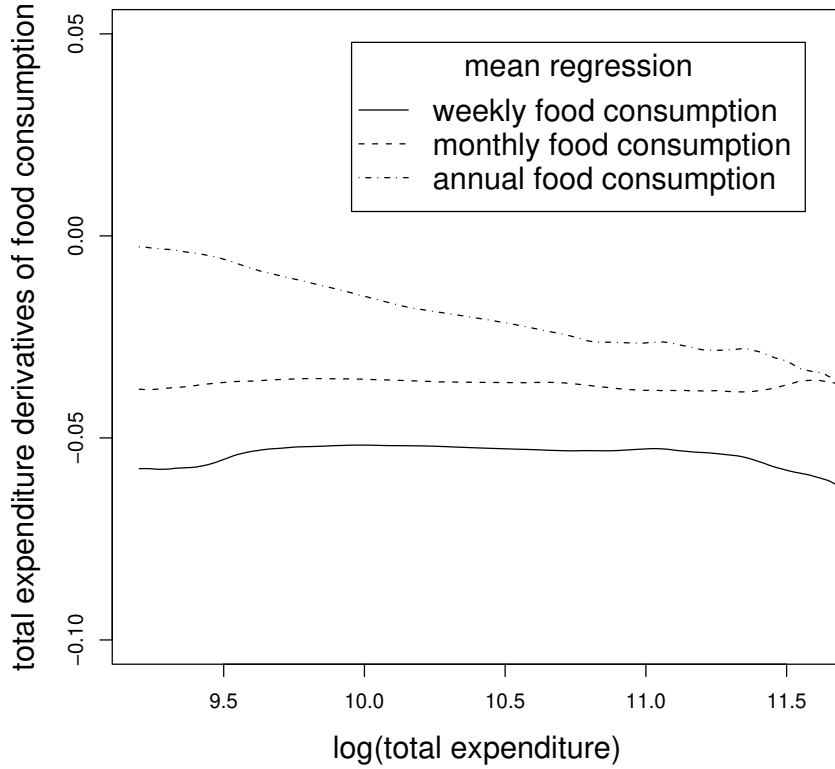


Figure 2: Derivatives of Total Food Consumption (at Home and outside Home) vs Total Expenditure based on Weekly, Monthly and Yearly Budget Share Data.

category. While plausible in principle, however, this is not the case. If we use a different food measure that includes dining out, the results remain unchanged, see fig.2.

Having established that there is a significant effect of the time horizon people choose, and in particular (as our theory predicts) an attenuation-like effect, we proceed by looking at the results of a formal test for information reduction shown in table 1. Our theory predicts that information reduction may create structures where there are previously none. To assess this, we use the test described above in section 3. The findings confirm the intuition that information reduction is more probable the further away in time things are, and the less mental capacity respondents have. Consequently, we may expect at least parts of the population to exhibit response behavior that cause left hand side recall errors of the problematic type.

Effect of recall errors on testing zero derivatives. In the light of the evidence on information reduction, it is appropriate to use the more conservative bounds put forward in theorem 2. To see the effect, we first test for homogeneity of degree zero without bounds. We apply a conventional nonparametric test described in section 3. We strongly reject homogeneity with a p-value of virtually zero. As reported frequently in the literature, the nonparametric tests are quite powerful and detect even small deviations from the null, see Lewbel (1999). However,

following our discussion, there is strong evidence for left hand side “information reduction” recall error which bias our results. To account for the possible effect of the latter on our results, we implement the test statistic $\hat{\Gamma}_4$ discussed in section 3, which accommodates for recall errors through bounds. The point estimate is $-2.11 * 10^{-6}$, and fig. 3 shows the bootstrap distribution of the test statistic. Not surprisingly, $\hat{\Gamma}_4$ is soundly negative, and we conclude that we cannot reject homogeneity any longer. Indeed, the bootstrap distribution in fig. 3 suggests that we might even reject $H_0 : \Gamma_4 = 0$ in favor of the one sided alternative that it is smaller, but in light of our theory this would have no added implication.

One may criticize the bounds on grounds that they are too wide, and the test hence lacks power. However, they are reasonably tight. The first squared term in $\hat{\Gamma}_4$ is $0.74 * 10^{-6}$ and the second is $-2.85 * 10^{-6}$, i.e., the bounds are four times bigger than the effect. While this seems wide, it is instructive to compare this with the squared average income effect we obtained before when we were not imposing homogeneity. This is in the order of magnitude of $c * 10^{-4}$, or roughly 100 times bigger than the effect when we impose homogeneity. We conclude that imposing homogeneity makes the coefficient already very weak, but in the absence of bounds the effect is significant largely because the employed L_2 -distance test is quite powerful. Summarizing, this is evidence that an explanation for the frequent rejection of homogeneity in consumer demand is indeed left hand side recall error. Accounting for this effect, we do not find homogeneity rejected any longer.

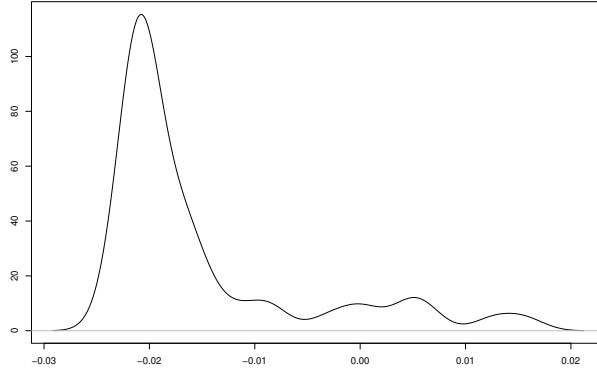


Figure 3: Bootstrap Distribution of $\hat{\Gamma}_4$ (scaled by 10^4)

5 Summary and Outlook

In this paper, we characterize the properties and consequences of measurement error in household surveys from a new perspective: We argue that econometricians should model the response itself since it has behavioral content rather than focusing on the properties of the response error. We focus on the role of respondents’ limited ability to recall information, but our analysis can potentially be extended to other aspects of the survey response processes, such as respondent-interviewer interaction. We leave this to future research.

One important practical implication of our findings is that in the social sciences, left-hand side measurement errors should be regarded as potentially as problematic as right-hand side errors (see also the discussion in Hausman, 2001). Also, quantiles may be less robust than means. Another interesting insight is that it is crucial whether respondents really use all the information the econometrician is interested in when they recall a quantity – including instruments, for example. Our results suggest that instruments may become useless if respondents do not use them actively in the survey response process.

For the various problems that appear when survey responses are subject to limited recall, we devise a number of remedies. The remedies are of two sorts. *Ex post*: We already pointed out immediate cures that an applied researcher could implement to accommodate recall problems in the analysis data from existing household surveys. *Ex ante*: Our approach also suggests survey techniques that help researchers and field institutions to reduce response errors at the design stage of a survey questionnaire. Specifically, the following strategies should be explored in future research. First, when asked for one specific variable, respondents should be encouraged to actively memorize *all* variables of importance for the economic decision that involves this variable. Second, when trying to recall their decisions, respondents should be encouraged to report the average decision rather than the median (“typical”) decision, provided these differ substantially. Third, if instruments are to be employed in the econometric analysis, these variables should be highlighted to the respondents so that they use this information when forming their response. Forth, a measure of precision of the answer should be elicited.

Of course, it remains to be determined how useful these remedies are in practice, but we believe that this is a fruitful area for future research. In an empirical application we have illustrated what type of problems arise due to limited recall, and in how far they may be tackled with the tools and remedies we propose in this paper.

6 Appendix 1: Proofs of the Identification Results

Regularity Conditions

Assumption 5. *Regularity Assumptions Quantiles:* The conditional distribution of \tilde{Y} given W is absolutely continuous w.r.t. the Lebesgue measure (for w). The conditional density $f_{\tilde{Y}|W}(y|w)$ of \tilde{Y} given $W = w$ is bounded in $y \in \mathbb{R}$. $k_\alpha \left[\tilde{Y} | W = w \right]$ is continuously partially differentiable with respect to the first component. The conditional distribution of $(\tilde{Y}, \partial_{x_1} \psi)$, given W , is absolutely continuous w.r.t. the Lebesgue measure. For the conditional density $f_{Y, \partial_{x_1} \psi | W}$ the following inequality holds with a constant C and a positive density g on \mathbb{R} with finite mean (i.e. $\int |y'|g(y') dy' < \infty$): $f_{Y, \partial_{x_1} \psi | W}(y, y'|w) \leq Cg(y')$. Finally, $\mathbb{E}[Y] = c < \infty$, where c is a generic constant.

Assumption 6. *Regularity Assumptions Mean:* The conditional distribution of \tilde{Y} given W is absolutely continuous w.r.t. the Lebesgue measure (for w). The conditional density $f_{\tilde{Y}|W}(y|w)$

of \tilde{Y} given $W = w$ is bounded in $y \in \mathbb{R}$. $\mathbb{E}[\tilde{Y}|W = w]$ is continuously partially differentiable with respect to the first component with uniformly bounded derivatives. $\mathbb{E}[Y] = c < \infty$.

Proof of Theorem 1

Start with the conditional quantile $k_\alpha(w)$, i.e. Theorem 1.1. Let $W = (X', S')'$, $A = (A'_1, A'_{-1})'$ and write $\mathbb{E}[Y|\mathcal{F}] = \varphi(W, A_1)$. Then, by the conditional independence assumption 2.1, and under the regularity assumptions 5, using a theorem of Hoderlein and Mammen (2007) we obtain that

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \varphi(A_1, W) | W = w, \tilde{Y} = k_\alpha(w)],$$

for all $w \in \mathcal{W}$. But $\partial_{x_1} \varphi(W, A_1) = \mathbb{E}[\partial_{x_1} \phi | \mathcal{F}]$, using uniform boundedness of $\partial_{x_1} \phi$, together with the facts that $\sigma(A_1, W, \tilde{Y}) = \sigma(A_1, W)$ and that assumption 2.1 implies that $A_{-1} \perp X | A_1, W$. Then, under assumption 3.1 specialized to $J = A_1$, we may apply the law of iterated expectations (LIE) to obtain $\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \phi | W = w, \tilde{Y} = k_\alpha(w)]$. To see Theorem 1.2, note that ϕ is not a function of s_1 , which combined with the same arguments produces the result.

For the conditional mean, note that by the regularity assumption 6, $m(w) = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}] | W = w] = \mathbb{E}[Y | W = w] = \mathbb{E}[\phi(X, A) | W = w]$, by assumption 3.2 and application of LIE. Then by bounded convergence in connection with the regularity assumption 6, we may interchange differentiation and integration. Finally, $\partial_{x_1} m(w) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w]$ by assumption 2.1. Note that this argument does not require J to be conditionally independent of X_1 . This contrasts with the derivative of the conditional quantile, cf. Theorem 1.5. To show the latter, we apply another theorem of Hoderlein and Mammen (2007). Under regularity conditions stated there,

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \varphi(J, W) | W = w, \tilde{Y} = k_\alpha(w)] + l_\alpha(w),$$

where $l_\alpha(w) = f_{\tilde{Y}|W}(k_\alpha(w); w)^{-1} \mathbb{E}[\mathbf{1}[\tilde{Y} \leq k_\alpha(w)] \partial_{x_1} \log f_{J|W}(J; w) | W = w]$. Rearranging terms produces the result. Q.E.D.

Proof of Statements in Extension 2 to Theorem 1

To see that $\partial_{x_1} \mathbb{E}[\tilde{Y} | W = w] \neq \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w]$ in general, observe that

$$\begin{aligned} \partial_{x_1} \mathbb{E}[\tilde{Y} | W = w] &= \mathbb{E}[\partial_{x_1} \text{med}_Y(W, A_1) | W = w], \\ &= \int_{\mathcal{A}_1} \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, A_1 = a_1, Y = \text{med}_Y(w, a_1)] F_{A_1|W}(da_1; w) \\ &\neq \int_{\mathcal{A}_1} \int_{\mathcal{Y}} \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, A_1 = a_1, Y = y] F_{Y|A_1W}(dy; a_1, w) F_{A_1|W}(da_1; w) \\ &= \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w], \end{aligned}$$

where we used first conditional independence assumption 2.1, and then again Hoderlein and Mammen's theorem. Hence, no law of iterated expectations applies. Example: let $A =$

(A_1, A_2) , $W = X_1$, $Y = \phi(X_1, A) = A_1^2 A_2 X_1$, and assume that $\text{med}_{A_2}(x_1, a_1) = 1$ from which $\text{med}_Y(x_1, a_1) = a_1^2 x_1$. Note that assumption 2.1 reduces to $A \perp X_1$. Then, observe that $\mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, A_1 = a_1, Y = \text{med}_Y(w, a_1)] = a_1^2$, implying that $\partial_{x_1} \mathbb{E}[\tilde{Y} | W = w] = \mathbb{E}[A_1^2]$. Next, $\mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, A_1 = a_1, Y = y] = y x_1^{-1}$ and hence $\mathbb{E}[\partial_{x_1} \phi(X, A) | W = w] = \mathbb{E}[Y | X_1 = x_1] x_1^{-1} = \mathbb{E}[A_1^2 A_2] \neq \mathbb{E}[A_1^2]$ unless e.g., $\text{Cov}(A_1, A_2) = 0$ and $\mathbb{E}[A_2] = 1$.

In fact, it is difficult to construct nontrivial examples where the equality holds. Similar arguments hold for the quantile in general. However, if we assume that $\tilde{Y} = \text{med}_Y(W, A_1) = \psi(W, A_1)$ is strictly monotone in the scalar first component, with 2.1 strengthened to $A_1 | W \sim U[0, 1]$, i.e. the conditions of equation (2.2) hold, then

$$\partial_{x_1} k_\alpha(w) = \partial_{x_1} \psi(w, \alpha) = \partial_{x_1} \text{med}_Y(w, \alpha) = \mathbb{E}[\partial_{x_1} \phi(X, A) | W = w, A_1 = \alpha, Y = \text{med}_Y(w, \alpha)].$$

This shows equation (2.2).

Q.E.D.

Theorem 3

The next theorem establishes what may be learned from quantiles under information reduction:

Theorem 3. *Let all the variables and functions be as defined above. Let assumptions 1, 2, 4.1 and 5 hold. Then follows $\partial_{x_1} k_\alpha(w) = 0$ (a.s.). If 4.1 is replaced by 4.2, then*

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\psi(X_1, S, A_1) | W = w, \tilde{Y} = k_\alpha(w)], \text{ for all } w \in \mathcal{W},$$

where $\psi(X_1, S, A_1) = \mathbb{E}[\partial_{x_1} \phi | X_1, S, A_1] + \text{Cov}[U, Q_{x_1} | X_1, S, A_1]$. Finally, if 4.2 is replaced by 4.3, then

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\eta(B, S, A_1) | W = w, \tilde{Y} = k_\alpha(w)], \text{ for all } w \in \mathcal{W},$$

where $\eta(B, S, A_1) = \mathbb{E}[\partial_{x_1} \phi | B, S, A_1] + \text{Cov}[U, Q_{x_1} | X_1, S, A_1]$.

Remark A.1: Discussion of Theorem 3. Compared to the mean, quantiles have the drawback that even if the respective bias term vanishes, the leading term does not produce a sensible average derivative. However, like in the case of the large information scenario, if we assume monotonicity of \tilde{Y} in scalar A_1 , then one can show that the leading term identifies again something meaningful, namely $\mathbb{E}[\partial_{x_1} \phi(X, A) | \mathcal{F}]$. Bounds may be derived as in theorem 2.

Proof of Theorems 2 and 3

To see the first statement, start by noting that by assumption 6, $\mathbb{E}[Y | \mathcal{F}]$ exists and is differentiable. Then, by assumption 4.1 and the LIE,

$$\mathbb{E}[\mathbb{E}[Y | \mathcal{F}] | W = w] = \mathbb{E}[\mathbb{E}[\phi(X, A) | W_{-1}, A_1] | W = w].$$

Hence,

$$\partial_{x_1} m(w) = \int_{\mathcal{A}_1} \mathbb{E}[\phi(X, A) | W_{-1} = w_{-1}, A_1 = a] \partial_{x_1} f_{A_1 | W}(a; w) v(da).$$

But by 2, $F_{A_1|X_{-1}W_{-1}}(a; w) = F_{A_1|W_{-1}}(a; w)$, and hence $\partial_{x_1} m(w) = 0$ follows. For the case of the quantile, note that by applying Hoderlein and Mammen (2007) to $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}] = \psi(W_{-1}, A_1)$ in connection with assumption 2, we obtain that

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \psi(W_{-1}, A_1) | W = w, \tilde{Y} = k_\alpha(w)] = 0,$$

where the last equality follows by $\partial_{x_1} \psi = 0$.

To the second statement, note again that by assumption 4.2 and the LIE that

$$\begin{aligned} \partial_{x_1} \mathbb{E} [\mathbb{E}[Y|\mathcal{F}] | X = x, S = s] &= \mathbb{E} [\partial_{x_1} \mathbb{E}[\phi(X, A) | X_1, S, A_1] | X = x, S = s] \\ &= \mathbb{E} [\partial_{x_1} \phi(X, A) | X_1 = x_1, S = s] \\ &\quad + \int_{\mathcal{A} \times \mathcal{X}_1} \phi(x_1, \xi_{-1}, a) \partial_{x_1} f_{AX_{-1}|X_1 S}(a, \xi_{-1}; x_1, s) v(da, d\xi_{-1}). \end{aligned}$$

where the first equality comes from assumption 2. This assumption implies also

$$\partial_{x_1} f_{A_{-1}X_{-1}|A_1 X_1 S}(a_{-1}, \xi_{-1}; a_1, x_1, s) = f_{A_{-1}|A_1 S}(a_{-1}; a_1, s) \partial_{x_1} f_{X_{-1}|X_1 S}(\xi_{-1}; x_1, s),$$

and using the transformation $\partial_{x_1} f_{X_{-1}|X_1 S} = [\partial_{x_1} \log f_{X_{-1}|X_1 S}] f_{X_{-1}|X_1 S}$, we obtain that the last term equals

$$\mathbb{E} [Y \partial_{x_1} \log f_{X_{-1}|X_1 S} | X_1 = x_1, S = s, A_1 = a_1].$$

Taking expectations with respect to A_1 and using $\mathbb{E} [\partial_{x_1} \log f_{X_{-1}|X_1 S} | X_1 = x_1, S = s] = 0$, this may be rewritten as $Cov [U, Q_{x_1} | X_1 = x_1, S = s]$. To see the bounds, note that with probability one

$$\begin{aligned} |\partial_{x_1} m(X, S) - \mathbb{E}[\partial_{x_1} \phi(X, A) | X_1, S]| &= \left| \mathbb{E} \left[(Y - \tilde{Y}) \partial_{x_1} \log f_{X_{-1}|X_1 S} | X_1, S \right] \right| \\ &\leq \mathbb{E} \left[\left| (Y - \tilde{Y}) \partial_{x_1} \log f_{X_{-1}|X_1 S} \right| | X_1, S \right] \\ &\leq \min_{(p,q) \in \mathcal{S}} \mathbb{E} \left[|Y - \tilde{Y}|^p | X_1, S \right]^{1/p} \mathbb{E} [|Q_{x_1}|^q | X_1, S]^{1/q}, \end{aligned}$$

by Hölder's inequality. For the case of the quantile, we apply again Hoderlein and Mammen (2007) to $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}] = \psi(W_{-1}, A_1)$ in connection with assumption 2.1. This produces

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1} \psi(X_1, S, A_1) | W = w, \tilde{Y} = k_\alpha(w)].$$

The result follows by using $\partial_{x_1} \psi(X_1, S, A_1) = \mathbb{E}[\partial_{x_1} \phi | X_1, S, A_1] + Cov [U, Q_{x_1} | X_1, S, A_1]$.

Finally, to see the last equality, note first that (with $b = x'\beta$),

$$\begin{aligned} \partial_{x_1} \mathbb{E} [\tilde{Y} | X = x, S = s] &= \partial_{x_1} \mathbb{E} [\mathbb{E} \{\phi(X, A) | B, S, A_1\} | X = x, S = s] \\ &= \mathbb{E} [\partial_{x_1} \mathbb{E} \{\phi(X, A) | B = b, S = s, A_1\} | X = x, S = s] \quad (6.1) \\ &= \beta_1 \mathbb{E} [\partial_b \mathbb{E} \{\phi(X, A) | B = b, S = s, A_1\} | X = x, S = s]. \end{aligned}$$

Next, observe that

$$\begin{aligned} & \mathbb{E} \{ \phi(X, A) | B = b, S = s, A_1 = a_1 \} \\ &= \int_{\mathcal{X}_{-1} \times \mathcal{A}_{-1}} \phi(\tau(b, \xi_{-1}), a) f_{A_{-1}X_{-1}|BSA_1}(a_{-1}, \xi_{-1}; b, s, a_1) v(da_{-1}, d\xi_{-1}), \end{aligned}$$

where $\tau(b, \xi_{-1}) = (\beta_1^{-1}(b - \beta_2\xi_2 - \dots - \beta_K\xi_K), \xi_2, \dots, \xi_K)$, and consequently,

$$\begin{aligned} & \partial_b \mathbb{E} \{ \phi(X, A) | B = b, S = s, A_1 = a \} \\ &= \int_{\mathcal{X}_{-1} \times \mathcal{A}_{-1}} \partial_b [\phi(\tau(b, \xi_{-1}), a) f_{A_{-1}X_{-1}|BSA_1}(a_{-1}, \xi_{-1}; b, s, a_1)] v(da_{-1}, d\xi_{-1}) \\ &= \beta_1^{-1} \mathbb{E} \{ \partial_{x_1} \phi(X, A) | B = b, S = s, A_1 = a \} \\ & \quad + \int_{\mathcal{X}_{-1} \times \mathcal{A}_{-1}} \phi(\tau(b, \xi_{-1}), a) \partial_b f_{A_{-1}X_{-1}|BSA_1}(a_{-1}, \xi_{-1}; b, s, a_1) v(da_{-1}, d\xi_{-1}). \end{aligned} \tag{6.2}$$

Now we focus on the second term on the last rhs of (6.2). To this end, note that by assumption 2.2

$$f_{A_{-1}X_{-1}B|SA_1}(a_{-1}, \xi_{-1}; b, s, a_1) = f_{X_{-1}B|S}(\xi_{-1}; b, s) f_{A_{-1}|SA_1}(a_{-1}; s, a_1),$$

and

$$\begin{aligned} f_{X_{-1}B|S}(\xi_{-1}; b, s) &= \beta_1^{-1} f_{X|S}(\tau(b, \xi_{-1}); s), \\ f_{B|S}(b; s) &= \beta_1^{-1} \int_{\mathcal{X}_{-1}} f_{X|S}(\tau(b, \xi_{-1}); s) d\xi_{-1} \\ f_{X_{-1}|BS}(\xi_{-1}; b, s) &= \frac{f_{X_{-1}B|S}(\xi_{-1}; b, s)}{f_{B|S}(b; s)} = \underbrace{\frac{\beta_1^{-1} f_{X|S}(\tau(b, \xi_{-1}); s)}{\beta_1^{-1} \int_{\mathcal{X}_{-1}} f_{X|S}(\tau(b, \xi_{-1}); s) d\xi_{-1}}}_{H(b, s)}, \end{aligned}$$

with derivatives

$$\begin{aligned} H'(b, s) &= \beta_1^{-1} \int_{\mathcal{X}_{-1}} \partial_{x_1} f_{X|S}(\tau(b, \xi_{-1}); s) d\xi_{-1} \\ \partial_b f_{X_{-1}|BS}(\xi_{-1}; b, s) &= H(b, s)^{-2} \{ \beta_1^{-1} \partial_{x_1} f_{X|S}(\tau(b, \xi_{-1}); s) H(b, s) - f_{X|S}(\tau(b, \xi_{-1}); s) H'(b, s) \}. \end{aligned}$$

Evaluated at $b = \beta'x$

$$\begin{aligned} H(\beta'x, s) &= \int_{\mathcal{X}_{-1}} f_{X|S}(\tau(b, x_{-1}); s) dx_{-1} = f_{X_1}(x_1) \\ H'(\beta'x, s) &= \beta_1^{-1} \int_{\mathcal{X}_{-1}} \partial_{x_1} f_{X|S}(\tau(b, x_{-1}); s) dx_{-1} = \beta_1^{-1} f'_{X_1|S}(x_1; s) \\ \beta_1 \partial_b f_{X_{-1}|BS}(x_{-1}; \beta'x, s) &= \frac{\partial_{x_1} f_{X|S}(x; s)}{f_{X_1|S}(x_1; s)} - f_{X|S}(x; s) \frac{f'_{X_1|S}(x_1, s)}{f_{X_1|S}(x_1; s)^2} \\ &= \partial_{x_1} \frac{f_{X|S}(x; s)}{f_{X_1|S}(x_1; s)} \\ &= \partial_{x_1} f_{X_{-1}|X_1S}(x_{-1}; x_1, s). \end{aligned} \tag{6.3}$$

Combining equations (6.1), (6.2) and (6.3) yields

$$\begin{aligned}\partial_{x_1}\mathbb{E}\left[\tilde{Y}|X=x, S=s\right] &= \mathbb{E}\left[\partial_{x_1}\phi(X, A)|B=b, S=s\right] \\ &+ \mathbb{E}\left[Y\partial_{x_1}\log f_{X_{-1}|X_1S}(x_{-1}; x_1, s)|B=b, S=s\right].\end{aligned}$$

Finally, the quantile follows as direct extension by similar arguments.

Q.E.D.

References

- [1] ALTONJI, J., and R. L. MATZKIN (2005): Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73, 1053–1103.
- [2] BATTISTIN, E., R. MINIACI, and G. WEBER (2003): What do we learn from recall consumption data? *Journal of Human Resources*, 38, 354–385.
- [3] BOUND, J., C. BROWN, and N. MATHIOWETZ (2001): Measurement error in survey data. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, 3705–3843. Amsterdam: Elsevier.
- [4] BROWNING, M., T. F. CROSSLEY, and G. WEBER (2003): Asking consumption questions in general purpose surveys. *Economic Journal*, 113, F540–F567.
- [5] CARROLL, R. J., D. RUPPERT, and L. A. STEFANSKI (1995): *Measurement Error in Nonlinear Models*. New York: Chapman & Hall.
- [6] DAS, M., J. DOMINITZ, and A. VAN SOEST (1999): Comparing predictions and outcomes: Theory and application to income changes. *Journal of the American Statistical Association*, 94(445), 75–85.
- [7] DATTA, S. (1995): On a Modified Bootstrap for Certain Asymptotically Nonnormal Statistics, *Statistics and Probability Letters*, 24, 91–98.
- [8] GROVES, R. M. (1989): *Survey Errors and Survey Costs*. New York: Wiley.
- [9] HAUSMAN, J. A. (2001): Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives*, 15(4), 57–67.
- [10] HAUSMAN, J. A., W. K. NEWKEY, and J. L. POWELL (1995): Nonlinear errors in variables estimation of some Engel curves. *Journal of Econometrics*, 65, 205–233.
- [11] HODERLEIN, S. (2005): Nonparametric demand systems, instrumental variables, and a heterogeneous population. Working Paper, University of Mannheim.
- [12] HODERLEIN, S. (2008): How many consumers are rational. Working Paper, University of Mannheim.

- [13] HODERLEIN, S., and E. MAMMEN (2007): Identification of local average structural derivatives in nonseparable models. *Econometrica*, 75(5), pages 1513-1518
- [14] HODERLEIN, S. and J. WINTER (2008): Recall errors in surveys. Working Paper, University of Mannheim and University of Munich.
- [15] HU, Y., and S. SCHENNACH (2006): Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions using instruments. Working Paper, University of Chicago.
- [16] HURD, M., and S. ROHWEDDER (2005): The Consumption and Activities Mail Survey: Description, data quality, and first results on life-cycle spending and saving. Working Paper, RAND, Santa Monica.
- [17] HYSLOP, D. R., and G. W. IMBENS (2001): Bias from classical and other forms of measurement error. *Journal of Business and Economic Statistics*, 19, 475–481.
- [18] IMBENS, G. W., and W. NEWHEY (2003): Identification and estimation of triangular simultaneous equations models without additivity. Working Paper, MIT.
- [19] JUSTER, F. T., and R. SUZMAN (1995): An overview of the Health and Retirement Study. *Journal of Human Resources*, 30, S7–S56.
- [20] MANSKI, C. (1990): The use of intentions data to predict behavior: A best-case analysis. *Journal of the American Statistical Association*, 85, 934–940.
- [21] ————— (2004): Measuring Expectations. *Econometrica*, 72, 1329–1376.
- [22] MATZKIN, R. L. (2007): Heterogeneous Choice, *Working Paper*, UCLA.
- [23] MATZKIN, R. L. (2008): Nonparametric survey response errors, *International Economic Review*, 48, 1411–1427.
- [24] McFADDEN, D. (2006): How consumers respond to incentives. Working Paper, UC Berkeley.
- [25] PHILIPSON, T. (1997): Data markets and the production of surveys. *Review of Economic Studies*, 64(1), 47–72.
- [26] ————— (2001): Data markets, missing data, and incentive pay. *Econometrica*, 69(4), 1099–1111.
- [27] TOURANGEAU, R., L. J. RIPS, and K. RASINSKI (2000): *The Psychology of Survey Response*. New York, NY and Cambridge, UK: Cambridge University Press.
- [28] WANSBEEK, T. and E. MEIJER (2000): *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier.

Appendix 2: Descriptive Statistics - For the Information of the Referees

Table A1: Descriptive statistics of three groups in CAMS reported weekly, monthly and annual food consumption

	Food	Income	couple	age	edyrs	hisp	male	race	shlt
mean	0.12	10.42	0.57	68.33	12.60	0.05	0.35	0.10	2.71
min	0.00	8.79	0.00	28.00	0.00	0.00	0.00	0.00	1.00
5p	0.02	9.15	0.00	54.00	8.00	0.00	0.00	0.00	1.00
50p	0.09	10.42	1.00	67.00	12.00	0.00	0.00	0.00	2.50
95p	0.32	11.69	1.00	85.00	17.00	0.00	1.00	1.00	4.50
max	0.59	12.13	1.00	100.00	17.00	1.00	1.00	1.00	5.00

Table A1 shows descriptive statistics of the population. The presented variables are food (budget share of food expenditure), income (log total household income per year in dollars), couple (0 = no, 1 = yes), age (years), education (years), hispanic (0 = no, 1 = yes), male (0 = no, 1 = yes), race (0 = nonblack, 1 = black), and self-reported health status (1 = excellent, ..., 5 = poor).

Table A2: Descriptive statistics of three groups in CAMS reported weekly, monthly and annual food consumption

	Food	Income	couple	age	edyrs	hisp	male	race	shlt
Weekly food consumption									
mean	0.14	10.44	0.57	68.11	12.45	0.05	0.33	0.11	2.70
min	0.00	8.80	0.00	28.00	0.00	0.00	0.00	0.00	1.00
5p	0.03	9.19	0.00	54.00	8.00	0.00	0.00	0.00	1.00
50p	0.11	10.42	1.00	67.00	12.00	0.00	0.00	0.00	2.50
95p	0.36	11.70	1.00	85.00	17.00	0.00	1.00	1.00	4.50
max	0.59	12.13	1.00	96.00	17.00	1.00	1.00	1.00	5.00
Monthly food consumption									
mean	0.11	10.35	0.54	68.06	12.65	0.06	0.31	0.12	2.77
min	0.00	8.80	0.00	34.00	0.00	0.00	0.00	0.00	1.00
5p	0.02	9.10	0.00	53.05	8.00	0.00	0.00	0.00	1.00
50p	0.08	10.37	1.00	67.00	12.00	0.00	0.00	0.00	2.50
95p	0.29	11.62	1.00	85.00	17.00	1.00	1.00	1.00	4.50
max	0.59	12.13	1.00	96.00	17.00	1.00	1.00	1.00	5.00
Annual food consumption									
mean	0.08	10.48	0.59	69.02	12.80	0.05	0.42	0.09	2.68
min	0.00	8.79	0.00	40.00	0.00	0.00	0.00	0.00	1.00
5p	0.01	9.15	0.00	55.00	8.00	0.00	0.00	0.00	1.00
50p	0.06	10.49	1.00	68.00	12.00	0.00	0.00	0.00	2.50
95p	0.21	11.71	1.00	85.00	17.00	0.00	1.00	1.00	4.50
max	0.55	12.13	1.00	100.00	17.00	1.00	1.00	1.00	5.00

Table A2 shows descriptive statistics of three groups in CAMS reported weekly, monthly and annual food consumption. The presented variables are food (budget share of food expenditure), income (log total household income per year in dollars), couple (0 = no, 1 = yes), age (years), education (years), hispanic (0 = no, 1 = yes), male (0 = no, 1 = yes), race (0 = nonblack, 1 = black) and self-reported health status (1 = excellent, ..., 5 = poor).

Table A3: Descriptive statistics of individuals reported weekly food consumption in CAMS and HRS Survey

	Food	Income	couple	age	edysr	hisp	male	race	shlt	srn
Weekly food consumption in CAMS										
mean	0.14	10.35	0.46	68.93	12.44	0.04	0.35	0.11	2.72	2.90
min	0.00	8.81	0.00	29.00	0.00	0.00	0.00	0.00	1.00	1.00
5p	0.03	9.21	0.00	56.00	8.00	0.00	0.00	0.00	1.00	2.00
50p	0.11	10.31	0.00	68.00	12.00	0.00	0.00	0.00	2.50	3.00
95p	0.36	11.58	1.00	85.00	17.00	0.00	1.00	1.00	4.50	4.00
max	0.58	12.13	1.00	96.00	17.00	1.00	1.00	1.00	5.00	5.00
Weekly food consumption in HRS										
mean	0.13	10.31	0.46	68.90	12.44	0.04	0.35	0.11	2.72	2.90
min	0.02	8.78	0.00	30.00	0.00	0.00	0.00	0.00	1.00	1.00
5p	0.03	9.18	0.00	56.00	8.00	0.00	0.00	0.00	1.00	2.00
50p	0.10	10.25	0.00	67.00	12.00	0.00	0.00	0.00	3.00	3.00
95p	0.32	11.56	1.00	86.00	17.00	0.00	1.00	1.00	5.00	4.00
max	0.59	12.16	1.00	97.00	17.00	1.00	1.00	1.00	5.00	5.00

Table A3 shows descriptive statistics of individuals reported weekly food consumption in CAMS and HRS Survey. The presented variables are food (budget share of food expenditure), income (log total household income per year in dollars), couple (0 = no, 1 = yes), age (years), education (years), hispanic (0 = no, 1 = yes), male (0 = no, 1 = yes), race (0 = nonblack, 1 = black), self-reported health status (1 = excellent, ..., 5 = poor) and self-rated memory (1 = excellent, ..., 5 = poor).