Recall Errors in Surveys

Stefan Hoderlein* Joachim Winter
University of Mannheim University of Munich

May 4, 2007

Abstract

This paper considers measurement errors from a different perspective. Concentrating on errors that are caused by the insufficient ability of individuals to acquire, process and recall information, we devise a structural model for the response behavior of individuals in surveys. Our aims are twofold: Using nonseparable models, we first explore the consequences of such a modeling approach for key econometric questions like the identification of marginal effects and economic restrictions. We establish that under general conditions recall measurement errors are likely to exhibit nonstandard behavior, in particular be nonclassical and differential, and we provide means to deal with this situation. Moreover, we obtain surprising findings indicating that conventional wisdom about measurement errors may be misleading in many economic applications. For instance, under certain conditions left hand side and right hand side errors will be equally problematic. The second aim of this paper is to provide a formal framework for understanding the actual response behavior of respondents in surveys. Specifically, we provide formal arguments and explanations for the frequently encountered substantial effects of the design of a survey. Also, we suggest a list of issues that should be considered when eliciting a survey to avoid adverse effects. Finally, we apply the main concepts put forward in this paper to real world data and to a controlled experiment, and find evidence that underscores the importance of focusing on individual response behavior, as well as taking the design of a survey seriously.

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

*Department of Economics, L7, 3-5, 68131 Mannheim, Germany, email: 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, Lars Nesheim, Carmit Segal, Melanie Schienle and Jörg Stoye. We would also like to thank particularly Daniel McFadden for encouraging us to pursue this line of research. Financial support by Landesstiftung Baden-Württemberg "Eliteförderungsprogramm" and the European Union (grant no. ERBFMRCT960016) is gratefully acknowledged.

1 Introduction

Consider a typical question in a survey questionnaire about a random variable Z like the following: "How much did you spend on food items in the last month?" 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 information to their immediate avail. Obviously, this recall $\tilde{Z}$ is only an imperfect measurement of the true food consumption Z, and any analysis based on this imperfect measurement may potentially be flawed. Traditionally, in order to deal with the difficulties arising in this situation, the Econometrics literature focussed on the error in the measurement of Z, ie on $Z - \tilde{Z}$.

In this paper we reverse the viewpoint, and focus on properties of the backcast $\tilde{Z}$ instead. Unlike in the natural sciences where (mis)measurements are given exogenously, in surveys in the social sciences individuals actively decide about their backcast, and the measurement error is correspondingly only a byproduct of this process. Somewhat provocatively we could say that people form backcasts, not errors. Consequently, we argue in this paper that the survey design and measurement error literature should be more concerned with the backcast $\tilde{Z}$.

Reversing the viewpoint has several advantages: First, it enables us to give new answers to old problems like the correct measurement of marginal effects. Second, we obtain new insights into objects previously considered unproblematic like mismeasurements in the dependent variable. Under our modeling framework, these objects will turn out to be likely to exhibit "nonstandard" behavior. Third, we provide a link to the literature on eliciting expectations and probabilistic questions as reviewed in Manski (2004). Finally, with the tools we suggest it is possible to address issues in the design of a survey formally, in order to provide means to minimize the arising distortions ex ante. 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 with the interviewer, nor do we consider the case where Z is elicited accurately, but is only taken as a proxy. Specifically, we are concerned with the respondents' limited ability to acquire, process, and recall information. Hence, we focus on this key type of measurement error in detail, and we propose a structural model of individual behavior to analyze the response behavior of individuals formally. The main feature of this model is bounded rationality. Individuals are boundedly rational in the sense that they still try to find an optimal solution, given their limitations to recall information. Moreover, we allow for heterogeneity of individuals in preferences and mental abilities in a very general fashion.

Relationship to the Literature: Even though our approach contributes to the literature on nonparametric identification using nonseparable models, the fact that we consider a structural model of response behavior distinguishes our approach from recent work in econometrics on mismeasured variables. In this literature, unspecified measurement errors in nonlinear models are treated under one general format. Key recent contributions are, inter alia, Hausman, Newey, and Powell (1995) for classical measurement errors, and Hu and Schennnach (2006) for nonclassical measurement errors. For additional literature, see references therein 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). In this literature, closely related in terms of the nonseparable framework employed is Matzkin (2007).

This literature is, at least compared to our approach, of fairly reduced form. There is no behavioral model that explains the respondents' error, and consequently the error is a rather 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 a crucially important one. However, with respect to this one type of measurement error we can describe more accurately determinants and implications. Consequently, we may also give guidance to those researchers and institutions who actually design survey questionnaires.

Thus, from the substance our approach is more closely related to work in survey research and psychology, for an overview see Tourangeau, Rips, and Rasinski (2000) and Moore, Stinson and Welniak (1999). Indeed, our work contributes a framework that allows to formalize various notions in this literature, and to link it to econometrics via nonparametric identification. There is also a strand of work in economic theory which considers the response behavior under a market perspective, see Philipson (1997, 2001). This approach is very different from ours because it focusses on the interaction between rational economic agents (the survey respondent on the one hand, and the researcher or survey field organization on the other), while we are primarily concerned with the boundedly rational behavior of a single individual, and its econometric implications.

Conceptually, there is some overlap between this paper and the work of Manski and his collaborators, see Manski (2004) for an overview. In particular, we will adopt the strong case he makes for the elicitation of entire conditional probability distributions when eliciting expectations (as opposed to point estimates), transfer this notion to the elicitation of responses to already occurred, but incompletely remembered variables, and establish formally their advantages. In addition to transfer and formalization, we will add some new ideas to this paradigm like the elicitation of co-distributions. What, however, distinguishes our approach materially from Manski's is our emphasis on limited rationality of individuals, and the corresponding consequences for the individuals' information set.

Key Issues Considered: In considering the behavior of respondents in detail and relating it to nonparametric identification, we aim specifically at four important issues:

1. *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? The second question is concerned with less basic requirements, namely: 2. *Can we identify core elements of economic theory as embodied in economic restrictions?* Of course, this question has to be answered for every restriction in isolation. However, we discuss very briefly the key integrability conditions arising from utility maximization as an example and highlight additional difficulties.

These two issues can be seen as classical identification questions. Since this paper is explicitly concerned with the consequences for survey design research, two additional issues arise: 3. *What is the role of survey design?* Specifically, what additional variables should be elicited and what is the role of embedded survey experiments? This question is obviously of great importance for the survey design literature and goes vis-a-vis the following: 4. *What kind of additional restrictions does a formal model of survey response behavior generate?* Indeed, formalizing the process of distortions caused by the design of a survey may lead to testable implications that allow judging the validity of the model as well as the implications for the above questions.

Illustration of Contributions: Instead of enumerating the contributions, we believe it to be more useful to illustrate our ideas using a simple example. While it is impossible to illustrate all ideas in this paper by a single example, some important issues may be found in the following textbook example of a linear model[1]. In this example we focus on the hitherto believed to be unproblematic case of a dependent (left hand side) variable being imperfectly recalled. We assume that the model is defined as follows: Let

$$Y = X'\beta + A, \tag{1.1}$$

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 $\mathcal{F}_m$. Loosely speaking, it is the richness of this information set that will be determined by the decision of the boundedly rational individuals to balance costs and profits of recalling. To this end, we introduce a structural model of recall behavior in section 2 which allows to understand the determinants of the individuals' decision and hence open up ways to influence it.

In this example, however, we focus on 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 $\mathbb{E}\left[\tilde{Y}|X\right] = X'\tilde{\beta}$. By standard arguments, $\tilde{\beta} = \mathbb{E}\left[XX'\right]^{-1}\mathbb{E}\left[X\tilde{Y}\right]$, provided $\mathbb{E}\left[XX'\right]$ is nonsingular. Suppose first that when backcasting individuals use the squared error loss, and assume that $\mathcal{F}_m \supseteq \sigma\left(X\right)$.

[1]Throughout 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 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.

Then, $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}_m]$, and

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

But due to $0 = \mathbb{E}[A|X] = \mathbb{E}[\mathbb{E}[A|\mathcal{F}_m]|X]$ 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 error does not matter for identification in linear models.

However, turn to the scenario when individuals do not actively use all information available to the econometrician. An example would be when an individual is being asked about his food consumption, but does not actively memorize the constellation of prices and incomes under which he has chosen a particular bundle. Say, the respondent has a fixed income-demand ratio, but forgot about a temporary reduction in income and stated that the associated demand stayed constant. When recording his income at a later stage of the survey, however, the respondent recalled it because he was reminded (say, by the wording of the question) of a spell of unemployment and the associated drop in income. What are the likely consequences? Suppose all individuals showed this pattern, and suppose these temporary income drops were the only movement in the income process. Then there would be no association between income and demand in the data whatsoever – but not because there is none, only because the individuals do not recall it.

Formally, consider the case where, in order to form a backcast individuals use the squared error loss in connection with $\mathcal{F}_m = \sigma(X'\gamma, A_1)$, where γ is a nonrandom nonzero vector and A_1 contains additional information not available to the econometrician. This means that for the purpose of backcasting individuals reduce information given to the econometrician, but they do not discard any single regressor (as is the case with income in the demand example). Assume for simplicity that $\mathbb{E}[A|\mathcal{F}_m] = \mathbb{E}[A|A_1] \neq 0$ and that A_1 is independent of X. Then,

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

where $V = X - \mathbb{E}[X|X'\gamma]$. Obviously, the 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.

Is this an effect of great importance in practise? We argue that this is indeed the case, also because this and the preceding demand example are parts of a wider class of phenomena caused by individuals reducing the information available to the econometrician. In the application, we will provide empirical and experimental evidence that similar phenomena are quantitatively important and may arise in many situations, including some where they arise rather unexpectedly. It is also surprising that quantiles are less robust than mean regressions in the presence of recall error.

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. Because of the importance of the decision to actively use all information of interest to the econometrician in the recall, it is imperative to understand the process by which this decision is made, both to give an estimate for the magnitude of the effect and also to come up with possible remedies. In section 2, we will do so by way of economic reasoning. The model we propose may for instance be used to design and focus economic incentives that allow to improve the quality of answers.

However, it is our opinion that there will always be limitations to the quality of the backcast. Indeed, recall errors are likely to occur even with improved surveys. Moreover, the recall errors are bound to be misbehaved. For instance, it is straightforward to see 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 in several theorems. 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:

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

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 $\mathbb{E}\left[Y^2\right]$ 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, $\mathbb{E}\left[U^2\right]$, then we may replace $\mathbb{E}\left[Y^2\right]$ by $\mathbb{E}\left[\tilde{Y}^2\right] + \mathbb{E}\left[U^2\right]$. Third, we may come up with alternative bounds which replace $\mathbb{E}\left[Y^2\right]$ by $\mathbb{E}\left[Var\left[Y^2|\mathcal{F}_m\right]\right]$, where $Var\left[Y^2|\mathcal{F}_m\right]$ may be obtained by eliciting a measure of the individuals' uncertainty. This line of thought connects our approach to that of Manski (2004). Third, the assumption about the structure of the recall information shows up in the second term $\mathbb{E}\left[Q_{x_1}^2\right]$, 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.

Implications for Survey Design: Other than providing a conceptual framework that allows fixing ideas, our contributions to the survey design literature can be summarized in the following guidance: 1. When asked for one specific variable, the respondents should be encouraged to actively memorize *all* variables of importance for the economic decision involving this variable. 2. When trying to recall their decisions, respondents should rather be encouraged to report the average decision than the median ("typical") decision, provided these differ substantially. 3. Recall problems are aggravated if a poorly recalled variable is the regressor of interest. Hence, these variable should be elicited with particular care. 4. If instruments are to be employed they should be highlighted to the respondent so that they really use it when forming their response. 5. A measure of precision of the answer should be elicited. 6. Ideally, the entire conditional distribution of the backcast should be elicited, at least for dependent variables and regressors of interest. 7. Measures of cognitive ability, time used for answering and time elapsed since decision should be elicited, alongside with some measure of opportunity costs of time. 8. If brackets are used then there should be variation in brackets across the population. 9. Co-distributions as defined in section 3.2.3. should be elicited. 10. If marginal effects are of interest, then they should be elicited directly as well.

Organization of the Paper: In terms of structuring the material, the subject of the paper presents quite a challenge. The reason is that the main contribution of the paper is a change of viewpoint, and not a single theorem. Hence, it is imperative to be make the structure of the paper as transparent as possible, given this difficulty.

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. 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

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

and $A \perp X$. In **section 2**, we provide a model of boundedly rational recall behavior that helps to understand the determinants and the process by which these backcasts $\tilde{Y}$, respectively $\tilde{X}_1$, are formed. We formalize the notion that individuals decide actively upon their backcast (which is either of $\tilde{Y}$ or $\tilde{X}_1$), given their limitations to process information. Readers solely interested in the statistical details may skim this section as it is mainly concerned with the decision theoretic background for modeling individuals' behavior.

In **section 3** we consider the implications of such a behavior for the measurement of the partial derivative ∂_{x_1}. We consider separately the case where the dependent variable Y is poorly recalled in **section 3.1**, and where the independent variable X_1 is poorly recalled in **section 3.2**. As it turns out, the crucial question is whether individuals reduce the econometricians' information set when backcasting. The answer to this question gives further structure to these two subsections. Finally, we briefly discuss additional difficulties that arise in the case of more complicated economic hypotheses in **section 3.3**.

In **section 4** we elaborate on the testable implications. Specifically, we show how various structural hypotheses lead to formal test statistics, and we suggest a method of hypothesis testing in the presence of bounds caused by information reduction. In **section 5** we suggest procedures based on the bootstrap to implement all test statistics. We provide rigorous large sample theory for one typical bootstrap test, and discuss the validity of the others heuristically.

But it is really only in **section 6** when all these various elements are brought together and their usefulness is analyzed through two applications. The first illustrates in a real world data set what can be learned in light of our theory about the distortions in response behavior of individuals, and how these distortions can be handled. The second is a controlled experiment, which illustrates the scope of the effect we are considering through variations in brackets. Finally, summary and outlook conclude.

2 A Formal Model of Boundedly Rational Survey Response Behavior

The purpose of this paper is to model the impact of imperfect recollection of a random variable Z, which will either be the dependent variable Y, or the regressor of interest X_1. We propose that rational individuals try to balance the costs and benefits of memorizing to obtain an optimal "amount" of memory. Formally, assume that there is a collection of infinitely many random variables (ξ_s), $\xi_s \in \mathbb{R}$, $s \in \mathbb{R}_{[0,1]}$, each of which can be thought of as giving one standardized "unit of information". Let $\mathcal{F}_m \equiv \sigma\{\xi_s | 0 \leq s \leq m\}$ denote the σ-algebra spanned by $(\xi_s)_{0 \leq s \leq m}$, and note that by construction $\mathcal{F}_{m+\delta} \supseteq \mathcal{F}_m \supseteq \mathcal{F}_{m-\delta}, \ldots$, $\delta > 0$, i.e. $\{\mathcal{F}_m, 0 \leq m \leq 1\}$ is a filtration. Let $\mathbb{E}[\cdot|\mathcal{F}_m]$ denote the conditional expectation given the σ-algebra $\mathcal{F}_m$. To determine the optimal amount of information, the individual chooses now a finite number m^*, where m^* is defined as

$$m^* = \arg\max_{m \in [0,1]} \mathbb{E}\left\{P_1\left[L_0 - L\left(Z, \tilde{Z}_m\right)\right] - c\left(m, P_2\right)|\mathcal{F}_0\right\}, \tag{2.1}$$

where $P = (P_1, P_2')' \in \mathbb{R} \times \mathcal{P}_2$ is a (possibly infinite dimensional) individual specific parameter which may vary across the population (think of prices), such that $\sigma(P) \subset \mathcal{F}_0$. Moreover, L is a standard loss function defined on $\mathbb{R} \times \mathbb{R}$ and $\tilde{Z}_m$ denotes the individual's forecast of Z for a fixed sigma algebra $\mathcal{F}_m$. Hence $\tilde{Z}_m = g\left((\xi_s)_{0 \leq s \leq m}\right)$, where g is a functional mapping the process $(\xi_s)_{0 \leq s \leq m}$ into $\mathbb{R}$, and $L_0 = \mathbb{E}\left[L\left(Z, \tilde{Z}_0\right)|\mathcal{F}_0\right]$. The leading example is when L is the squared error loss so that $L(z, z_m) = (z - z_m)^2$, and $\tilde{Z}_m = \mathbb{E}[Z|\mathcal{F}_m]$ (i.e., g is the conditional expectation operator). Finally, c is a nonrandom cost function giving the minimal costs of building up memory m for every p_2.

Because of the law of iterated expectations we may rewrite the optimization problem (2.1):

$$m^* = \arg\max_{m\in[0,1]} \{P_1 l_0(m) - c(m, P_2)\},$$

where

$$l_0(m) = \mathbb{E}\left\{L_0 - \mathbb{E}\left[L\left(Z, \tilde{Z}_m\right)|\mathcal{F}_m\right]|\mathcal{F}_0\right\}.$$

Note that $l_0(\cdot)$ is a monotonically increasing function of m because $\mathbb{E}\left[L\left(Z, \tilde{Z}_m\right)|\mathcal{F}_0\right] \geq \mathbb{E}\left[L\left(Z, \tilde{Z}_{m+\delta}\right)|\mathcal{F}_0\right]$ for every $\delta > 0$ as the set of potential optimizers is increasing. Moreover, we maintain the assumption that P_1 and P_2 have no elements in common which allows us to separate both parts of the optimization problem. This is plausible since the rewards individuals obtain should not enter the cost function. Hence, for fixed m individuals first minimize the expected loss $\mathbb{E}\left[L\left(Z, \tilde{Z}_m\right)|\mathcal{F}_m\right]$ by choosing $\tilde{Z}_m$ for every fixed m, and then pick the m that minimizes the whole expression.

We now discuss the building blocks of the individual's optimization problem.

- $\Pi(m) = P_1 l_0(m)$ can be interpreted as the profit associated with choosing m. For all commonly used loss functions, Π is a concave function of m. A further implication is that $\mathbb{E}\left[L\left(Z, \tilde{Z}_m\right)|\mathcal{F}_m\right] \leq \mathbb{E}\left[L\left(Z, g\left((\xi_s)_{0\leq s\leq m}\right)\right)|\mathcal{F}_m\right]$, for all other functionals g, and all m. Hence, $\tilde{Z}_m$ is the optimal predictor for fixed m, and (under some differentiability and interiority conditions) the following well known (and principally testable) first order condition holds:
$$\mathbb{E}\left[\partial_{z_m} L\left(Z, \tilde{Z}_m\right)|\mathcal{F}_m\right] = 0. \tag{2.2}$$
Other than the squared error loss, we also consider the case when L is the absolute error loss, in which case the function of the data under consideration is the conditional median (denoted $med_Z(G)$, given the random vector G). We focus on these two since we expect the individuals to use some notion of average when being asked about point predictions of past behavior[2].

- The cost function $c(m, p_2)$ can be seen as the optimizer of the cost minimization problem of building up memory m, i.e. it solves the problem:
$$\min_{z\geq 0} \ b'z \quad \text{s.t.} \quad h(z, \lambda) \geq m,$$
where $h : \mathbb{R}^l \times \mathcal{L} \to (0, 1)$ denotes the memory production function that maps the l-vector of input factors $z \in \mathbb{R}^l$ and parameters $\lambda \in \mathcal{L}$ into the unit interval, and $b \in \mathbb{R}^l_+$ denotes

[2] Of course, this could be easily extended. The idea to attach different interpretations to individuals' response behavior regarding "expectations" is explored empirically in Das, Dominitz, and van Soest (1999).

the prices associated with these inputs. Note that $p_2 = (b', \lambda')'$, and that because of standard producer theory the factor demands $z = \varphi(m, b, \lambda)$ obtain some structure, e.g. that the matrix of price derivatives for fixed λ and m is negative semidefinite or that demands be zero homogenous in b.[3]

- The parameter P : we have chosen the letter P to denote parameters to emphasize the economic association between the parameters and prices. An example for P_2 is the price or opportunity cost for the time needed to answer the survey, an example for P_1 is the price (or the reward) an individual obtains from answering correctly. We think primarily of money, as proposed in Philipson (2001) or McFadden (2006).

Our notion of bounded rationality is a formal one, and we believe that individuals still try to behave optimally, given their constraints. This assumption may indeed be criticized as requiring individuals to act overly rational – they have to solve a potentially complicated optimization problem. However, as most economic theory this should be seen as approximation of reality, where individuals choose the effort to "backcast" according to some intention.

The advantages of setting up a formal model instead of a specifying a response heuristic are manifold: First, if individuals act (at least approximately) as our model assumes them to do, then we may obtain strong testable implications. These are in particular the rational demand structure on the factors needed to build up memory, as well as the optimality condition for the optimal backcast. If we find these rejected using nonparametric tests detailed below on real data, then we may conclude that individuals do not balance costs and profits of memorizing when responding to a survey, and some other explanation would have to be considered. However, we will see below that the empirical regularities are reasonably well explained by the model. We call for experiments to determine incentives P_1 and loss function L, while being aware of the problem of validation of the answer.

Second, if accurate, we may use the model to obtain structural predictions. In particular, it may provide us with a welfare or money measure of the incentive we would have to provide to improve the individual's response behavior (see also McFadden, 2006). Ultimately, this may allow to assess the total costs of improving the quality of the data, and help set up a decision problem for researchers or survey field agencies that administer household surveys. We will leave such an analysis for future research, but touch upon estimation of the cost function in our application.

Finally, a formal model allows to structure the thoughts about the problem. For instance, a key insight of formalizing individual behavior as above is that individuals may find it optimal to choose a sigma algebra $\mathcal{F}_m$ coarser then the one spanned by, say, their stochastic income

[3]These two elements formalize the notion of optimization, and make the "economic" association clear. There are also some parallels with existing concepts in statistical decision theory: for fixed m and p, Π_0 is formally similar to the Bayes Risk. However, note that in the Bayesian framework it is a (random) parameter that is of interest whereas in our case it is precisely the random vector Z. Moreover, the dependence on m and the focus on heterogeneity via the parameter vector P is novel.

process, or spanned by other data an econometrician has to his disposal. Or, more likely, choose an information set that contains a reduced form of the econometricians' information, but also additional information. Working out the (testable) implications will be the topic of the next sections.

3 Effects of Boundedly Rational Survey Response Behavior on Identification of Marginal Effects and Economic Restrictions

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. More formally, let $Y \in \mathbb{R}$ denote the dependent variable (for simplicity a scalar), and $X \in \mathbb{R}^K$ denote regressors. Throughout this paper, 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 $\mathcal{A}$. For instance, in the case of demand for a single good, Y would be a budget share, 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:

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. Therefore, a large class of models falls into this category. Analysis of this model has been considered by Altonji and Matzkin (2006), Imbens and Newey (2005) and Hoderlein (2006) for mean regression, as well as by Hoderlein and Mammen (2006) for quantiles. Like these papers, we will be interested in local average structural derivatives (LASD) with respect to some σ-algebra, e.g., $\mathbb{E}\left[\partial_{x_1}\phi(X, A)|X = x\right]$. 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 welll as $\mathbb{E}[A] = 0$, then $\mathbb{E}\left[\partial_{x_1}\phi(X, A)|X = x\right] = \beta_1$. 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 $\mathbb{E}\left[\partial_{x_1}\phi(X, A)|X = x\right] = \bar{\beta}_1$. Finally, if the underlying model is an index model, $Y = G(X'\beta, A)$ with $A \perp X$, then $\mathbb{E}[\nabla_x \phi(X, A)|X = x] = \beta$, up to scale normalization. Note that from the LASD average marginal effects $\mathbb{E}[\partial_{x_1}\phi(X, A)]$ are straightforwardly obtained by simply averaging.

In this section we give results on the first two questions posed in the introduction. We will consider two types of scenarios that roughly correspond to the classical left-hand-side versus right-hand-side measurement error distinction in parametric econometrics: Z, the variable to be recalled, is either $Y \in \mathcal{Y} \subseteq \mathbb{R}$ or $X_1 \in \mathcal{X} \subseteq \mathbb{R}$. Note that we focus on scalars[4]. Throughout this paper, we need the following notation: Let F_G be the cumulative distribution function of a random variable G, and denote by $F_{G|H}$ the conditional cdf of G given H, where 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). For the optimal backcast $\tilde{Z}_m$, for simplicity of exposition we drop the subscript m, so that answers (ie, optimal backcasts) will be denoted either $\tilde{Y}$ or $\tilde{X}_1$.

3.1 Marginal Effects if the Dependent Variable is not Fully Recalled

3.1.1 Motivation and Assumptions

In this scenario, as a result of incomplete recollection we observe the data $D = (\tilde{Y}, X', S')'$. Here, $\tilde{Y}$ denotes either $\mathbb{E}[Y|\mathcal{F}_m]$ if individuals use the squared error loss, or the median $med_Y(\mathcal{F}_m)$ if they use the absolute error loss, X denotes variables in the economic relationship of interest, S denote factors employed in the recall (eg, time used or mental capacity). Moreover, let $W = (X', S')'$ and introduce the partitions $X = (X_1, X_{-1}')'$, $W = (X_1, W_{-1}')'$, i.e. $W_{-1} = (X_{-1}, S')'$, and $A = (A_1', A_{-1}')'$. The first question posed in the introduction may be reformulated as follows: what can we learn from data about the derivative with respect to a typical economic variable of interest, i.e. $\partial_{x_1}\phi$? To answer this question, all we have to our disposal is the joint distribution of the data F_D. We focus on implications for the mean regression function $\mathbb{E}\left[\tilde{Y}|W = w\right] = m(w)$, 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. $F_{A|XS} = F_{A|S}$.

[4]Good parts of the theory we present in this paper could be extended towards random vectors of higher dimension. However, we will extensively consider quantiles, a concept which is clearly not straightforward to generalize to higher dimensions, see Hoderlein and Mammen (2006) for a discussion in the context of the multivariate nonseparable model.

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 the working people 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 short experience *within this subpopulation*, the proportion of type 1 and 2 must be identical. The second assumption strengthens the first so that all X are independent of unobservables.

The final obstacle we have to remove for the model of section 2 to be operational in applications is to relate the theoretical object $\mathcal{F}_m$ to observables. As will turn out, 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 $\mathcal{F}_m$ and the variables at the disposal of the econometrician is given by either of*

1. $\mathcal{F}_m = \sigma(W, J) \supseteq \sigma(W)$.
2. $\mathcal{F}_m = \sigma(\varphi(W), J)$ *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 information[5]. In order to assess all implications of our assumptions, we require some regularity conditions, which may be found in the appendix. They essentially restrict our analysis to continuously distributed (recalled) dependent variables. In the case of mean regressions we could, of course, also consider discrete random variables, but since we want to have unified and comparative treatment with the conditional quantile (where a differentiable quantile does not make much sense), we desist from this straightforward extension.

3.1.2 Individuals Actively use all Information Available to 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,

[5] The 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.

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 12 hold. 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 13 holds instead of assumption 12, we obtain for derivatives of the conditional mean, for all w:

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

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**).

Extensions: 1. Endogenous Regressors and Instruments: It is interesting to consider 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. In the mean regression case,

$$\partial_{x_1} m(x_1, w_{-1}) = \mathbb{E}[\partial_{x_1}\phi(X, A)|X_1 = x_1, W_{-1} = w_{-1}] + Cov[Y, \bar{Q}_{x_1}|X_1 = x_1, W_{-1} = w_{-1}], \quad (3.1)$$

where $\bar{Q}_{x_1} = \partial_{x_1} \log f_{A|X_1W_{-1}}(a; x_1, w_{-1})$, 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), 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 $\mathcal{F}_m^* = \sigma(W^*, J) \supseteq \sigma(W^*)$ 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.

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 $\partial_{x_1}\mathbb{E}\left[\tilde{Y}|W = w\right] \neq \mathbb{E}\left[\partial_{x_1}\phi(X, A)|W = w\right]$ 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 $\tilde{Y} = med_Y(W, A_1) = \psi(W, A_1)$ is strictly monotone in the scalar second component, with assumption **2.1** strengthened to $A_1|W \sim \mathcal{U}[0, 1]$, then

$$\partial_{x_1} k_\alpha(w) = \mathbb{E}\left[\partial_{x_1}\phi(X, A)|W = w, A_1 = \alpha, Y = med_Y(w, \alpha)\right], \quad (3.2)$$

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

3. Eliciting Entire Distributions: Nevertheless, it is possible to extract information about sensible averages from quantiles, even without assuming monotonicity. The interesting point, however, is that in order to do so the individuals have to report all α-quantiles of their conditional recall distribution, not just the median. Average effects may then be obtained by averaging over α. Elicitation of the entire conditional distribution is also put forward in Manski (2004).

4. Heterogeneity in σ-Algebras: A final question that remains to be answered is the question about heterogeneity of the size of the individuals' sigma algebras[6]. How do we formalize this notion in our setup? Suppose there are two groups, one with low information $\underline{m}$, one with high information $\overline{m}$. Suppose further that the group an individual falls into is assumed to depend on whether a random variable ξ falls into a set G, e.g. whether net profits of memorizing exceed a certain threshold or not. Assume that $\mathcal{F}_{\overline{m}} = \sigma(W, A_1)$, $\mathcal{F}_{\underline{m}} = \sigma(W)$, and moreover that ξ is independent of A_1 conditional on W (in Horowitz and Manski (1995) a related notion is called "contaminated sample"). Then, for a prediction $\tilde{Y}$,

$$\tilde{Y} = \mathbf{1}\{\xi \in G\}\,\mathbb{E}\left[Y|\mathcal{F}_{\overline{m}}\right] + \mathbf{1}\{\xi \notin G\}\,\mathbb{E}\left[Y|\mathcal{F}_{\underline{m}}\right],$$

it is straightforward to see that as long as $\mathcal{F}_{\overline{m}}, \mathcal{F}_{\underline{m}} \supseteq \sigma(W)$, our results continue to hold. We conclude that as long as all individuals use at least as much information as the econometrician we are still in a benign scenario.

3.1.3 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 $\mathcal{F}_m$ and the variables to the disposal of the econometrician is given by either of*

1. $\mathcal{F}^1_m = \sigma(X_{-1}, S, A_1)$
2. $\mathcal{F}^2_m = \sigma(X_1, S, A_1)$
3. $\mathcal{F}^3_m = \sigma(B, S, A_1)$, *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 all information in X_1, but discard other variables of interest that are potentially correlated. Finally, the last scenario is when all information contained in X is shrunk, but in a smooth fashion and without, in particular completely discarding the variable of interest.

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 $\mathcal{S} = \left\{(p,q) \in \mathbb{R}^2_+ | 1/p + 1/q = 1,\ \mathbb{E}\left[\left|Y - \tilde{Y}\right|^p\right] < \infty, \mathbb{E}\left[|Q_{x_1}|^q\right] < \infty\right\}$ and $\mathcal{T} = \left\{(p,q) \in \mathbb{R}^2_+ | 1/p + 1/q = 1,\ \mathbb{E}\left[|Y|^p\right] < \infty, \mathbb{E}\left[|Q_{x_1}|^q\right] < \infty\right\}$

[6]Up to now we have assumed that the size of the sigma algebra is the same for every individual. Loosely speaking, this may be translated as saying that people have in principle access to the same sources of information.

Theorem 2. *Let all the variables and functions be as defined above. Let assumptions 1, 2.2, 4.1 and 13 hold. Then follows* $\partial_{x_1} m(W) = 0$ *(a.s.). If 4.1 is replaced by 4.2, then* $\partial_{x_1} m(W) = \mathbb{E}[\partial_{x_1}\phi(X,A)|X_1,S] + Cov\left[Y - \tilde{Y}, Q_{x_1}|X_1,S\right]$*. Moreover, for the local average structural derivative,*

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

and for the average structural derivative,

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

Finally, if 4.2 is replaced by 4.3, then $\partial_{x_1} m(W) = \mathbb{E}[\partial_{x_1}\phi(X,A)|B,S] + Cov\left[U, Q_{x_1}|X_1,S\right]$*, 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 $\mathbb{E}[\partial_{x_1}\phi(X,A)|\mathcal{F}_\alpha]$, $\mathcal{F}_\alpha \subseteq \mathcal{F}_m$ 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 $\mathbb{E}[YQ_{x_1}|X_1,S] = Cov\left[Y - \tilde{Y}, Q_{x_1}|X_1,S\right]$ 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. $Var\left[Y|\mathcal{F}_m\right] = Var\left[U|\mathcal{F}_m\right]$, we may by Cauchy Schwarz give at least an upper bound for the absolute value of the bias, because the term $\mathbb{E}\left[Q_{x_1}^2|X_1,S\right]$ 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.

3.2 Marginal Effects if a Regressor is not Fully Recalled

3.2.1 Setup and Problem

Given that left hand side recall problems caused so many difficulties, it should hardly come as surprise that right hand side recall errors present an even more serious challenge. To look at the details, let us first define the scenario: As a result of incomplete recollection we observe the data $D = (Y, \tilde{X}_1, X_{-1}, S')'$, where $\tilde{X}_1$ is the poorly recalled regressor. In this section we will start with the best scenario where individuals use the squared error loss to predict X_1. Note first that there is no reason why the dependent variable Y should not be part of the individuals' information set. Consequently, the assumptions about the information set are as follows:

Assumption 5. *Let all variables be as defined above. The relationship between the individuals' σ-algebra $\mathcal{F}_m$ and the variables at the disposal of the econometrician is given by either of*

1. $\mathcal{F}_m = \sigma(Y, W_{-1}, J)$.
2. $\mathcal{F}_m = \sigma(\varphi(Y, W_{-1}), J)$ *such that* $\sigma(\varphi(Y, W_{-1}), J) \not\supseteq \sigma(Y, W_{-1})$ *where φ is a nonrandom Borel function.*

For the following discussion, it is imperative to recall that we are interested in average structural marginal effects, given our information set, i.e. $\mathbb{E}[\partial_{x_1}\phi(X, A)|D]$, or $\mathbb{E}[\partial_{x_1}\phi(X, A)|\Phi]$, where Φ is some coarser sigma algebra than $\sigma(D)$. To this end, we start out by analyzing the object that would appear most natural, i.e. the mean regression $\mathbb{E}\left[Y|\tilde{X}_1 = \tilde{x}_1, W_{-1} = w_{-1}\right] = m(\tilde{x}_1, w_{-1})$, and take derivatives w.r.t. the poorly recalled variable $\tilde{X}_1$. Note that for m being a nontrivial function of $\tilde{X}_1$, we require (at least) that $\tilde{X}_1$ is not a function of W_{-1} alone, and that therefore individuals are required to use some additional information in their backcast. Moreover, let $K = X_1 - \tilde{X}_1$ and $\tilde{Q}_{\tilde{x}_1} = \partial_{\tilde{x}_1} \log f_{KA|\tilde{X}_1 W_{-1}}(K, A, \tilde{X}_1, W_{-1})$, where we focus on the (benign!) case of $J = A_1$. Observe that K can be thought of as a measurement error. It is straightforward to show

$$\partial_{\tilde{x}_1} m(\tilde{x}_1, w_{-1}) = \mathbb{E}[\partial_{x_1}\phi(X, A)|\tilde{X}_1 = \tilde{x}_1, W_{-1} = w_{-1}] + Cov[Y, \tilde{Q}_{\tilde{x}_1}|X_1 = \tilde{x}_1, W_{-1} = w_{-1}], \quad (3.3)$$

i.e., the mean regression provides only a biased version of the object of interest (a similar result holds for the quantile). The bias depends on the conditional correlation between Y (or a function thereof) and $\tilde{Q}_{\tilde{x}_1}$. Note that the bias expression $\tilde{Q}_{\tilde{x}_1}$ contains two sources of potential correlation: K (the measurement error) and A (the unobserved heterogeneity), both of which are systematically related to $\tilde{X}_1$. This limits the scope of instruments, which have been suggested as a possible way out of right hand side measurement error problem. As mentioned in extension **1** after theorem **1**, instruments may be integrated in this framework in control function fashion as in Imbens and Newey (2005), or Hoderlein (2006). But, firstly, note that the individuals have to use the information contained in the instruments in the backcast. Moreover, secondly, in this case the control function residuals T would have to set $Cov[Y, P_{\tilde{x}_1}|W = w]$ to zero (where $P_{\tilde{x}_1} = \partial_{\tilde{x}_1} \log f_{KA|\tilde{X}_1 W_{-1} T}(K, A, \tilde{X}_1, W_{-1}, T)$). Because of the complicated joint dependence, we view this as an unlikely event and hence voice our scepticism against the use of instruments in this scenario.

Moreover, note that bounds are hard to provide as $\tilde{Q}_{\tilde{x}_1}$ is very intractable, and it is not clear what reasonable assumptions are in this case. Finally, one can show that the bias converges to zero as the distribution of K tends to a point measure (normalized at zero). So it is not a question of whether individuals use more or less information than the econometrician. It is a question of perfect versus imperfect recall. Thus, we conclude that any direct attack on the problem of estimating average structural derivatives does not seem promising.

3.2.2 Possible Solutions I: Reverse Regression

The scope of the difficulty faced when the regressors of interest are not accurately recalled suggests that it would be advantageous to be back in the scenario when a left hand side variable was not recalled properly. Moreover, the intuitive information reduction structure is lost in the direct attack. Under certain circumstances it is actually possible to switch back to the left hand side scenario and retain a more tractable problem having the previous structure.

Suppose the function ϕ would be strictly monotonic in X_1, the poorly recalled regressor. Then we may rewrite the model defined by assumption **1** as in the following assumption:

Assumption 6. *1. There exists a representation of the relationship of interest of the form*

$$X_1 = \psi(Y, X_{-1}, A),$$

where ψ as a nonrandom, scalar valued Borel-measurable function defined on $\mathbb{R}^K \times \mathcal{A}$, continuously differentiable in y, with derivatives $\partial_y \psi = \partial_{x_1} \phi^{-1}$ that are uniformly bounded from above and below.

2. In addition, $A = (A_1, A_2)$ where A_1 and A_2 are both uniformly distributed random scalars and ψ is strictly monotone in the last two arguments for fixed values of the other arguments.

Recall that Y is by definition correlated with the unobserved heterogeneity in A, and hence will be endogenous. However, we show below that this situation is better than the previous one. In particular, monotonicity in unobservables will be a very helpful assumption, but we will also discuss the implications of not assuming monotonicity. Suppose that we are in case where individuals use the squared loss when backcasting. Let $\tilde{Q}_y = \partial_y \log f_{A|JYW_{-1}}$ and $\eta(x_{1;}y, w_{-1}) = \left[f_{X_1|YW_{-1}}(k_\alpha(y, w_{-1}), y, w_{-1})\right]^{-1} \mathbf{1}[x_1 \leq k_\alpha(y, w_{-1})]$. With the obvious modification in regularity conditions (denoted by double primes), we obtain the following result:

Theorem 3. *(Consequences of Large Information under Squared Error Loss): Let all the variables and functions be as defined above. Let assumptions 2, 5.1, 6.1 and 12" hold. Then we obtain for all w and all α:*

1. $\partial_y k_\alpha(y, w_{-1}) = \mathbb{E}[\partial_{x_1}\phi(X, A)^{-1}|Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})]$
$\quad + Cov\left[\eta(X_{1;}Y, W_{-1}), \tilde{Q}_y|Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})\right]$

If instead of assumption 12", assumption 13" holds, we obtain that for all w:

2. $\partial_y m(y, w_{-1}) = \mathbb{E}[\partial_{x_1}\phi(X, A)^{-1}|Y = y, W_{-1} = w_{-1}] + Cov\left[X_1, \tilde{Q}_y|Y = y, W_{-1} = w_{-1}\right].$

Discussion of Theorem 3: Despite the superficial similarities, note that the situation is much better than in equation (3.3): The bias depends only on the unobserved heterogeneity component A. Moreover, there is a deeper parallel to the left hand side recall error discussed in theorem **1**. Recall that if X_1 were endogenous too, and the relationship between the endogenous variables is of main interest[7], then we obtain for the mean regression

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

Note the striking similarity between this result, and theorem **3.2**. Specifically, note that the first term gives a best approximation given the information both respondent and econometrician use. However, in the lhs poorly recalled case it is the marginal effect of interest, while here it is the inverse of this quantity that is approximated. The second is a bias term, which depends on the correlation between either $\tilde{Q}_y$ (which is the density of A given (J, Y, W_{-1})) and X_1, or on the correlation between Y and $\bar{Q}_{x_1}$. In any of the two cases, if J already captures most of the variation of A, then $\tilde{Q}_y$ (or $\bar{Q}_{x_1}$) might not be very sensitive to variations in Y (or X_1), respectively, and this term will vanish, or at least its correlation with X_1 (or Y) will be small. In this case, $\partial_y m(y, w_{-1})^{-1}$ may provide a reasonable approximation to $\mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]$, provided the relationship of interest is close to being linear.

In any case, this suggest that right hand side recall errors can be decomposed into three components: 1. A left hand side recall error component, which can again be classified according to whether respondents use the information the econometrician employs. In particular, if individuals use very little information, this part exhibits the same problems as discussed at length above. 2. A "standard" endogeneity, i.e., once we invert the relationship of interest Y behaves like an ordinary endogenous regressor. 3. An additional expression arising from the effects of nonlinearity of ϕ in the inversion.

If we abstract from the third problem (as would be justified in the case of approximate linearity), right hand side recall errors behave materially like left hand side recall errors in an equation where the regressors are additionally endogenous (e.g., through simultaneity). We conclude that once we can invert the relationship of interest, *all endogenous variables behave materially similar with respect to recall error, whether they are on the left hand or on the right hand side.* This suggest to select the dependent variable in a regression with endogenous regressors and suspected recall error according to whether it is poorly recalled, and whether we rather find instruments that set modified versions of $\tilde{Q}_y$ or $\bar{Q}_{x_1}$ to zero[8].

Finally, as a caveat, even after the bias due to endogeneity is eliminated, the leading term exhibits differences. Then, $\partial_y k_\alpha(y, w_{-1})^{-1}$ is only the best projection to $\mathbb{E}[\partial_{x_1}\phi(X, A)^{-1}|Y = y, W_{-1} = w_{-1}]^{-1}$, not to $\mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]$. But taking the approximation $\mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}] \cong \mathbb{E}[\partial_{x_1}\phi(X, A)^{-1}|Y = y, W_{-1} = w_{-1}]^{-1}\{1 + \mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]^{-2} Var[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]\}$, which is accurate provided higher order terms in a Taylor expansion are small, we obtain that $sign\,\{\mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]\} = sign\,\{\partial_y k_\alpha(y, w_{-1})^{-1}\}$ and $|\mathbb{E}[\partial_{x_1}\phi(X, A)|Y = y, W_{-1} = w_{-1}]| \geq |\partial_y k_\alpha(y, w_{-1})^{-1}|$.

[7]In Economics, this is perhaps the rule rather than the exception. Specifically, think of A as capturing mainly unobserved preference heterogeneity, as is common in Microeconometrics (e.g., consumer demand), meaning that both X_1 and Y are endogenous because they are related to the same underlying preference ordering.

[8]The latter sentence is meant in the sense that control function residuals be included in W^* as in extension 1 to theorem **1.**

Extensions: 1. Eliciting the Entire Distribution: The structure of the bias term suggests there is some hope of reducing it further by eliciting the entire random variable X_1. A (strong) set of assumptions that helps to achieve this involves additionally monotonicity, cf. assumption **6.2**. Suppose further that we elicit again the entire conditional distribution via their quantiles. Let $\partial_y k_\alpha^\gamma(y, w_{-1})$ be the conditional α quantile of the backcast of the conditional γ quantile of X_1, denoted $X_1^\gamma = k^\gamma(Y, X_{-1}, A_1)$. By monotonicity of ψ, $X_1^\gamma = \psi(Y, X_{-1}, A_1, \gamma)$. Hence

$$\partial_y k_\alpha^\gamma(y, w_{-1}) = \mathbb{E}[\partial_y \psi(Y, X_{-1}, A_1, \gamma)|Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})] = \partial_y \psi(y, x_{-1}, \alpha, \gamma),$$

for $\alpha, \gamma \in (0,1)$, and consequently $\partial_y k_\alpha^\gamma(y, w_{-1})^{-1} = \partial_{x_1} \phi(x_1(y), x_{-1}, \alpha, \gamma)$. If assumption 6.2 does not hold, we obtain by the same arguments as in theorem **3** only that

$$\begin{aligned} \partial_y k_\alpha^\gamma(y, w_{-1}) &= \mathbb{E}[\partial_{x_1} \phi(X, A)^{-1}|Y = y, W_{-1} = w_{-1}, X_1^\gamma = k_\alpha^\gamma(y, w_{-1})] \\ &\quad + Cov\left[X_1, \tilde{Q}_y|Y = y, W_{-1} = w_{-1}, X_1^\gamma = k_\alpha(y, w_{-1})\right], \end{aligned}$$

but the second term may be expected to be small: Conditioning on (A_1, A_2), the covariance is only over the remaining heterogeneity in A. Consequently, $\partial_y k_\alpha^\gamma(y, w_{-1})^{-1}$ may again provide a reasonable approximation to the local average derivative of interest. Note that there are two routes out of the endogeneity problem that causes the bias: instruments, potentially combined with bounds, and eliciting quantiles. Eliciting quantiles may hence be seen as novel route out of this problem, akin to the treatment of endogeneity in nonseparable systems of equations.

2. Bounds when Individuals use less Information than the Econometrician: The clear parallels to the scenario with left hand side recall errors and endogenous regressors suggest that in order to treat the case when individuals use less information than the econometrician as specified in assumption **5.2**, we may come up with bounds that are analogous to those derived in section 3.1.3. Indeed, much of the discussion would simply repeat what we have summarized in theorem **2,** and we do not present it here. At this point we simply emphasize that the bounds may be obtained in practise by using aggregate information, or measures of individuals' uncertainty.

3.2.3 Possible Solutions II: Eliciting Co-Distributions

If strict monotonicity of ϕ in the regressor of interest is ex ante considered to be a problematic assumption then the reverse regression may not be a good idea, and we have to search for alternatives. In this and the following subsection we suggest two alternatives that have (again to the best of our knowledge) not been suggested in the literature before.

We call the first co-distributions. It is very closely related to the discussion of the previous subsection. The idea is the following: instead of eliciting a point backcast, we ask people to report their dependent variable at different quantiles of the regressor of interest. Say, we were interested in the effect of income on food consumption. After explaining the concept of a quantile in a survey, a respondent could be asked: "All other factors held constant, suppose your income were at the 25th percentile of its usual range. How much would you spend on food?"

This would have a direct and two indirect positive effects. The direct one may be formalized as follows. Suppose the model is again given by

$$Y = \phi(X_1, X_{-1}, A) = \phi(\tau(W_{-1}, B), X_{-1}, A),$$

where all variables are as defined above, and τ is strictly monotonic in the scalar B, where $B|W_{-1} \sim \mathcal{U}[0, 1]$. This representation of X_1 exists under fairly unrestrictive conditions, it is a reduced form much like the quantile. Then $Y^\gamma = \phi(\tau(W_{-1}, \gamma), X_{-1}, A)$, with $\gamma = 0.25$ is the random variable we are eliciting by the above question. Consequently, we could identify the effect of X_1 on Y by looking at variations in B and recall production factors S (which are a part of W_{-1}). There are two related caveats: 1. Whether individuals truly provide us with B, or whether we obtain $\phi(\tau(W_{-1}, B_1, B_2), X_{-1}, A)$, where the above assumption only applies to B_1 and there is still a rest of uncertainty B_2 left (e.g., individuals do not fully account for all variations in their income) is not clear. 2. The result depends on τ being the same function for every individual.

Possible positive indirect effects are that variables are highlighted to the individual, and hence more effort is put in memorizing (in our language, m is bigger). Moreover, we get additional variation in the data. This may be particularly useful when we have regressors with insufficient variation like prices in consumer demand.

3.2.4 Possible Solutions III: Eliciting Marginal Effects

Another way out of the problem we are facing is eliciting directly the quantity of interest. So, instead of asking people for their demand and their income separately and then estimate the derivative, we may ask them directly for a prediction of their marginal effect. For instance, if individuals use the squared error loss, this provides us directly with $\mathbb{E}[\partial_{x_1}\phi(X, A)|\mathcal{F}_m]$, and all tools of the theory of the second section may be applied to $Z = \partial_{x_1}\phi$.

One obvious advantage is that we may, by simply aggregating, obtain the average structural derivative $\mathbb{E}[\partial_{x_1}\phi(X, A)]$. This holds regardless of how good individuals recall the regressor (i.e., we have a right hand side measurement error), and even holds if individuals vary arbitrarily in the coarseness of their sigma algebra (i.e., we have differential measurement error), provided all individuals use the squared error loss to determine their prediction. An obvious disadvantage is that we may not identify local average structural derivatives (i.e., the average derivative of a subpopulation defined by a certain combination of values of regressors), if (some) individuals reduce the information necessary to define the subpopulation, be it when recalling the dependent variables or the regressor.

3.3 Effect on Economic Restrictions

Thus far we have only discussed the effects of survey design on the derivatives. In this subsection we briefly discuss the implications for a structural model in the case of consumer demand. Note first that homogeneity of degree zero in income and prices (i.e., $g(p/y, y) = g(p/y)$) is easily tested because it amounts to testing whether the last derivative is zero. Indeed, we will pursue this issue in the application below. For more complicated objects like the Slutsky matrix we conclude that even in the most benign of cases, as in Theorem **1,** we are only able to identify average effects. But averages of nonlinear structures pose additional difficulties compared to derivatives. A thorough analysis is beyond the scope of the paper and we pursue it elsewhere. Here we only refer to Hoderlein (2006), where similar complications are discussed when the only source of uncertainty is unobserved preference heterogeneity.

4 Empirical Research 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 determine whether individuals use a boundedly rational model of memorizing? 3. Which loss function do respondents use. In particular, do respondents use the squared error loss? 4. Can we find the quantities necessary to construct the bounds, and how do we incorporate the bounds when evaluating economic hypotheses?

Evaluating these questions will be the topic of this section. First, we will consider testable hypotheses of the formal model of recollection, then we discuss testing for squared error loss, show how to test for information reduction, and finally we establish how to test the hypothesis of a zero marginal effect.

4.1 Testing the Survey Design Model via Implications of Cost Minimization on Factor Demands

To describe the implications of a formal model of survey design we have to distinguish between the two elements Π_0 and c. As mentioned, the latter can be seen as the optimizer of the cost minimization problem of building up memory m. This implies that the factor demands $z = \varphi(m, q, \lambda)$ obtain some structure. In particular, factor demands should be zero homogenous in prices. Moreover, the matrix of price derivatives for fixed λ and m is negative semidefinite and symmetric. In difference to consumer demand there is no wealth effect to be taken into account.

The hypothesis of homogeneity is part of a set of hypotheses of the form

$$H_0 \colon \quad \mathbb{P}\left\{\mu(W_1, W_2) = \eta(W_1)\right\} = 1,$$

where $\mu : \mathbb{R}^{d_{W_1}+d_{W_2}} \to \mathbb{R}^{d_Y}$ and $\eta : \mathbb{R}^{d_{W_1}} \to \mathbb{R}^{d_Y}$ are smooth functions of their respective arguments, usually mean regressions and $W = (W_1', W_2')'$ denotes a partition, not necessarily identical to $\left(X_1, W_{-1}'\right)$. The alternative is that both regressions differ on a subset of their support of positive measure. Examples are the zero homogeneity of factor demands, and the hypothesis $\mathbb{P}\{\partial_{s_1} m(W) = 0\} = 1$, which is an important part of theorem **1**. The null is equivalent to the condition that the L_2 distance of the two functions is zero. Using a nonzero and bounded weighting function a this condition can be written as

$$\Gamma_0 = \mathbb{E}\Big(\sum_{j=1}^{d_Y}\big(\mu^j(W_1, W_2) - \eta^j(W_1)\big)^2 a(W_1, W_2)\Big) = 0. \tag{4.1}$$

We may use sample counterparts to check whether Γ_0 is significantly different from 0. To obtain sample counterparts, one simply has to replace all functions by nonparametric estimates, as we will discuss below.

4.2 Testing the Specification of the Loss Function

To characterize testable implications associated with our modeling of Π_0, we have to specify the setup further provided we are given a correctly measured Y alongside with the individuals recollection $\tilde{Y}$. If we invoke the assumption of squared error loss, because the conditional expectation $\mathbb{E}[Y|\mathcal{F}_m]$ is specified as the function of the data individuals form as backcast "unbiasedness" is a consequence, i.e., $\mathbb{E}[\mathbb{E}[Y|\mathcal{F}_m]] = \mathbb{E}[Y]$, even in the case when individuals use less information than the econometrician, which amounts to comparing two different means. But we can make even stronger statements if individuals use more information, in particular $\mathbb{E}[Y|W] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_m]|W] = \mathbb{E}\left[\tilde{Y}|W\right]$ with probability one. If we let $m(w)$ and $M(w)$ denote continuous versions of $\mathbb{E}\left[\tilde{Y}|W = w\right]$, respectively $\mathbb{E}[Y|W = w]$, we base the test statistic on the null hypothesis

$$H_0\colon \quad \mathbb{P}(M(W) = m(W)) = 1,$$

while the alternative is that they differ on a subset of the support of W of positive measure. The null is equivalent to the condition that the L_2 distance of the two functions is zero. Using a nonzero and bounded weighting function a this condition can be written as

$$\Gamma_2 = \mathbb{E}\Big(\sum_{j=1}^{d}\big(M^j(W) - m^j(W)\big)^2 a(W)\Big) = 0, \tag{4.2}$$

where the superscript j denotes the j-th component function.

Suppose instead that the individuals use the absolute error loss to determine the backcast. Then, the conditional means do not have to coincide any longer.

4.3 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 assumption that A_1 is independent of X given S, this model is within reach of the framework discussed in Hoderlein and Mammen (2006). Although we could devise a test for dimension reduction more generally, for the purpose of this paper it suffices to assume that dimension reduction takes the form of the index, ie $\mathcal{F}_m = \sigma(B, S, A_1)$, where $B = X'\beta$. Let $\tilde{Y}_k$ denote the k-th component. Then, we may estimate the parameter β (up to scale) in the model $\tilde{Y}_k = \psi(X'\beta, S, A_1)$ by use of the equations

$$\partial_x k_\alpha(w) = \beta \mathbb{E}[\partial_l \psi_k(X'\beta, S, A_1)|W = w, \tilde{Y}_k = k_\alpha(w)],\ k = 1, .., d,\ \ (a.s), \tag{4.3}$$

where ∂_l denotes the derivative with respect to the index. Imposing a normalization condition, we obtain $\mathbb{E}\left[\int \partial_x k_\alpha(W) w(\alpha, W) d\alpha\right] = \beta$ for some weight function w. This identifies β. For testing the hypothesis $\varphi_k(X, A_1) = \psi_k(X'\beta, A_1),\ \ k = 1, .., d,\ \ (a.s)$, we use that from equation (4.3) follows that $\partial_{x_l} k_\alpha(w)/\partial_{x_j} k_\alpha(w) = \beta^l/\beta^j,\ \ \ k = 1, .., d,\ \ l, j = 1, .., p$. This suggests to form

$$\Gamma_3 = \mathbb{E}\left[\sum_k \sum_{j>l} \int \left[\partial_{x_l} k_\alpha(W)\beta^j - \partial_{x_j} k_\alpha(W)\beta^l\right]^2 \gamma(W, \alpha) d\alpha\right], \tag{4.4}$$

for some bounded nonzero weighting function γ.

4.4 Testing for Zero Marginal Effect in the Presence of Information Reduction

If we detect information reduction, as outlined in theorem **2**, we might be able to provide bounds on the marginal effects. In this section we will consider as examples those introduced in the introduction, i.e.,

$$\left(\mathbb{E}\left[\partial_{x_1} m(X)\right] - \mathbb{E}\left[\partial_{x_1} \phi(X, A)\right]\right)^2 \leq \mathbb{E}\left[Y^2\right] \mathbb{E}\left[Q_{x_1}^2\right],$$

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 $\beta_1 = \mathbb{E}\left[\partial_{x_1} \phi(X, A)\right]$, and assume for the time that $\mathbb{E}\left[Y^2\right]$ 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 $\mathbb{E}\left[\partial_{x_1} m(X)\right]^2 - \mathbb{E}\left[Y^2\right] \mathbb{E}\left[Q_{x_1}^2\right] \leq 0$. This suggests to use

$$\Gamma_4 = \mathbb{E}\left[\partial_{x_1} m(X)\right]^2 - \mathbb{E}\left[Y^2\right] \mathbb{E}\left[Q_{x_1}^2\right],$$

and test $H_0 : \Gamma_4 \leq 0$ versus the (one sided) alternative $H_1 : \Gamma_4 > 0$. Note that we may rewrite $\Gamma_4 = g(\mathbb{E}\left[\partial_{x_1} m(X)\right], \mathbb{E}\left[Y^2\right], \mathbb{E}\left[Q_{x_1}^2\right])$, where $g : \mathbb{R}^3 \rightarrow \mathbb{R}$, $g(x, y, z) = x^2 - yz$ is a smooth function of the sample averages. Assume $\mathbb{E}\left[\partial_{x_1} m(X)\right] \neq 0$, and $\mathbb{E}\left[Y^2\right] > 0, \mathbb{E}\left[Q_{x_1}^2\right] > 0$ which rules out the degenerate case when g is not being differentiable at $(\mathbb{E}\left[\partial_{x_1} m(X)\right], \mathbb{E}\left[Y^2\right], \mathbb{E}\left[Q_{x_1}^2\right])$, see Datta (1995) for a discussion. While assuming $\mathbb{E}\left[Y^2\right] > 0$ and $\mathbb{E}\left[Q_{x_1}^2\right] > 0$ is fairly innocuous, $\mathbb{E}\left[\partial_{x_1} m(X)\right] \neq 0$ is less obvious. But note that homogeneity of degree zero implies (only) that $\partial_{x_1} m(X) + \mathbb{E}\left[Y Q_{x_1}|X\right] = 0$. Since $\mathbb{E}\left[Y Q_{x_1}|X\right] \neq 0$ in general under the conditions of theorem

2 this implies that $\partial_{x_1} m(X) \neq 0$ and hence $\mathbb{E}\left[\partial_{x_1} m(X)\right] \neq 0$ generically. This implies actually that for the sample counterpart statistic

$$\widehat{\Gamma}_4 = \left[n^{-1} \sum_i \partial_{x_1} \widehat{m}(X_i)\right]^2 - \left[n^{-1} \sum_i Y_i^2\right] \left[n^{-1} \sum_i Q^2_{x_1,i}\right],$$

it holds that $\sqrt{n}\left(\widehat{\Gamma}_4 - \Gamma_4\right)$ is asymptotically normally distributed. To see this, apply a Taylor expansion to g:

$$\begin{aligned} \widehat{\Gamma}_4 - \Gamma_4 &= 2\mathbb{E}\left[\partial_{x_1} m(X)\right] \left\{ n^{-1} \sum_i \partial_{x_1} \widehat{m}(X_i) - \mathbb{E}\left[\partial_{x_1} m(X)\right] \right\} \\ &+ \mathbb{E}\left[Q^2_{x_1}\right] \left\{ n^{-1} \sum_i Y_i^2 - \mathbb{E}\left[Y^2\right] \right\} + \mathbb{E}\left[Y^2\right] \left\{ n^{-1} \sum_i Q^2_{x_1 i} - \mathbb{E}\left[Q^2_{x_1 i}\right] \right\} + o_p\left(n^{-1/2}\right). \end{aligned}$$

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 $n^{-1} \sum_i \partial_{x_1} \widehat{m}(X_i)$ by an U-statistic CLT, we require higher order smoothness assumptions on m and have to employ a higher order Kernel. However, if this is assumed, by a theorem of Mammen (1992) the bootstrap is consistent.

5 Nonparametric Tests in this Framework

5.1 Overview

In this subsection we discuss the bootstrap versions of the test statistics we will implement. Given the large number of different tests our model suggests, discussing all thoroughly would be beyond the scope of this paper. Even though they are not treatable under one consistent format, they share structural similarities. Indeed, there would be ample repetition if we were to discuss all of them. Hence we will proceed as follows: we will discuss the asymptotic behavior of the first test in detail, and show formally why the bootstrap works. For the rest we then provide the bootstrap procedures and give heuristic explanations why the work.

5.2 Homogeneity: From Hypothesis to Sample Counterpart

Instead of using Γ_0, for bias reduction purposes we propose:

$$\Gamma_1 = \mathbb{E}\Big(\sum_{j=1}^{d} \big(\mu^j(W_1, W_2) - \mathbb{E}\left[m^j(W_1)|W_1, W_2\right]\big)^2 a(W_1, W_2)\Big).$$

The sample counterpart of Γ_1 serves as test statistic. Given a sample of n independent and identically distributed random vectors $(Y_1, W_{11}, W_{21}), \dots, (Y_n, W_{1n}, W_{2n})$, we replace the unknown functions $m(w_1)$ and $\mu(w_1, w_2)$ by their Nadaraya-Watson estimators $\hat{m}_{\tilde{h}}(w_1)$ and $\hat{\mu}_h(w_1, w_2)$.

Formally, these are defined as vectors with the one-dimensional estimators, $\hat{m}^j_{\tilde{h}}(w_1) = \sum_{i=1}^n K_{\tilde{h}}(w_1 - W_{1i})Y_i^j / \sum_{i=1}^n K_{\tilde{h}}(w_1 - W_{1i})$ and $\hat{\mu}^j_h(w_1, w_2) = \sum_{i=1}^n K_h(w_1 - W_{1i}, w_2 - W_{2i})Y_i^j / \sum_{i=1}^n K_h(w_1 - W_{1i}, w_2 - W_{2i})$, where $K_h(u) = K(u/h)/h$ with a kernel K and bandwidths h and $\tilde{h}$. as individual elements. As an estimator for $\mathbb{E}(m^j(W_1) \mid W_1 = w_1, W_2 = w_2)$ we propose

$$\widehat{\mathcal{K}_n m^j_{\tilde{h}}}(w_1, w_2) = \frac{\sum_{i=1}^n K_h(w_1 - W_{1i}, w_2 - W_{2i})\hat{m}^j_{\tilde{h}}(W_{1i})}{\sum_{i=1}^n K_h(w_1 - W_{1i}, w_2 - W_{2i})}.$$

Then, the statistic is given by

$$\hat{\Gamma}_{\mathcal{K}} = \frac{1}{n}\sum_{j=1}^{d_Y}\sum_{i=1}^{n}\left(\hat{\mu}^j_h(W_{1i}, W_{2i}) - \widehat{\mathcal{K}_n m^j_{\tilde{h}}}(W_{1i}, W_{2i})\right)^2 A_i \tag{5.1}$$

with $A_i = a(W_{1i}, W_{2i})$. The additional smoothing step associated with $\widehat{\mathcal{K}_n m_{\tilde{h}}}(w_1, w_2)$ eliminates the bias arising from $\hat{\mu}_h(w_1, w_2)$. This test is a nonparametric omission of variables test related to work of – inter alia -Aït-Sahalia, Bickel and Stoker (2002). Compared to this line of work, there are three modifications: first we have an additional smoothing step, second we consider a system of equations, and third we focus on the bootstrap below.

5.3 Asymptotic Distribution of the Test Statistic

In order to treat the asymptotic distribution of the test statistic, we introduce the following assumptions. The first two assumptions are concerned with the data generating process.

Assumption 7. *The data $(Y_i, W_{1i}, W_{2i}), i = 1, \ldots, n$ are independent and identically distributed with density $f(y, w_1, w_2)$.*

Assumption 8. *For the data generating process*

1. *The continuously differentiable weighting function $a(w_1, w_2)$ is nonzero and bounded with compact support $\mathcal{A} \subset \mathbb{R}^{d_{W_1}+d_{W_2}}$.*

2. *$f_{YW}(y, w_1, w_2)$ is r-times continuously differentiable ($r \geq 2$). f_{YW} and its partial derivatives are bounded and square-integrable on $\mathcal{A}$.*

3. *$\mu(w_1, w_2)$ and $m(w_1)$ are $r + 1$-times continuously differentiable.*

4. *$f_W(w_1, w_2) = \int f_{YW}(y, w_1, w_2)\, dy$ is bounded from below on $\mathcal{A}$, i. e. $\inf_{(w_1,w_2)\in\mathcal{A}} f_{YW}(w_1, w_2) = b > 0$.*

5. *The covariance matrix*

   $$\Sigma(w_1, w_2) = (\sigma^{ij}(w_1, w_2))_{1\leq i,j\leq d_Y} = \mathbb{E}((Y - \mu(W_1, W_2))(Y - \mu(W_1, W_2))' \mid W_1 = w_1, W_2 = w_2)$$

   is square-integrable (elementwise) on $\mathcal{A}$ and.

6. $\mathbb{E}((Y^j - \mu^j(W_1, W_2))^2(Y^k - \mu^k(W_1, W_2))^2) < \infty$ *for every* $1 \leq j, k \leq d_Y$*.*

Assumption **7** may be relaxed to allow for dependent data. Assumption **8** contains standard differentiability and integrability assumptions that do not deserve further mentioning. The next assumptions are concerned with the kernel and the bandwidth sequences. For simplicity, we assume product kernels in both regressions. Therefore we formulate our assumptions for one-dimensional kernel functions. To simplify things further, instead of bandwidth vectors $\mathbf{h} \in \mathbb{R}^{d_{W_1}+d_{W_2}}$ and $\tilde{\mathbf{h}} \in \mathbb{R}^{d_{W_1}}$ we assume that we have only one single bandwidth for each regression $(h, \tilde{h})$. We shall make use of the following notation: Define kernel constants

$$\kappa_k = \int u^k K(u)\,\mathrm{d}u \qquad \text{and} \qquad \kappa_k^2 = \int u^k K(u)^2\,\mathrm{d}u$$

$$\kappa_* = \int \Big(\int K(u)K(u-v)\,\mathrm{d}u\Big)^2\,\mathrm{d}v$$

Then, our assumptions regarding kernels and bandwidths are as follows:

Assumption 9. *The one-dimensional kernel is Lipschitz continuous, bounded, has compact support, is symmetric around 0 and of order* r *(i. e.* $\int u^k K(u)\,du = 0$ *for all* $k < r$ *and* $\int u^r K(u)\,du < \infty$*).*

Assumption 10. *For the bandwidths*

1. For $n \to \infty$*, the bandwidth sequence* $h = O(n^{-1/\delta})$ *satisfies*

$$d_{W_1} + d_{W_2} < \delta \tag{5.2}$$

2. For $n \to \infty$*, the bandwidth sequence* $\tilde{h} = O(n^{-1/\tilde{\delta}})$ *satisfies*

$$2\delta \frac{d_{W_1}}{d_{W_1} + d_{W_2}} < \tilde{\delta} \tag{5.3}$$

3. For the order r *of the kernel holds*

$$\tilde{\delta}\frac{2\delta - d_{W_1} - d_{W_2}}{4\delta} < r \tag{5.4}$$

Observe that the optimal rate for the bandwidth in the estimation of the full dimensional regression function $\mu(w_1, w_2)$, given by $\delta_{opt} = (d_{W_1} + d_{W_2}) + 2r$ is not excluded from inequality (5.2), but due to the restriction $\delta < \tilde{\delta}$, $\tilde{h}$ of the dimension-reduced regression m may not be chosen optimally.

To discuss the first theorem, we introduce the following quantities

$$\sigma_\Gamma^{ij} = \iint \sigma^{ij}(w_1, w_2)^2 a(w_1, w_2)^2\,\mathrm{d}w_1\,\mathrm{d}w_2 \qquad b_\Gamma^i = \iint \sigma^{ii}(w_1, w_2) a(w_1, w_2)\,\mathrm{d}w_1\,\mathrm{d}w_2.$$

The asymptotic distribution of the test statistic is given by the following

Theorem 4. *Let assumptions 7–10 hold. Then we have that under* H_0

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_{W_1}+d_{W_2})/2}\hat{\Gamma}_{\mathcal{K}} - h^{-(d_{W_1}+d_{W_2})/2}B_{\mathcal{K}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where

$$\Sigma_{\mathcal{K}}^2 = 2(\kappa_*)^{d_{W_1}+d_{W_2}}\Big(\sum_{i=1}^{d_Y}\sigma_{\Gamma}^{ii} + 2\sum_{i<j}\sigma_{\Gamma}^{ij}\Big) \qquad B_{\mathcal{K}} = (\kappa_0^2)^{d_{W_1}+d_{W_2}}\sum_{i=1}^{d_Y} b_{\Gamma}^i.$$

Simplifying the proofs in the appendix to one line, the test statistic can be written as

$$\hat{\Gamma}_{\mathcal{K}} = \Gamma_n + I_n + U_n \tag{5.5}$$

where $\Gamma_n = 0$ under H_0, U_n depends upon the uniform rate of convergence of the restricted estimator, and I_n is a degenerated U-statistic which dominates asymptotically. This U-statistic converges at the rate $nh^{(d_{W_1}+d_{W_2})/2}$, which is faster than $n^{1/2}$, under the admissible bandwidth sequence. The local power properties of the tests are discussed in Haag and Hoderlein (2006).

5.4 Bootstrap-Implementation of the Test for Homogeneity

The direct way to implement the test is to estimate the expected value $B_{\mathcal{K}}$ and the variance $\Sigma_{\mathcal{K}}^2$. This requires the estimation of integrals like

$$\int \sigma^{jj'}(w_1,w_2)^k a(w_1,w_2)^k \, \mathrm{d}w_1 \, \mathrm{d}w_2 \qquad k=1,2,\ j,j'=1,\ldots,d_Y, \tag{5.6}$$

whose estimation may be cumbersome. To avoid this problem we propose the following wild bootstrap procedure:

1. Calculate (multivariate) residuals $\hat{\varepsilon}_i = Y_i - \hat{m}_{\tilde{h}}(W_{1i})$.
2. For each i randomly draw $\varepsilon_i^* = (\varepsilon_i^{1*},\ldots,\varepsilon_i^{d_Y,*})'$ from a distribution $\hat{F}_i$ that mimics the first three moments of $\hat{\varepsilon}_i$.
3. Generate the bootstrap sample $(Y_i^*, W_{1i}^*, W_{2i}^*), i=1,\ldots,n$ by $Y_i^* = \hat{m}_{\tilde{h}}(W_{1i}) + \varepsilon_i^*$ and $W_{1i}^* = W_{1i}, W_{2i}^* = W_{2i}$.
4. Calculate $\hat{\Gamma}_{\mathcal{K}}^*$ from the bootstrap sample $(Y_i^*, W_{1i}^*, W_{2i}^*), i=1,\ldots,n$.
5. Repeat steps 2 to 4 often enough to obtain critical values for $\hat{\Gamma}_{\mathcal{K}}$.

Assumption 11. *For the bootstrap distribution*
The bootstrap residuals $\varepsilon_i^*, i=1,\ldots,n$ *are drawn independently from distributions* $\hat{F}_i$*, such that* $\mathbb{E}_{\hat{F}_i}\varepsilon_i^* = 0, \mathbb{E}_{\hat{F}_i}\varepsilon_i^*(\varepsilon_i^*)' = \hat{\varepsilon}_i\hat{\varepsilon}_i'$ *and that Cramer's conditions are fulfilled marginally, i. e. there exists a constant* $c > 0$ *such that* $\mathbb{E}_{\hat{F}_i}|\varepsilon_i^{j*}|^p \le c^{p-2}p!\mathbb{E}_{\hat{F}_i}(\varepsilon_i^{j*})^2 < \infty$ *for all* $p = 3,4,\ldots$ *and for all* $j = 1,\ldots,d_Y$.

This set of admissible distributions is very general and includes two-point distribution, discrete distributions, distributions with compact support and the normal distribution, which are the most commonly used distributions in practice.

The theoretical result concerning this bootstrap procedure is given in

Theorem 5. *Let assumptions 7–11 be true. Under* H_0,

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_{W_1}+d_{W_2})/2}\hat{\Gamma}_{\mathcal{K}}^{*} - h^{-(d_{W_1}+d_{W_2})/2}B_{\mathcal{K}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

conditional on the data $(Y_1, W_{11}, W_{21}), \ldots, (Y_n, W_{1n}, W_{2n})$.

To prove theorem **5** it is sufficient to assume that the bootstrap distribution $\hat{F}_i$ mimics the first two moments of $\hat{\varepsilon}_i$.

5.5 Bootstrap-Implementation of the Specification Tests of Loss Function, of Index Structure, and of Zero Marginal Effect in the Presence of Information Reduction

First we start with the test of Loss function defined by (4.2). Applying a similar logic as above,

1. Calculate (multivariate) residuals $\hat{\varepsilon}_i = Y_i - \hat{M}_{\tilde{h}}(W_i)$ and $\hat{\eta}_i = \tilde{Y}_i - \hat{m}_{\tilde{h}}(W_i)$.
2. For each i randomly draw $\varepsilon_i^* = (\varepsilon_i^{1*}, \ldots, \varepsilon_i^{d_Y *})'$ and $\eta_i^* = (\eta_i^{1*}, \ldots, \eta_i^{d_{\tilde{Y}} *})'$ from a distribution $\hat{F}_i$ that mimics the first three moments of $\hat{\varepsilon}_i$, respectively $\hat{\eta}_i$.
3. Generate the bootstrap samples (Y_i^*, W_i^*) and $(\tilde{Y}_i^*, W_i^*), i = 1, \ldots, n$ by $Y_i^* = 0.5\left[\hat{M}_{\tilde{h}}(W_i) + \hat{m}_{\tilde{h}}(W_i)\right] + \varepsilon_i^*$, $\tilde{Y}_i^* = 0.5\left[\hat{M}_{\tilde{h}}(W_i) + \hat{m}_{\tilde{h}}(W_i)\right] + \eta_i^*$ and $W_i^* = W_i$.
4. Calculate $\hat{\Gamma}_2^*$ from the bootstrap samples (Y_i^*, W_i^*) and $(\tilde{Y}_i^*, W_i^*), i = 1, \ldots, n$.
5. Repeat steps 2 to 4 often enough to obtain critical values for $\hat{\Gamma}_2$.

This procedure is quite similar to the previous one, and the analysis would follow essentially along very similar lines. The main difference is perhaps the appearance of two sets of observations and two error terms which require additional assumptions. Finally, to obtain the bootstrap for the index restriction,

1. Calculate (multivariate) single indices $X_i'\hat{\beta}$, using the estimated coefficients $\hat{\beta}$ from equation (4.3).
2. Randomly draw α_i^* from $U[0,1]$.
3. Generate the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \ldots, n$ 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 $\hat{\Gamma}_3^*$ from the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \ldots, n$.

5. Repeat steps 2 to 4 often enough to obtain critical values for $\hat{\Gamma}_3$.

What makes this test substantially different from the previous ones is the use of the conditional quantile. The large sample theory of such tests can nevertheless be tackled along similar lines, using a Bahadur expansion as in Hoderlein and Mammen (2006).

Finally, as suggested in section 4.4, we propose to use the bootstrap to obtain critical values of Γ_4. Specifically, an iid bootstrap based estimator of the distribution function of $\sqrt{n}\left(\hat{\Gamma}_4^* - \widehat{\Gamma}_4\right)$ yields a consistent estimate of the distribution of $\sqrt{n}\left(\widehat{\Gamma}_4 - \Gamma_4\right)$. To test whether $\Gamma_4 < 0$, we propose to use a hybrid bootstrap based on the one sided confidence interval using $\underline{\Gamma_4}$, where $\underline{\Gamma_4}$ 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 $\sqrt{n}\left(\hat{\Gamma}_4^* - \widehat{\Gamma}_4\right)$ we calculate $\underline{\Gamma_4} = \widehat{\Gamma}_4 - 1/\sqrt{n}\gamma_\tau^* = 2\widehat{\Gamma}_4 - k_{.95}\left(\hat{\Gamma}_4^*\right)$, where $k_{.95}\left(\hat{\Gamma}_4^*\right)$ denotes the .95 percentile of $\hat{\Gamma}_4^*$. To obtain the distribution of $\sqrt{n}\left(\hat{\Gamma}_4^* - \widehat{\Gamma}_4\right)$, we run the following bootstrap procedure:

1. Generate the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \ldots, n$ by i.i.d. resampling from the data.

2. Calculate $\sqrt{n}\left(\hat{\Gamma}_4^* - \widehat{\Gamma}_4\right)$ from the bootstrap sample $(\tilde{Y}_i^*, X_i^*, S_i^*), i = 1, \ldots, n$.

3. Repeat steps 1-2 often enough to obtain percentiles of $F_{\sqrt{n}\left(\hat{\Gamma}_4^* - \widehat{\Gamma}_4\right)}$.

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 $\mathbb{E}\left[\tilde{Y}^2\right] \leq \mathbb{E}\left[Y^2\right]$. Hence, with probability approach one, $n^{-1}\sum_i \tilde{Y}_i^2 \leq n^{-1}\sum_i Y_i^2$. Let $\hat{\Gamma}_5 = n^{-2}\left[\left\{\sum_i \partial_{x_1} m(X_i)\right\}^2 - \sum_i \tilde{Y}_i^2 \sum_i Q_{x_1 i}^2\right]$. Then, $\mathbb{P}\left[\widehat{\Gamma}_4 \leq c\right] \geq \mathbb{P}\left[\widehat{\Gamma}_5 \leq c\right]$ 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 $\hat{\Gamma}_5^* = n^{-2}\left[\left\{\sum_i \partial_{x_1}\widehat{m}(X_i^*)\right\}^2 - \sum_i \tilde{Y}_i^{*2} \sum_i Q_{x_1 i}^{*2}\right]$, and obtain that it is below zero then we conclude that $\Gamma_4 \leq 0$ is not rejected, because a test based on the bootstrap distribution of $\hat{\Gamma}_4^*$ would squarely not reject.

6 Recall Effects in Real World Data: Empirical Evidence from the HRS and a Survey Experiment

In this section, we try to answer the following question: "What can be learned about our model from a real world data set?" Moreover, we are also concerned with the impact of recall effects on the evaluation of a typical economic property like zero homogeneity of food consumption. Response behavior in questions on consumption items has a long 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.

For most of the empirical analysis in this section, we use data from the Health and Retirement Study (HRS). Certain specific features of the HRS data collection process make these data particularly amenable to our analysis, for example the fact that the same quantity (food expenditure) has been elicited in various different ways. Nevertheless, it should go without saying that the HRS survey has not been exactly been designed for our purpose. Hence, it only serves as an example of how the results derived in this paper can be applied to real-world data.

This section is organized as follows: We start by describing the data and why it is useful for our purposes. Then, we try to assess the degree of recall error in the data. After that, we look at zero homogeneity which can be phrased as a derivative being zero. Finally, we give a striking example of the effect of the survey design on recall using data from a survey experiment.

Data Description: The data come from the Health and Retirement Study (HRS) and a supplemental survey to the HRS, 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 but additional cohorts were added later (see Juster and Suzman, 1995). In 2000 the HRS interviewed about 20,000 subjects in 13,100 households. While the HRS main surveys are conducted as computer-assisted telephone or personal interviews, the CAMS supplements to the HRS are self-administered mail questionnaires. For the first wave of CAMS, a random sample of 5,000 HRS households was contacted with a response rate of 77.3 percent, see Hurd and Rohwedder (2005) for more details.

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 the internet, but some descriptive statistics can be found in tables A1-A3 in the Appendix.

Income Concept: 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. However, this is only possible in the CAMS part of the data, as this variable is not elicited in the HRS main survey. Hence, if total expenditure is not available, we use RAND's measure of total household income, defined as the sum of all income in the household (labor income, social security and other pension income, asset income, and other income). In what follows when we refer to income, we mean the latter concept, while when we talk about total expenditure (totexp) we mean the former. Total expenditure is believed to suffer frequently from endogeneity, and income as we defined it is used as instruments. As shown formally in Hoderlein (2006), in this case we may obtain the average structural derivative (in our case the total expenditure semielasticity) by dividing the effect we obtain by the derivative of a regression of total expenditure on income. For the purpose of this section - illustrating the effect of left hand side measurement error - the difference between these two concepts does actually not matter too much.

Evidence on Recall Measurement Error: As outlined above, we attempt to quantify the error induced by imperfect recall. To this end, we first try to determine whether survey respondents really employ the kind of rationale analyzed in our theoretical model. Specifically, we look at the factor "time used for answering the survey" as a function of the price of time, for which we construct a proxy. If our model is correct, this derivative should be negative once we control for household technology parameters like mental capacity. Note that a complete list of all cost factors and prices would allow us to identify the cost function of the individuals' optimization problem (up to a constant). Applying Hoderlein's (2006) test for negativity of the derivative, we obtain the results displayed in Panel A of table 1. It shows the percentage of the population for which a pointwise test does not reject the null of a negative derivative of a suitably estimated nonparametric regression. A global L_2-distance test yields a materially similar result.

This result clearly suggests that some optimizing with respect to time use is going on, as our theory predicts. Of course, this could still mean that all respondents' optimal amounts of memory are higher than the one required to nest the sigma algebra used by the econometrician. But it could also be an indication of information reduction – with the implication of all the problems we discussed above. To answer this question, we exploit the fact that in the CAMS, respondents can decide the reporting horizon for consumption items themselves. After having been informed that all quantities are to be reported at the household level, the respondents read the following introduction to the questions on consumption items: "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, for a review of the psychological literature on recall). Hence we would expect to find some differences between any two mean regressions that use alternative time horizons. Moreover, we may hope to find some information reduction structure, in particular in the answer to the more distant yearly consumption question.

Turn to the first hypothesis first. Using the specification test statistic $\hat{\Gamma}_2$ and the associated

Table 1: Comparison Weekly, Monthly, Yearly

	Panel A: Negativity		
	weekly	monthly	annual
$mc = 1$	1	0.97	1
$mc = 2$	1	0.97	1
$mc = 3$	0.87	0.98	1
	Panel B: Comparison of Regressions		
	weekly-monthly	monthly-annual	weekly-annual
$mc = 1$	0	0	0
$mc = 2$	0	0	0
$mc = 3$	0	0	0.16
	Panel C: Index Test for Information Reduction		
	weekly	monthly	annual
$mc = 1$	0	0.21	0.10
$mc = 2$	0	0.09	0.24
$mc = 3$	0	0	0.18

In Panel A we test for a negative effect of cost of time on the time needed for answering the CAMS questionnaire. We stratify the population in three groups according to how they rate their memory: excellent or very good ($mc = 1$), good ($mc = 2$), fair or poor ($mc = 3$). The percentage of the population for which negativity is not rejected is reported. Panel B presents p-values of the null hypothesis that the L_2 distance of the conditional expectations of weekly and monthly, monthly and annual, and weekly and annual food consumption (respectively) is zero. Panel C shows p-values of the test for information reduction for the three groups of reported food consumption.

bootstrap critical values, we obtain that almost all regression functions differ. More precisely, Panel B in table 1 shows the p-values of the hypothesis that the two regression functions are equal. They are almost always virtually zero, and do not show much variation, even after stratifying according to different mental capacity.

The following two graphs illustrate more precisely what is happening: The first shows a scatterplot of CAMS data of food expenditure and income. A particularly noticeable feature are the "curved streets" in the data. These are due to rounding effects, see also figures 5 and 6 below for food expenditure, which is strong evidence of information reduction.

Based on psychological research, we would expect that yearly data are more likely to be "estimated" than monthly or in particular weekly data. For our theory, this means that the association between income and food consumption should be the weaker the longer the 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 behaves according to theorem **1** (i.e., uses more information than the econometrician when backcasting), while another fraction discards the

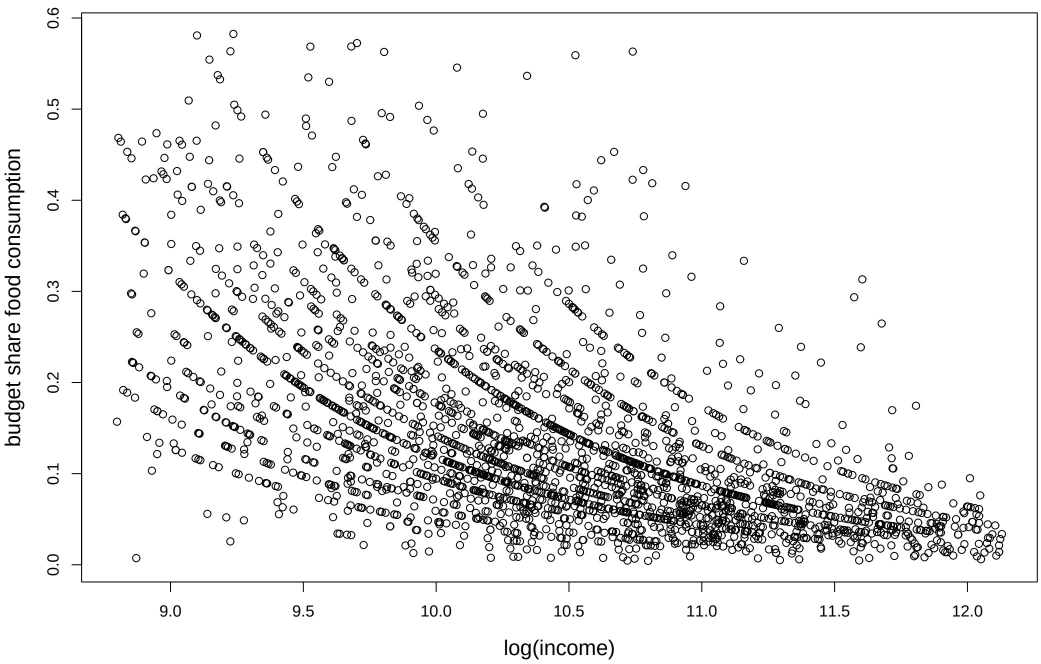

Figure 1: Data CAMS – Budget Share Food Consumption (Weekly) vs log Income.

information in the regressor of interest completely as in the first part of theorem **2**. Then by similar arguments as in point **4** in extensions to theorem **1**, we would expect the marginal effects of the mean regression (a local average of the two subpopulations!) to attenuate. Indeed, the longer the time horizon, the more likely people are to encode things, and consequently the stronger this attenuation effect should be.

This is exactly what we find in the figure 2, which shows nonparametric regression of the food budget share on totexp and some covariables, using different time horizons in the answers:

This picture as well as the associated test statistics in table 1, panel B, illustrate that left-hand side recall error matters dramatically for the economic issue we want to consider. Moreover, it looks as if we obtain the expected effect. The average derivatives are: weekly -0.065026, monthly -0.044245, yearly -0.033798. The picture changes slightly if we control for endogeneity of totexp. In this case, we obtain the ADEs: weekly -0.070767, monthly -0.029048, yearly -0.031103. As the associated fig. 3 seems to suggest as well, there is a persistent difference between the remembered weekly expenditures, and the "estimated" monthly and yearly ones. Note that the totexp effect 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. Moreover, we do find the attenuation effect we would predict if with increasing horizon increasing parts of the population do not take totexp variations into account when backcasting. But 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

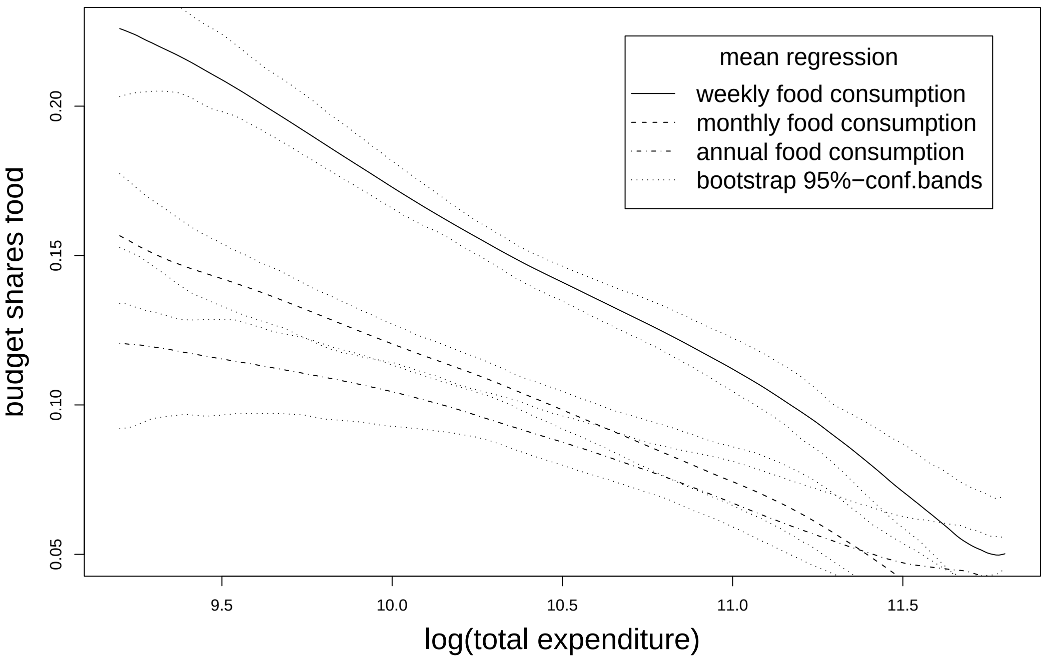

Figure 2: Regressions of Food Consumption vs Total Expenditure based on Weekly, Monthly and Yearly Budget Share Data.

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 doing so. If we compare the various groups in our data using the descriptive statistics in table A2 in the appendix, we find little or no obvious 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 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 as is seen by the following graph (figure 3). 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.

Another very interesting finding is the result on information reduction shown in Panel C of table 1. These results are with income as regressor (with totexp the results are similar, though less pronounced). Our theory predicts that information reduction may create structures where there are previously none. We use the test described above in sections 4.3 and 5.5. 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 fractions of the population to exhibit response behavior that cause left hand side recall errors of the problematic type.

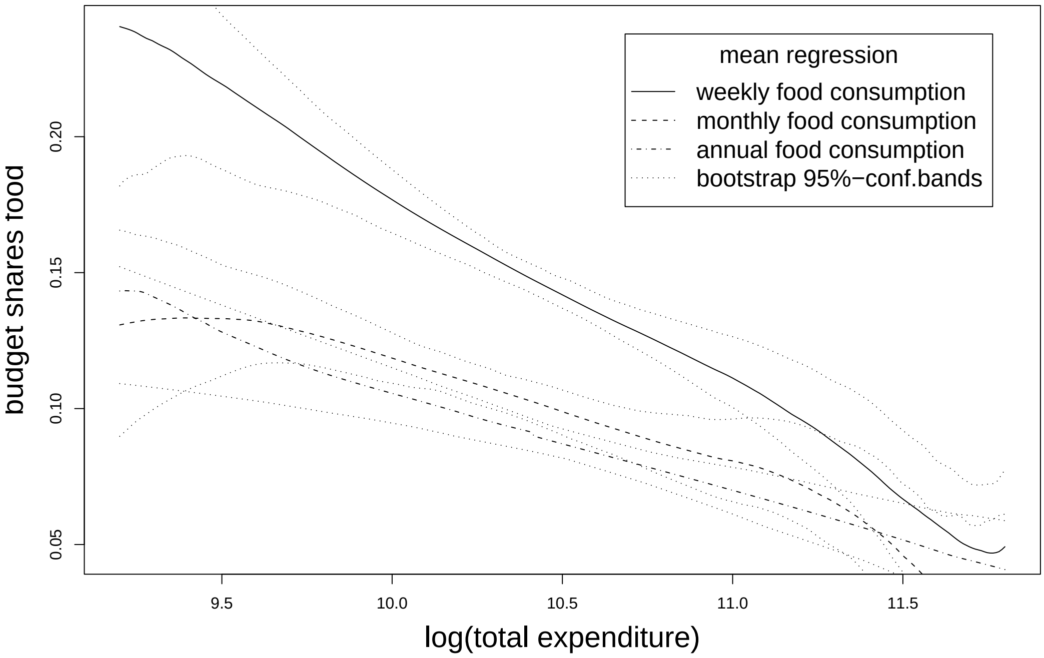

Figure 3: Regressions of Food Consumption vs Total Expenditure based on Weekly, Monthly and Yearly Budget Share Data, Controlled for Endogeneity of Total Expenditure.

A further interesting comparison is the one between similar food expenditure questions in CAMS and in the HRS main survey, which are both targeted on weekly food expenditures[9]. One hypothesis is that in the CAMS, respondents are (through the large number of consumption questions contained in the mail questionnaire) more aware of their current consumption, and hence less prone to information reduction. A related hypothesis is that respondents have more time to answer questions in the CAMS mail questionnaire than in the HRS telephone surveys (which would be called a "mode effect" in survey research). With the data at hand, we cannot distinguish between these two channels that both lead to the same observed effect. As already mentioned, in this comparison we have the less theory consistent income variable at our disposal, but the aim of the exercise is only illustrative for recall errors. The results we find are less pronounced, but a modest effects is detectable too, cf. figure 4.

The result suggest that there are some differences between the two regression functions, even though we have exploited the panel dimension and are effectively using the same individuals (and can hence rule out selection effects, as well as some time effects, because we normalize by the growth in income). A formal test underscores the differences (Panel A of table 2), and gives some indication of information reduction in the HRS (Panel B of table 2), as we would have expected. More specifically, some of the differences can actually be attributed to more rounding

[9]The precise wording in the HRS is: "About how much do you (and other family members living there) spend on food that you use at home in an average week?"

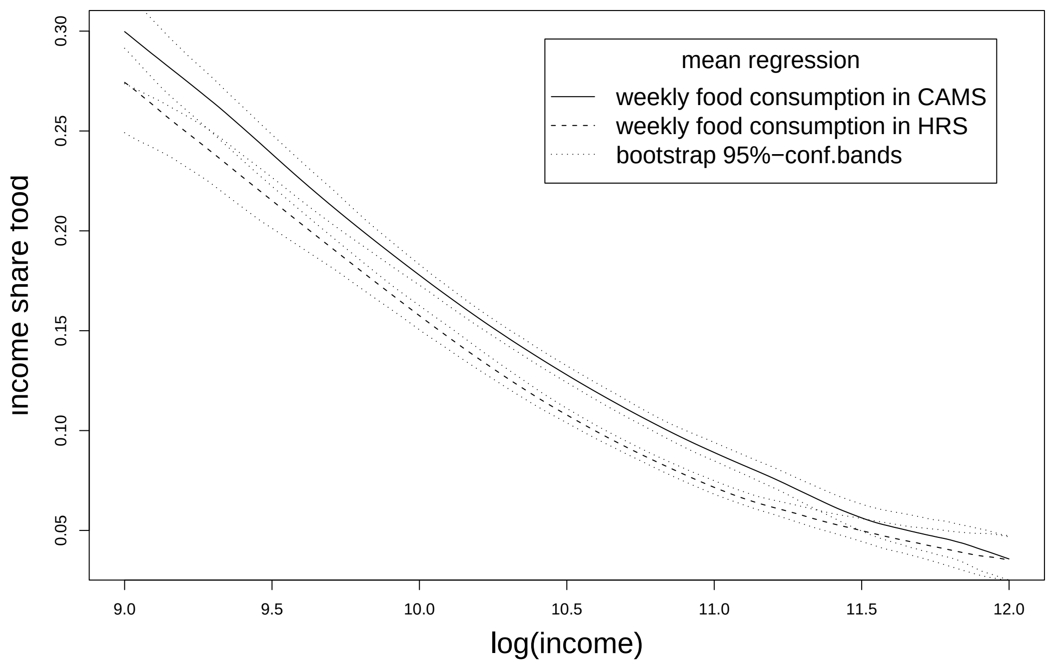

Figure 4: Regressions of Food Consumption vs Income based on Weekly Income Share Data in CAMS and HRS.

Table 2: Comparison CAMS-HRS

Panel A: Comparison of food consumption	
CAMS-HRS	
0	
Panel B: Single Index Test	
CAMS	HRS
0	0.14

Panel A of table 4 presents p-values of the null hypothesis that the L_2 distance of the conditional expectations of weekly food consumption in CAMS and HRS is zero. Panel B shows p-values of the test for information reduction in both surveys.

in the HRS than in the CAMS, as can be seen in figures 5 and 6 which display the relative frequencies of all reported values. Obviously, we have more individuals picking multiples of 50 (in particular 100), in the HRS than in CAMS.

The Effect on Testing Homogeneity: The strong evidence of differences in recall in the weekly, monthly, and yearly answers is suggestive of major recall problems in the data. Moreover, in the light of the obtained evidence on information reduction, it is appropriate to use the more conservative bounds put forward in theorem 2.

To test for homogeneity of degree zero (i.e. absence of money illusion, which is a core

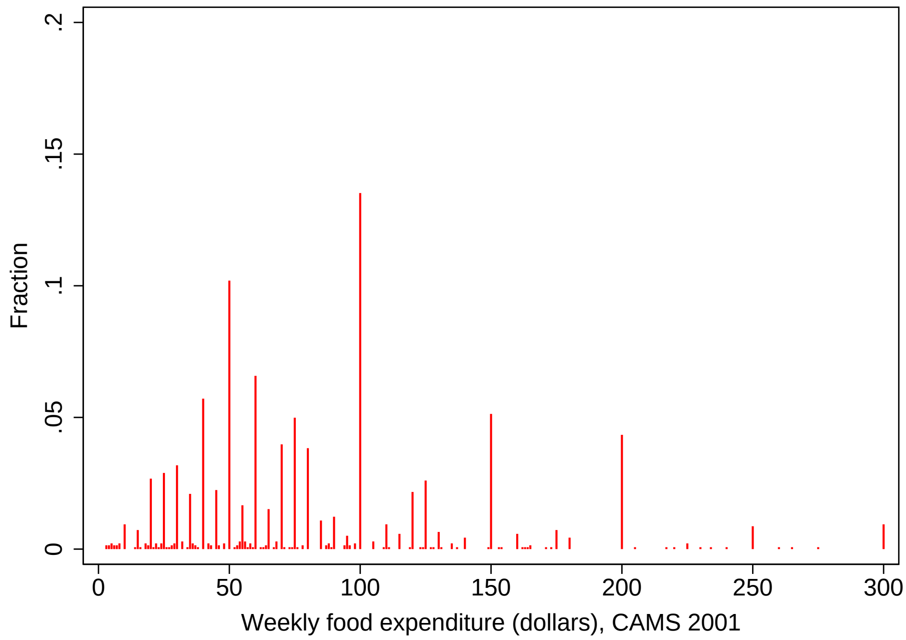

Figure 5: Data CAMS – Weekly Food Consumption Expenditure.

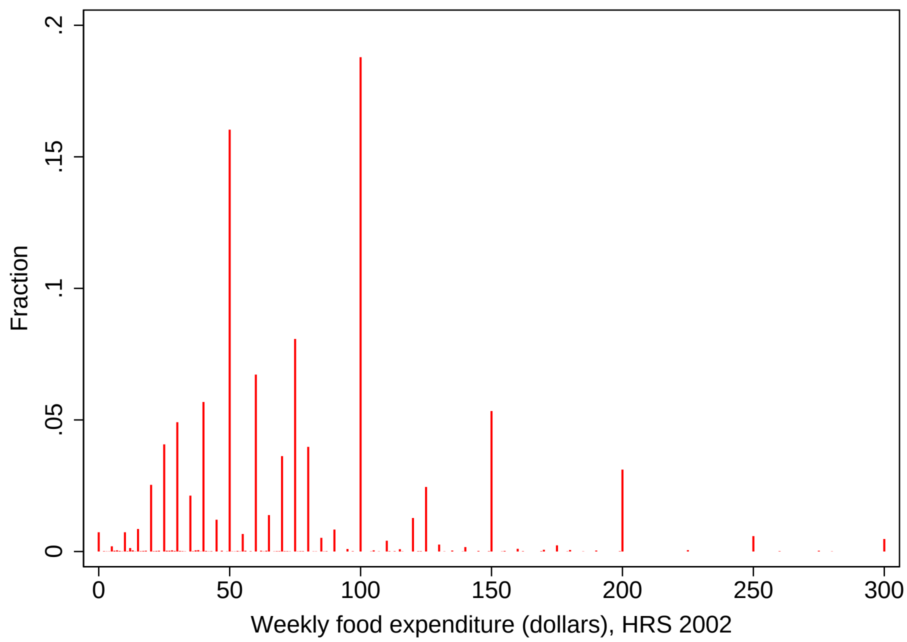

Figure 6: Data HRS – Weekly Food Consumption Expenditure.

economic property) without bounds, we apply a conventional homogeneity test using methods put forward in section **5.4**. We strongly reject homogeneity with a p-value of virtually zero. This coincides with evidence gathered elsewhere in the literature. Based on our analysis thus far, there are two potential reasons for this rejection: First, here we are using the HRS data since we require time series variations in prices. But this means that we have to employ the economically less plausible income definition. Second, following our discussion there is strong evidence for left hand side "information reduction" recall error. To account for the possible effect of the latter on our results, we implement the test statistic $\widehat{\Gamma}_5$ discussed in sections **4.4** and **5.5,** which allows for recall errors through bounds. The point estimate is -0.021, and fig. 7 shows the bootstrap distribution of the test statistic. Not surprisingly, $\underline{\Gamma_5}$ is soundly negative, and we conclude that we cannot reject homogeneity any longer. Indeed, the bootstrap distribution in fig. 7 suggests that we might even reject $H_0 : \Gamma_5 = 0$ in favor of the one sides alternative that it is smaller, but in light of our theory this would have no added implication. We conclude that a possible explanation for the frequent rejection of homogeneity in consumer demand is indeed left hand side recall error; controling for this effect we do not find homogeneity rejected any longer.

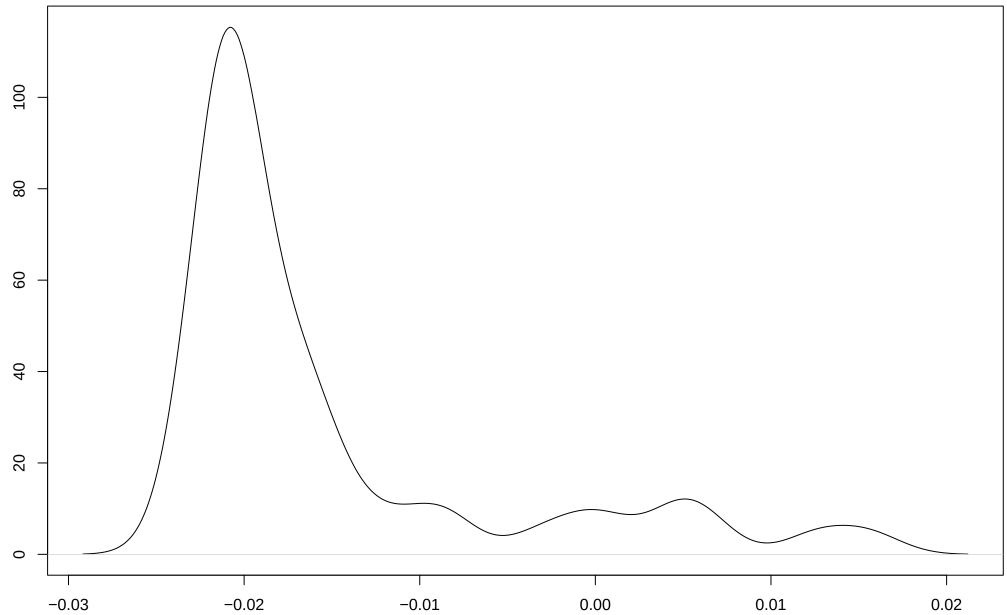

Figure 7: Bootstrap Distribution of $\widehat{\Gamma}_5$.

The Design of a Survey: Our theory underscores the importance of the design of a survey for information gathering respondents. In this subsection we briefly illustrate the scope of the effect using data from a controlled survey experiment on bracketing effects that has been conducted in February 2001 (with an existing Dutch internet panel survey, the CentER Panel). To keep the exposition consistent, we focus on an experiment concerning monthly food expenditure. Recall that food data is usually considered to be relatively accurately measured. Similar, but often stronger response effects can be found in questions on other items. Winter (2002) contains a detailed discussion of these experiments. Here we just note that other than a typical set of household covariates we use again a similar definition as income as above.

The wording of the food expenditure question in this experiment was straightforward: "What amount, in guilders, did your household spend on food last month?" In the survey experiment, respondents were randomly assigned to one of three treatments. Each group was presented with a differently scaled range card, each with eight brackets. The seven break points ranged, in steps of 50 guilders, from 275 to 575 (low treatment), 450 to 750 (medium treatment), and 675 to 975 (high treatment). The three treatment groups contained 31, 40, and 31 observations, respectively. We also have data from a control group that reported their food expenditure using an open-ended response format (294 observations).

If all respondents reported correctly, that is, if they ignored information provided in the brackets and selected themselves into the bracket in which true food expenditure falls, the distributions generated by the three treatments should be (almost) identical. Figure 8 shows that they are astonishingly different.[10]

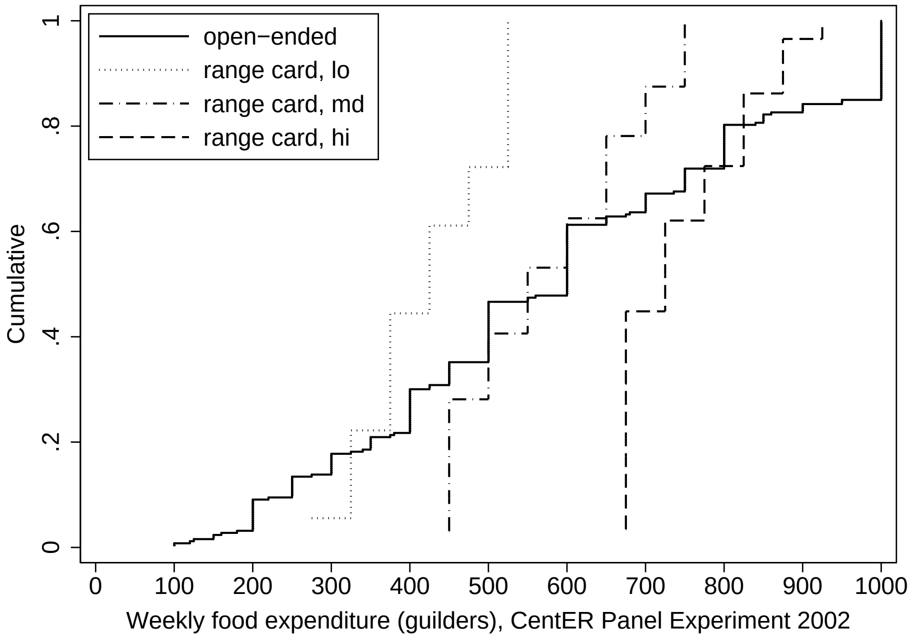

Figure 8: Data from CentER Experiment – Empirical Cumulative Distribution Functions of Responses on a Weekly Food Expenditure Question under Variations in Brackets.

In particular, the distributions generated by range-card questions are much more compressed than the distribution of the continuous responses to the open-ended question. Also, the location of those distributions is strongly influenced by the level of the brackets used in the range cards.

Evidently, respondents use information provided by the brackets. In particular, they seem to select themselves into an "average" bracket – regardless of where that average actually is. In Hoderlein and Winter (2007), we extend our theoretical framework to explain this and

[10] As a side note, this finding indicates a violation of the "correct interval reporting" assumption made by Manski and Tamer (2002) in the derivation of nonparametric bounds for regressions with interval data.

similar rule-of-thumb phenomena formally. Here we just note that in our setup substituting own information by misleading survey information obscures the effect we are interested in, and most likely reduces it. The parametric result as summarized in table 3 seems to confirm this intuition. Table 3 shows the results of several regressions with log food expenditure as the dependent variable and some household covariates (which are contained in the CentER Panel background data and were not elicited in the experiment). The first four columns contain Maximum Likelihood estimates of a linear index model for interval data with known thresholds. The first three columns use the data from the three treatments with low, medium, and high brackets separately. In column (4), data from all three treatments are pooled. Columns (5) and (6) are based on data from the control group whose respondents answered an open-ended question. For column (5), we coarsened the continuous data provided by the control group artificially, using the thresholds from the medium brackets. In column (6), we use the continuous responses as they were given by the control-group respondents and estimate the model by OLS. This last column is our benchmark (since one may argue that in this survey experiment, respondents answer the open-ended question with negligible error). Indeed, if we consider the first row, the income semi-elasticity is significantly smaller in the bracketed case than in the open-ended one. The same holds also true for most other effects.

Table 3: Regressions of log weekly food expenditure, CentER Panel Experiment 2002

	(1)	(2)	(3)	(4)	(5)	(6)
Response format	range card	range card	range card	range card	open-ended	open-ended
Scale of dependent variable	low brackets	medium brackets	high brackets	pooled brackets	medium brackets (artificial)	continuous
Estimation method	ML interval	ML interval	ML interval	ML interval	ML interval	OLS
Log net HH income	0.037	0.206	0.178	0.168	0.364	0.327
	(0.158)	(0.109)	(0.099)	(0.076)*	(0.103)**	(0.085)**
Respondent's age	0.021	0.022	0.006	0.008	0.023	0.023
	(0.027)	(0.033)	(0.013)	(0.016)	(0.017)	(0.015)
Respondent's age, squared	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Respondent retired	-0.160	-0.063	-0.204	-0.214	-0.177	-0.122
	(0.142)	(0.300)	(0.165)	(0.149)	(0.133)	(0.122)
Household size	0.429	0.117	-0.008	0.104	0.190	0.174
	(0.167)*	(0.052)*	(0.037)	(0.037)**	(0.049)**	(0.034)**
Number of children in HH	-0.006	-0.035	0.013	-0.010	-0.073	-0.077
	(0.190)	(0.070)	(0.039)	(0.045)	(0.064)	(0.045)
Low brackets				-0.110		
				(0.070)		
High brackets				0.228		
				(0.066)**		
Constant	4.863	3.845	4.910	4.494	2.425	2.762
	(0.865)**	(1.049)**	(0.731)**	(0.583)**	(0.814)**	(0.670)**
Number of observations	31	40	31	102	294	294
Log likelihood	-41.23	-70.67	-49.42	-171.72	-433.05	
R-squared						0.30

Notes: Robust standard errors in parentheses. The symbols * and ** denote significance at 5% and 1%, respectively.

Comparing the first three columns, we see that the coefficients of household income are strikingly different depending on what treatment is used; unfortunately, they are also estimated imprecisely due to the small sample sizes in the experiment. Nevertheless, the magnitudes of these differences are such that substantive conclusions a researcher would draw may be affected. In a real world application, the data would of course come only from one range-card, and the analyst would have no way of detecting whether and how strong the effects induced by the location of bracket bounds are. We leave a more detailed analysis, including further experiments, for future research. What is already strikingly obvious from this example is the sheer extend to which respondents react to survey design, suggesting that the individuals' information set is very small indeed. Moreover, it is urging stronger research efforts in this area to determine consequences and implications of this phenomena. In this, the structural model proposed in this paper may provide some guidance.

7 Summary and Outlook

In this paper, we propose to look at measurement error in household surveys from a new perspective. We stress the role of a respondent's limited ability to recall information. We argue, through a large number of individual points and issues, that this change in viewpoint is an useful exercise. One important issue is that in the social sciences, left-hand side measurement errors should be regarded as potentially problematic (see also the discussion in Hausman, 2001). Another interesting insight obtained by this change of viewpoint is that it appears to be crucial whether individuals really use all information the econometrician is interested in – including instruments, for example. Our results imply that instruments may become useless if individuals do not use them actively in the recall process.

For the various problems that appear, we devise a number of remedies. The remedies are of two sorts. *Ex post*: We suggest immediate cures that an applied researcher could implement to accommodate recall problems in the analysis data from existing household surveys. *Ex ante*: We provide suggestions to researchers and institutions actually designing surveys in order to minimize arising problems ex ante. Of course, it remains to be determined how useful these remedies are. In two applications, 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.

Appendix 1: Proofs for Identification Results

Regularity Conditions

Assumption 12. *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 13. *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* $y \in \mathbb{R}$. $\mathbb{E}\left[\tilde{Y}|W = w\right]$ *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 $\mathbb{E}[Y|\mathcal{F}_m] = \varphi(W, A_1)$. Then, by the conditional independence assumption 2.1, and under the regularity assumptions 12, using a theorem of Hoderlein and Mammen (2006) 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 $\partial_{x_1}\varphi(W, A_1) = \mathbb{E}[\partial_{x_1}\phi|\mathcal{F}_m]$, using uniform boundedness of $\partial_{x_1}\phi$, together with the 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 $\partial_{x_1} k_\alpha(w) = \mathbb{E}[\partial_{x_1}\phi|W = w, \tilde{Y} = k_\alpha(w)]$. Note that under the weaker assumption 3.2 we may not apply the LIE, and consequently, $\partial_{x_1} k_\alpha(w) \neq \mathbb{E}[\partial_{x_1}\phi|W = w, \tilde{Y} = k_\alpha(w)]$. To see Theorem **1.2**, note that in this case 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 13, $m(w) = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_m]|W = w] = \mathbb{E}[Y|W = w] = \mathbb{E}[\phi(X, A)|W = w]$, by assumption 3.2 and application of LIE. Then by bounded convergence in connection with the regularity assumption 13, 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 (2006). 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 $l_\alpha(w) = f_{\tilde{Y}|W}(k_\alpha(w); w)^{-1}\mathbb{E}\left[\mathbf{1}[\tilde{Y} \leq k_\alpha(w)]\partial_{x_1}\log f_{J|W}(J; w)|W = w\right]$. Rearranging terms produces the result. *Q.E.D.*

Proof of Statements in Extension 2 to Theorem 1

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

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

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 $A=(A_1,A_2)$, $W=X_1$, $Y=\phi(X_1,A)=A_1^2A_2X_1$, and assume that $med_{A_2}(x_1,a_1)=1$ from which $med_Y(x_1,a_1)=a_1^2x_1$. Note that assumption 2.1 reduces to $A \perp X_1$. Then, observe that $\mathbb{E}[\partial_{x_1}\phi(X,A)|W=w, A_1=a_1, Y=med_Y(w,a_1)] = a_1^2$, implying that $\partial_{x_1}\mathbb{E}\left[\tilde{Y}|W=w\right] = \mathbb{E}\left[A_1^2\right]$. Next, $\mathbb{E}[\partial_{x_1}\phi(X,A)|W=w,A_1=a_1,Y=y]=yx_1^{-1}$ and hence $\mathbb{E}[\partial_{x_1}\phi(X,A)|W=w] = \mathbb{E}[Y|X_1=x_1]x_1^{-1} = \mathbb{E}[A_1^2A_2] \neq \mathbb{E}\left[A_1^2\right]$ unless e.g., $Cov(A_1,A_2)=0$ 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 $\tilde{Y}=med_Y(W,A_1)=\psi(W,A_1)$ is strictly monotone in the scalar first component, with 2.1 strengthened to $A_1|W \sim U\left[0,1\right]$, i.e. the conditions of equation (3.2) hold, then

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

This shows equation (3.2). *Q.E.D.*

Theorem 6

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

Theorem 6. *Let all the variables and functions be as defined above. Let assumptions 1, 2, 4.1 and 12 hold. Then follows $\partial_{x_1}k_\alpha(w)=0$ (a.s.). If 4.1 is replaced by 4.2, then*

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

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

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

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

Remark A.1: Discussion of Theorem 6 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 $\mathbb{E}[\partial_{x_1}\phi(X,A)|\mathcal{F}_m]$. Bounds may be derived as in theorem 2.

Proof of Theorems 2 and 6

To see the first statement, start by noting that by assumption 13, $\mathbb{E}[Y|\mathcal{F}_m]$ exists and is differentiable. Then, by assumption 4.1 and the LIE,

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

Hence,

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

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 (2006) to $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}_m] = \psi(W_{-1}, A_1)$ 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

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

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

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

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

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

by Hölder's inequality. For the case of the quantile, we apply again Hoderlein and Mammen (2006) to $\tilde{Y} = \mathbb{E}[Y|\mathcal{F}_m] = \psi(W_{-1}, A_1)$ 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 $\partial_{x_1}\psi(X_1,S,A_1)=\mathbb{E}[\partial_{x_1}\phi|X_1,S,A_1]+Cov\left[U,Q_{x_1}|X_1,S,A_1\right].$ Finally, to see the last equality, note first that (with $b=x'\beta$),

$$\begin{aligned}
\partial_{x_1}\mathbb{E}\left[\tilde{Y}|X=x,S=s\right] &= \partial_{x_1}\mathbb{E}\left[\mathbb{E}\left\{\phi(X,A)|B,S,A_1\right\}|X=x,S=s\right] \\
&= \mathbb{E}\left[\partial_{x_1}\mathbb{E}\left\{\phi(X,A)|B=b,S=s,A_1\right\}|X=x,S=s\right] \qquad (7.1)\\
&= \beta_1\mathbb{E}\left[\partial_b\mathbb{E}\left\{\phi(X,A)|B=b,S=s,A_1\right\}|X=x,S=s\right].
\end{aligned}$$

Next, observe that

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

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

$$\begin{aligned}
&\partial_b\mathbb{E}\left\{\phi(X,A)|B=b,S=s,A_1=a\right\}\\
=&\int_{\mathcal{X}_{-1}\times\mathcal{A}_{-1}}\partial_b\left[\phi(\tau(b,\xi_{-1}),a)f_{A_{-1}X_{-1}|BSA_1}(a_{-1},\xi_{-1};b,s,a_1)\right]v(da_{-1},d\xi_{-1})\\
=&\beta_1^{-1}\mathbb{E}\left\{\partial_{x_1}\phi(X,A)|B=b,S=s,A_1=a\right\} \qquad (7.2)\\
&+\int_{\mathcal{X}_{-1}\times\mathcal{A}_{-1}}\phi(\tau(b,\xi_{-1}),a)\partial_b f_{A_{-1}X_{-1}|BSA_1}(a_{-1},\xi_{-1};b,s,a_1)v(da_{-1},d\xi_{-1}).
\end{aligned}$$

Now we focus on the second term on the last rhs of (7.2) . 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

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

with derivatives

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

Evaluated at $b = \beta' x$

$$
\begin{aligned}
H(\beta' x, s) &= \int_{\mathcal{X}_{-1}} f_{X|S}(\tau(b, x_{-1}); s) dx_{-1} = f_{X_1}(x_1) \\
H'(\beta' x, s) &= \beta_1^{-1} \int_{\mathcal{X}_{-1}} \partial_{x_1} f_{X|S}(\tau(b, x_{-1}); s) dx_{-1} = \beta_1^{-1} f'_{X_1|S}(x_1; s) \\
\beta_1 \partial_b f_{X_{-1}|BS}(x_{-1}; \beta' x, s) &= \frac{\partial_{x_1} f_{X|S}(x; s)}{f_{X_1|S}(x_1; s)} - f_{X|S}(x; s) \frac{f'_{X_1|S}(x_1, s)}{f_{X_1|S}(x_1; s)^2} \\
&= \partial_{x_1} \frac{f_{X|S}(x; s)}{f_{X_1|S}(x_1; s)} \qquad (7.3) \\
&= \partial_{x_1} f_{X_{-1}|X_1 S}(x_{-1}; x_1, s).
\end{aligned}
$$

Combining equations (7.1), (7.2) and (7.3) yields

$$
\begin{aligned}
\partial_{x_1} \mathbb{E}\left[\tilde{Y} | X = x, S = s\right] &= \mathbb{E}\left[\partial_{x_1} \phi(X, A) | B = b, S = s\right] \\
&\quad + \mathbb{E}\left[Y \partial_{x_1} \log f_{X_{-1}|X_1 S}(x_{-1}; x_1, s) | B = b, S = s\right].
\end{aligned}
$$

Finally, the quantile follows as direct extension by similar arguments. *Q.E.D.*

Proof of Theorem 3

Start with the first statement,

$$
\begin{aligned}
& \partial_y \mathbb{E}\left[\tilde{X}_1 | Y = y, W_{-1} = w_{-1}\right] \\
= & \partial_y \mathbb{E}\left[\mathbb{E}\left[X_1 | Y, W_{-1}, A_1\right] | Y = y, W_{-1} = w_{-1}\right] \\
= & \mathbb{E}\left[\partial_y \mathbb{E}\left[X_1 | Y, W_{-1}, A_1\right] | Y = y, W_{-1} = w_{-1}\right] \\
= & \mathbb{E}\left[\partial_y \psi(Y, X_{-1}, A) | Y = y, W_{-1} = w_{-1}\right] \\
& + \int_{\mathcal{A}} \psi(y, x_{-1}, a) \partial_y f_{A|A_1, Y, W_{-1}}(a; a_1, y, w_{-1}) \, v(da) f_{A_1|YW_{-1}}(a_1; y, w_1) \, v(da_1).
\end{aligned}
$$

For the first term,

$$
\mathbb{E}\left[\partial_y \psi(Y, X_{-1}, A) | Y = y, W_{-1} = w_{-1}\right] = \mathbb{E}\left[\partial_{x_1} \phi(X, A)^{-1} | Y = y, W_{-1} = w_{-1}\right],
$$

while the second term on the last rhs rewrites again as $Cov\left[X_1, \tilde{Q}_y | Y = y, W_{-1} = w_{-1}\right]$, where $\tilde{Q}_y = \partial_y \log f_{A|A_1, Y, W_{-1}}$. Finally, for the quantile, we obtain

$$
\begin{aligned}
\partial_y k_\alpha(y, w_{-1}) &= \mathbb{E}\left[\partial_y \mathbb{E}\left[X_1 | Y, W_{-1}, A_1\right] | Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})\right] \\
&= \mathbb{E}\left[\partial_{x_1} \phi(X, A)^{-1} | Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})\right] \\
&\quad + Cov\left[X_1, \tilde{Q}_y | Y = y, W_{-1} = w_{-1}, \tilde{X}_1 = k_\alpha(y, w_{-1})\right],
\end{aligned}
$$

where the first equality uses again a theorem of Hoderlein and Mammen (2007). *Q.E.D.*

References

[1] AIT-SAHALIA, Y., P. BICKEL, and T. STOKER (2001): Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105, 363–412.

[2] ALTONJI, J., and R. MATZKIN (2005): Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73, 1053–1103.

[3] BATTISTIN, E., R. MINIACI, and G. WEBER (2003): What do we learn from recall consumption data? *Journal of Human Resources*, 38(2), 354–385.

[4] 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.

[5] BROWNING, M., T. F. CROSSLEY, and G. WEBER (2003): Asking consumption questions in general purpose surveys. *Economic Journal*, 113, F540–F567.

[6] CARROLL, R. J., D. RUPPERT, and L. A. STEFANSKI (1995): *Measurement Error in Nonlinear Models*. New York: Chapman & Hall.

[7] 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.

[8] DATTA, S. (1995): On a Modified Bootstrap for Certain Asymptotically Nonnormal Statistics, *Statistics and Probability Letters*, 24, 91–98.

[9] GROVES, R. M. (1989): *Survey Errors and Survey Costs*. New York: Wiley.

[10] 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.

[11] HAUSMAN, J. A., W. K. NEWEY, and J. L. POWELL (1995): Nonlinear errors in variables estimation of some Engel curves. *Journal of Econometrics*, 65, 205–233.

[12] HODERLEIN, S. (2006): Nonparametric demand systems, instrumental variables, and a heterogeneous population. Working Paper, University of Mannheim.

[13] HODERLEIN, S., and E. MAMMEN (2006): Identification of local average structural derivatives in nonseparable models. Working Paper, University of Mannheim.

[14] HOROWITZ, J., and C. MANSKI, (1995): Identification and robustness with contaminated and corrupted data, *Econometrica*, 63(2), 281–302.

[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] JUSTER, F. T., and R. SUZMAN (1995): An overview of the Health and Retirement Study. *Journal of Human Resources*, 30, S7–S56.

[18] IMBENS, G., and W. NEWEY (2003): Identification and estimation of triangular simultaneous equations models without additivity. Working Paper, MIT.

[19] MAMMEN, E. (1992): *When does the Bootstrap Work? Asymptotic Results and Simulations*, New York, Springer.

[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] MANSKI, C. and E. TAMER (2002): Inference on regressions with Interval Data on a Regressor or Outcome, *Econometrica*, 70, 519–546.

[23] MATZKIN, R. (2007), Nonparametric Survey Response Errors, *Working Paper*, Northwestern University.

[24] McFADDEN, D. (2006): How consumers respond to incentives. Working Paper, UC Berkeley.

[25] MOORE, J. C., L. L. STINSON, and E. J. WELNIAK, Jr. (1999): Income reporting in surveys: Cognitive issues and measurement error. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research*, 155–174. New York: Wiley.

[26] PHILIPSON, T. (1997): Data markets and the production of surveys. *Review of Economic Studies*, 64(1), 47–72.

[27] ———— (2001): Data markets, missing data, and incentive pay. *Econometrica*, 69(4), 1099–1111.

[28] TOURANGEAU, R., L. J. RIPS, and K. RASINSKI (2000): *The Psychology of Survey Response*. New York, NY and Cambridge, UK: Cambridge University Press.

[29] WANSBEEK, T. and E. MEIJER (2000): *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier.

[30] WINTER, J. (2002): Bracketing effects in categorized survey questions and the measurement of economic quantities. Discussion Paper No. 02-35, Sonderforschungsbereich 504, University of Mannheim.

Separate Appendices

Appendix 2: Descriptive Statistics

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	edyrs	hisp	male	race	shlt	srm
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).

Appendix 3: Proofs of Large Sample Results

Proof of Theorem 4

In this separate appendix which is intended as supplementary material, we use as an alternative notation for the random vectors $W_1 = X$ and $W_2 = Z$. The reasons are that this is more in line with the standard nonparametric testing literature, and it makes the exposition clearer. As should be obvious from the main text, these variables have nothing in common with similarly denoted ones in the main text. Hence, from now on $Y \in \mathbb{R}^{d_Y}$ is a d_Y-dimensional dependent variable, and $X \in \mathbb{R}^{d_X}, Z \in \mathbb{R}^{d_Z}$ are the predictors. The hypothesis to be tested is whether Z can be omitted from the regression of Y on (X, Z).

For abbreviation we introduce $V_i = (X_i, Z_i)$ and $W_i = (Y_i, X_i, Z_i)$ and decompose the statistic in the following way

$$\begin{aligned}\hat{\Gamma}_{\mathcal{K}} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}\frac{K_h(V_i - V_j)}{\hat{f}_h(V_i)}(Y_j^k - \hat{m}_{\tilde{h}}^k(X_j))\Big)^2 A_i \\ &= \hat{\Gamma}_{\mathcal{K}1} + \hat{\Gamma}_{\mathcal{K}2} + \hat{\Gamma}_{\mathcal{K}3} + \hat{\Gamma}_{\mathcal{K}4}, \end{aligned} \tag{7.4}$$

where

$$\hat{\Gamma}_{\mathcal{K}1} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{Y_j^k - \mu^k(V_j)}{\hat{f}_h(V_i)}\Big)^2 A_i$$

$$\hat{\Gamma}_{\mathcal{K}2} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{\mu^k(V_j) - m^k(X_j)}{\hat{f}_h(V_i)}\Big)^2 A_i$$

$$\hat{\Gamma}_{\mathcal{K}3} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{m^k(X_j) - \hat{m}_{\tilde{h}}^k(X_j)}{\hat{f}_h(V_i)}\Big)^2 A_i$$

and $\hat{\Gamma}_{\mathcal{K}4}$ contains all cross terms. Note that under H_0 we have that $\hat{\Gamma}_{\mathcal{K}2} = 0$ almost surely. For the third term it holds that

$$\begin{aligned}|\hat{\Gamma}_{\mathcal{K}3}| &\le \max_{k=1,\ldots,d_Y}\sup_{x\in\mathcal{A}}|m^k(X_j) - \hat{m}_{\tilde{h}}^k(X_j)|^2\sup_{v\in\mathcal{A}}|a(v)| \\ &= O_P(\tilde{h}^{2r} + \frac{\log n}{n\tilde{h}^{d_X}}) \\ &= o_P\big(n^{-1}h^{(d_X+d_Z)/2}\big)\end{aligned}$$

under assumption 10.3. If $\hat{\Gamma}_{\mathcal{K}1} = O_P(n^{-1}h^{-(d_X+d_Z)/2})$, it follows by an application of Cauchy-Schwarz that $\hat{\Gamma}_{\mathcal{K}4} = o_P(n^{-1}h^{-(d_X+d_Z)/2})$. Then, it remains to show that the first term has the requested asymptotic distribution. For this we write

$$\begin{aligned}\hat{\Gamma}_{\mathcal{K}1} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{Y_j^k - \mu^k(V_j)}{f(V_i)}\Big)^2\Big(\frac{f(V_i)}{\hat{f}_h(V_i)}\Big)^2 A_i \\ &= (I_{\mathcal{K}n} + \Delta_{\mathcal{K}n})(1 + o_P(1))\end{aligned}$$

where we have defined

$$I_{\mathcal{K}n} = \int \sum_{k=1}^{d_Y} \Big(\frac{1}{n} \sum_{j=1}^{n} K_h(v - V_j) \frac{Y_j^k - \mu^k(V_j)}{f(v)} \Big)^2 a(v) f(v) \, \mathrm{d}v \tag{7.5}$$

$$\Delta_{\mathcal{K}n} = \int \sum_{k=1}^{d_Y} \Big(\frac{1}{n} \sum_{j=1}^{n} K_h(v - V_j) \frac{Y_j^k - \mu^k(V_j)}{f(v)} \Big)^2 a(v) (\hat{f}_e(v) - f(v) \, \mathrm{d}v \tag{7.6}$$

and $\hat{f}_e = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(V_i)}(v)$ denotes the empirical distribution of the sampled data. Starting with the leading term, we rearrange $I_{\mathcal{K}n}$ to obtain

$$\begin{aligned} I_{\mathcal{K}n} &= \frac{1}{n^2} \sum_{i<j} \sum_{k=1}^{d_Y} \int K_h(v - V_i) \frac{Y_i^k - \mu^k(V_i)}{f(v)} K_h(v - V_j) \frac{Y_j^k - \mu^k(V_j)}{f(v)} a(v) f(v) \, \mathrm{d}v \\ &\quad + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{d_Y} \int \big(K_h(v - V_i) \frac{Y_i^k - \mu^k(V_i)}{f(v)} \big)^2 a(v) f(v) \, \mathrm{d}v \\ &= I_{\mathcal{K}n,1} + I_{\mathcal{K}n,2} \end{aligned} \tag{7.7}$$

Now it remains to show

$$nh^{(d_X+d_Z)/2} I_{\mathcal{K}n,1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{\mathcal{K}}^2) \tag{7.8}$$

$$nh^{(d_X+d_Z)/2} I_{\mathcal{K}n,2} - h^{-(d_X+d_Z)/2} B_{\mathcal{K}} \xrightarrow{P} 0 \tag{7.9}$$

$$nh^{(d_X+d_Z)/2} \Delta_{\mathcal{K}n} \xrightarrow{P} 0. \tag{7.10}$$

From this the statement of the theorem follows.. *Q.E.D.*

Proof of (7.8) Write

$$I_{\mathcal{K}n,1} = \sum_{i<j} h_n(W_i, W_j)$$

as U-statistic with kernel

$$h_n(W_i, W_j) = \frac{2}{n^2 h^{d_X+d_Z}} \sum_{k=1}^{d_Y} (Y_i^k - \mu^k(V_i))(Y_j^k - \mu^k(V_j)) \int K(u) K(u + (V_i - V_j)/h) \frac{a(V_i + uh)}{f(V_i + uh)} \, \mathrm{d}u.$$

where a change of variables has been applied. Asymptotic normality is shown by using a central limit theorem for generalized U-statistics (theorem 3 in de Jong, 1987) with a kernel function $h_n(\cdot, \cdot)$ that is symmetric, centered ($\mathbb{E}\, h_n(W_1, W_2) = 0$) and degenerate ($\mathbb{E}(h_n(W_1, W_2) \mid W_1) = \mathbb{E}(h_n(W_1, W_2) \mid W_2) = 0,\ \mathbb{P}$-a.s.). Under the conditions

$$\frac{\max_{1 \le i \le n} \sum_{j=1}^{n} \mathbb{E}\, h_n(W_i, W_j)}{Var\,[I_{\mathcal{K}n,1}]} \xrightarrow{P} 0 \qquad \text{and} \qquad \frac{\mathbb{E}\, I_{\mathcal{K}n,1}^4}{(Var\,[I_{\mathcal{K}n,1}])^2} \xrightarrow{P} 3 \tag{7.11}$$

it follows that

$$\sqrt{2}\frac{I_{\mathcal{K}n,1}}{\sqrt{Var\left[I_{\mathcal{K}n,1}\right]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

It is immediate to see that the kernel is degenerate, symmetric and centered. No, we introduce $\sigma_n^2 = \mathbb{E}\, h_n(W_i, W_j)^2$. As we have independent and identically distributed data we can write

$$\max_{1\le i\le n} \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{E}\, h_n(W_i, W_j)^2 = (n-1)\sigma_n^2$$

and

$$\begin{aligned} Var\left[I_{\mathcal{K}n,1}\right] &= \sum_{i_1<i_2} Var\left[h_n(W_{i_1}, W_{i_2})\right] + \sum_{i_1<i_2} \sum_{\substack{i_3<i_4\\ (i_3,i_4)\neq(i_1,i_2)}} Cov\left[h_n(W_{i_1}, W_{i_2}), h_n(W_{i_3}, W_{i_4})\right] \\ &= \frac{n(n-1)}{2}\sigma_n^2 \end{aligned}$$

because $h_n(\cdot,\cdot)$ is centered. From these two results the first condition in equation (7.11) is established. For the second calculate

$$\begin{aligned} \mathbb{E}\, I_{\mathcal{K}n,1}^4 = \sum_{i_1<i_2} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^4 &+ 3\sum_{i_1<i_2} \sum_{\substack{i_3<i_4\\ (i_3,i_4)\neq(i_1,i_2)}} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^2 h_n(W_{i_3}, W_{i_4})^2 \\ &+ 24 \sum_{i_1<i_2} \sum_{i_3\neq i_1,i_2} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^2 h_n(W_{i_1}, W_{i_3}) h_n(W_{i_2}, W_{i_3}) \\ + 3\sum_{i_1}\sum_{i_2\neq i_1}\sum_{i_3\neq i_1,i_2}\sum_{i_4\neq i_1,i_2,i_3} &\mathbb{E}\, h_n(W_{i_1}, W_{i_2}) h_n(W_{i_2}, W_{i_3}) h_n(W_{i_3}, W_{i_4}) h_n(W_{i_4}, W_{i_1}) \end{aligned} \tag{7.12}$$

where all vanishing terms (with $\mathbb{E}\, h_n(W_{i_1}, W_{i_2}) = 0$) are omitted. To show the second condition, the remaining terms have to be calculated. Starting with the denominator, we have to calculate

$$\sigma_n^2 = \mathbb{E}\, h_n(W_1, W_2)^2. \tag{7.13}$$

Resolving the square and changing variables[11] to $\tilde{v} = (v - v_1)/h$ together with expanding $a(\cdot)$ and $f(\cdot)$ yields

$$\begin{aligned} \sigma_n^2 = \frac{4}{n^4 h^{2(d_X+d_Z)}} \sum_{k,k'} \iint K(\tilde{v}) \frac{y_1^k - \mu^k(v_1)}{f(v_1)} K(\tilde{v} + (v_1 - v_2)/h) \frac{y_2^k - \mu^k(v_2)}{f(v_1)} a(v_1) f(v_1)\, \mathrm{d}\tilde{v} \\ \times \int K(\tilde{v}) \frac{y_1^{k'} - \mu^{k'}(v_1)}{f(v_1)} K(\tilde{v} + (v_1 - v_2)/h) \frac{y_2^{k'} - \mu^{k'}(v_2)}{f(v_1)} a(v_1) f(v_1)\, \mathrm{d}\tilde{v} \\ f(y_1, v_1) f(y_2, v_2)\, \mathrm{d}y_1\, \mathrm{d}v_1\, \mathrm{d}y_2\, \mathrm{d}v_2 (1 + O(h)). \end{aligned}$$

[11] Here the notation is simplified. As v_1 is $d_X + d_Z$-dimensional one has to apply $d_X + d_Z$ substitutions.

Now substitute $\tilde{\tilde{v}} = (v_1 - v_2)/h$ to obtain

$$
\begin{aligned}
&= \frac{4(\kappa_*)^{d_X+d_Z}}{n^4 h^{d_X+d_Z}} \sum_{k,k'} \int (y_1^k - \mu^k(v_1))(y_2^k - \mu^k(v_1))(y_1^{k'} - \mu^{k'}(v_1))(y_2^{k'} - \mu^{k'}(v_1)) \\
&\qquad \times \Big(\frac{a(v_1)}{f(v_1)}\Big)^2 f(y_1, v_1) f(y_2, v_1)\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}v_1 (1 + O(h)) \\
&= \frac{4(\kappa_*)^{d_X+d_Z}}{n^4 h^{d_X+d_Z}} \sum_{k,k'} \int \Big(\int (y_1^k - \mu^k(v_1))(y_1^{k'} - \mu^{k'}(v_1)) \frac{f(y_1, v_1)}{f(v_1)}\,\mathrm{d}y_1\Big)^2 a(v_1)^2\,\mathrm{d}v_1 (1 + O(h)) \\
&= \frac{2}{n^4 h^{d_X+d_Z}} \Sigma_{\mathcal{K}}^2 (1 + O(h)).
\end{aligned}
$$

Similar calculations show that

$$
\begin{aligned}
\mathbb{E}\, h_n(W_1, W_2)^4 &= O(n^{-8} h^{-3(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_1, W_3)^2 &= O(n^{-8} h^{-2(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_1, W_3) h_n(W_2, W_3) &= O(n^{-8} h^{-2(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2) h_n(W_2, W_3) h_n(W_3, W_4) h_n(W_1, W_4) &= O(n^{-8} h^{-(d_X+d_Z)})
\end{aligned}
$$

Using some combinatorics we can establish from equation (7.12) that $\mathbb{E}\, I_{\mathcal{K}n,1}^4$ is asymptotically dominated by terms with $\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_3, W_4)^2 = (\mathbb{E}\, h_n(W_1, W_2)^2)^2$. Therefore the second condition in equation 7.11 is fulfilled as

$$
\frac{\mathbb{E}\, I_{\mathcal{K}n,1}^4}{(Var\,[I_n])^2} = \frac{12 n^{-4} h^{-2(d_X+d_Z)} \Sigma_{\mathcal{K}}^4 (1 + o(1))}{(2 n^{-2} h^{-(d_X+d_Z)} \Sigma_{\mathcal{K}}^2 (1 + o(1)))^2} \longrightarrow 3
$$

and weak convergence of $I_{\mathcal{K}n,1}$ is established.. *Q.E.D.*

Proof of (7.9) The expected value of the test statistic is given by

$$
\mathbb{E}\, I_{\mathcal{K}n,2} = \frac{1}{n} \sum_{k=1}^{d_Y} \iint \Big(K_h(v - v_1) \frac{y_1^k - \mu^k(v_1)}{f(v)}\Big)^2 a(v) f(v)\,\mathrm{d}v f(y_1, v_1)\,\mathrm{d}y_1\,\mathrm{d}v_1.
$$

Changing variables and expanding yields

$$
\begin{aligned}
&= \frac{\kappa_0^2}{n h^{d_X+d_Z}} \sum_{k=1}^{d_Y} \int (y_1^k - \mu^k(v_1))^2 \frac{a(v_1)}{f(v_1)} f(y_1, v_1)\,\mathrm{d}v_1 (1 + O(h)) \\
&= n^{-1} h^{-(d_X+d_Z)} B_{\mathcal{K}} (1 + O(h^r)).
\end{aligned}
$$

Stochastic Convergence follows from Markov's inequality with second moments, which requires to calculate

$$
\frac{1}{n^4} \Big(\int \sum_{k=1}^{d_Y} \big(K_h(v - v_1)(y_1^k - \mu^k(v_1))\big)^2 \frac{a(v)}{f(v)}\,\mathrm{d}v\Big)^2 f(y_1, v_1)\,\mathrm{d}y_1\,\mathrm{d}v_1.
$$

Changing variables as before results in

$$\frac{\kappa_0^2}{n^4 h^{2(d_X+d_Z)}} \sum_{k,k'} \int (y_1^k - \mu^k(v_1))^2 (y_1^{k'} - \mu^{k'}(v_1))^2 \frac{a(v_1)^2}{f(v_1)^2} f(y_1, v_1)\, \mathrm{d}y_1\, \mathrm{d}v_1 (1 + o(1))$$

which is bounded by assumption 8. In total this yields

$$\mathbb{E}\, I_{\mathcal{K}n,2}^2 = O(n^{-3} h^{-2(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$$

and stochastic convergence of $I_{\mathcal{K}n,2}$ follows.. *Q.E.D.*

Proof of (7.10) For this statement we will restrict to the case when $d_Y = 1$. Then stochastic convergence has to be shown for

$$\Delta_{\mathcal{K}n} = \frac{1}{n^3} \sum_{i,j,k} \gamma_n(W_i, W_j, W_k)$$

where

$$\gamma_n(W_i, W_j, W_k) = \tilde{\gamma}_n(W_i, W_j, W_k) - \int \tilde{\gamma}_n(W_i, W_j, w) f(w)\, \mathrm{d}w$$

with

$$\tilde{\gamma}_n(W_i, W_j, W_k) = K_h(V_k - V_i) \frac{Y_i - \mu^1(V_i)}{f(V_k)} a(V_k) K_h(V_k - V_j) \frac{Y_j - \mu^1(V_j)}{f(V_k)} a(V_k).$$

First we show that the expectation tends to zero

$$\mathbb{E}\, \Delta_{\mathcal{K}n} = \frac{1}{n^3} \sum_{i,j,k} \mathbb{E}\, \gamma_n(W_i, W_j, W_k) = o(n^{-1} h^{-(d_X+d_Z)/2})$$

where only the cases $i = k \neq j, j = k \neq i$ and $i = j = k$ have to be considered, all others have expectation zero. In the remaining cases, two (resp. one) substitution can be applied and their total contribution is $O(n^{-1} h^{2(d_X+d_Z)} + n^{-2} h^{d_X+d_Z})$.

To show stochastic convergence, Markov's inequality is applied with the second moments and we have to investigate

$$\mathbb{E}\, \Delta_{\mathcal{K}n}^2 = \frac{1}{n^6} \sum_{ijk} \mathbb{E}\, \gamma_n(W_i, W_j, W_k)^2 + \frac{2}{n^6} \sum_{ijk} \sum_{i'j'k'} \mathbb{E}\, \gamma_n(W_i, W_j, W_k) \gamma_n(W_{i'}, W_{j'}, W_{k'}).$$

The covariance parts vanish, whenever $k \neq k'$. If $k = k'$ the covariance terms are zero by the conditional independence of the error terms, in all cases where $i \neq i'$ or $j \neq j'$. For the remaining cases we have to distinguish if the number of different indices is $N = 2, 3$. Then, the overall contribution of these terms is $O(n^{N-6} h^{-4(d_X+d_Z)} h^{N(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Y)})$.

Next, consider the variance terms. If there are three different indices, two changes of variables can be applied and the overall contribution is $O(n^{-3} h^{-2(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$. If there are two different indices, one change of variables can be applied and we obtain terms of order $O(h^{-3(d_X+d_Z)})$ with a total contribution of $O(n^{-4} h^{-3(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$. If $i = j = k$ one change of variables is still possible and the contribution is $O(n^{-5} h^{-3(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$. This completes the proof of equation (7.10). *Q.E.D.*

Proof of Theorem 5

In the proof of this theorem we use the notation $\mathbb{E}^*$ (Var^* and $\mathbb{P}^*$) to denote expectation (variance and probability) conditional on the data. Decompose

$$\begin{aligned}
\hat{\Gamma}^*_{\mathcal{K}} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j, Z_i - Z_j)\frac{Y_j^{k,*} - \hat{m}^{k,*}_{\tilde{h}}(X_i)}{\hat{f}_h(X_j, Z_j)}\Big)^2 A_i \\
&= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j, Z_i - Z_j)\frac{\varepsilon^{k,*}}{\hat{f}_h(X_j, Z_j)}\Big)^2 A_i \\
&\quad + \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j, Z_i - Z_j)\frac{\hat{m}^{k}_{\tilde{h}}(X_i) - \hat{m}^{k,*}_{\tilde{h}}(X_i)}{\hat{f}_h(X_j, Z_j)}\Big)^2 A_i \\
&= (I^*_{\mathcal{K}n} + \Delta^*_{\mathcal{K}n})(1 + o_P(1)) + \Gamma^*_{\mathcal{K}3}
\end{aligned}$$

where $I^*_{\mathcal{K}n}$ and $\Delta^*_{\mathcal{K}n}$ are defined as in (7.5) and (7.6) by replacing $Y_j^k - \mu^k(X_j)$ with $\varepsilon_j^{k,*}$. $\Gamma^*_{\mathcal{K}3}$ can be bounded by showing that conditional on the data

$$\sup_{x\in\mathcal{A}} |\hat{m}^k_{\tilde{h}}(x) - \hat{m}^{k,*}_{\tilde{h}}(x)| = O_P\Big(\tilde{h}^r + \Big(\frac{\log n}{nh^{d_X}}\Big)^{1/2}\Big). \tag{7.14}$$

Decomposing $I^*_{\mathcal{K}n}$ as in equation (7.7) into $I^*_{\mathcal{K}n,1}$ and $I^*_{\mathcal{K}n,2}$ it remains to show that

$$nh^{(d_X+d_Z)/2} I^*_{\mathcal{K}n,1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2_{\mathcal{K}}) \tag{7.15}$$

$$nh^{(d_X+d_Z)/2} I^*_{\mathcal{K}n,2} - h^{-(d_X+d_Z)/2} B_{\mathcal{K}} \xrightarrow{P} 0 \tag{7.16}$$

$$nh^{(d_X+d_Z)/2} \Delta^*_{\mathcal{K}n} \xrightarrow{P} 0 \tag{7.17}$$

conditional on the data. Then the statement of the theorem follows.. *Q.E.D.*

Proof of (7.14) First note that

$$\begin{aligned}
\sup_{x\in\mathcal{A}} |\hat{m}^k_{\tilde{h}}(x) - \hat{m}^{k,*}_{\tilde{h}}(x)| &= \sup_{x\in\mathcal{A}} |(\hat{f}^k_{\tilde{h}}(x))^{-1}\frac{1}{n}\sum_{i=1}^{n} K_{\tilde{h}}(x - X_i)(Y_i^k - Y_i^{k,*})| \\
&\le \sup_{x\in\mathcal{A}} |m^k(x) - \hat{m}^k_{\tilde{h}}(x)| + \sup_{x\in\mathcal{A}} |(\hat{f}^k_{\tilde{h}}(x))^{-1}\sum_{i=1}^{n}\frac{1}{n}K_{\tilde{h}}(x - X_i)\varepsilon_i^{k,*}| \\
&\qquad + \sup_{x\in\mathcal{A}} |(\hat{f}^k_{\tilde{h}}(x))^{-1}\sum_{i=1}^{n}\frac{1}{n}K_{\tilde{h}}(x - X_i)\varepsilon_i^{k}|
\end{aligned}$$

The first term already has the desired rate. Because $\hat{f}^k_{\tilde{h}}(x)$ is consistent and $f(x)$ is bounded from below on $\mathcal{A}$ further analysis can be restricted to the numerator. Covering the compact set $\mathcal{A}$ with N cubes $\mathcal{A}_l = \{x \mid \|x - x_l\| < \eta_N\}, l = 1, \ldots, N, \eta_N = O(N^{-1/d_X})$ we write for the second term (the third is analyzed in the same fashion)

$$\sup_{x\in\mathcal{A}}\Big|\frac{1}{n}\sum_{i=1}^{n}K_{\tilde h}(x-X_i)\varepsilon_i^{k,*}\Big| \leq \max_{l}\Big|\frac{1}{n}\sum_{i=1}^{n}K_{\tilde h}(x_l-X_i)\varepsilon_i^{k,*}\Big| + \max_{l}\sup_{x\in\mathcal{A}_l}\Big|\frac{1}{n}\sum_{i=1}^{n}(K_{\tilde h}(x-X_i)-K_{\tilde h}(x_l-X_i))\varepsilon_i^{k,*}\Big|.$$

Using the Lipschitz-continuity of the kernel, one directly obtains that the second term is of $O_P(\eta_N n^{-1/2})$. The first term is bounded using Bonferroni's inequality first and then Bernstein's inequality

$$\mathbb{P}^*\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}K_{\tilde h}(x_l-X_i)\varepsilon_i^{k,*}\Big| > \Big(\frac{\log n}{n\tilde h^{d_X}}\Big)^{1/2}\frac{c}{2}\Big) \leq 2\exp\Big(-\frac{c^2(\log n)/(4n\tilde h^{d_X})}{4\sum_{i=1}^{n}\mathbb{E}^*(\frac{1}{n}K_{\tilde h}(x-X_i)\varepsilon_i^{k,*})^2+\tilde c\frac{(\log n)^{1/2}}{(n\tilde h^{d_X})^{3/2}}}\Big)$$

where $\tilde c$ is the constant arising from Cramer's conditions on the distribution of ε^*. It follows from standard arguments that $\sum_{i=1}^{n}\mathbb{E}^*(\frac{1}{n}(K_{\tilde h}(x-X_i)\varepsilon_i^{k,*})^2 = O_P(n^{-1}\tilde h^{-d_X})$ and so we get that

$$\mathbb{P}^*\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}K_{\tilde h}(x-X_i)\varepsilon_i^{k,*}\Big| > \Big(\frac{\log n}{n\tilde h^{d_X}}\Big)^{1/2}\frac{c}{2}\Big) \leq \frac{1}{n}O_P(1).$$

Then, for $N = o(n)$ the desired rate of convergence is obtained.. *Q.E.D.*

Proof of (7.15) To derive the asymptotic distribution of

$$I^*_{\mathcal{K}n,1} = \sum_{i<j} h_n(W_i^*, W_j^*)$$

given the data, again the central limit theorem by de Jong (1987) will be applied. This is done by showing that the conditions hold with probability tending to one, i.e.

$$\frac{\max_{1\leq i\leq n}\sum_{j=1}^{n}\mathbb{E}^*\, h_n(W_i^*,W_j^*)^2}{Var^*\left[I^*_{\mathcal{K}n1}\right]} \xrightarrow{P} 0 \qquad \frac{\mathbb{E}^*(I^*_{\mathcal{K}n,1})^4}{(Var^*\left[I^*_{\mathcal{K}n,1}\right])^2} \xrightarrow{P} 3$$

Here

$$h_n(W_i^*,W_j^*) = \frac{2}{n^2}\int K_h(v-V_i)K_h(v-V_j)\frac{a(v)}{f(v)}\,\mathrm{d}v\sum_{k=1}^{d_Y}\varepsilon_i^{k,*}\varepsilon_j^{k,*}$$

First we analyze

$$
\begin{aligned}
\mathbb{E}^* h_n(W_i^*, W_j^*)^2 &= \frac{4}{n^4}\Big(\int K_h(v-V_i)K_h(v-V_j)\frac{a(v)}{f(v)}\,\mathrm{d}v\Big)^2 \\
&\quad \times\Big(\sum_{k=1}^{d_Y}(Y_i^k - \hat m_{\tilde h}^k(X_i))(Y_j^k - \hat m_{\tilde h}^k(X_j))\Big)^2 \\
= \frac{4}{n^4}\Big(\int K_h(v-V_i)K_h(v-V_j)\frac{a(v)}{f(v)}\,\mathrm{d}v\Big)^2 & \\
\times\Big(\sum_{k=1}^{d_Y}(Y_i^k - \mu^k(V_i))(Y_j^k - \mu^k(V_j))\Big)^2 &\Big(1 + O_P\Big(\tilde h^r + \Big(\frac{\log n}{n\tilde h^{d_X}}\Big)^{1/2}\Big)\Big) \\
&= h_n(W_i, W_j)^2 + o_p(n^{-4}h^{-2(d_X+d_z)}).
\end{aligned}
\tag{7.18}
$$

This holds because under H_0 we have that $m^k(X_i) = \mu^k(X_i, Z_i)$ almost surely. Starting with the numerator, we utilize the conditional independence of the bootstrap residuals to see that

$$
Var^*\left[I_{\mathcal{K}n,1}\right] = \sum_{i<j}\mathbb{E}^* h_n(W_i^*, W_j^*).
$$

To bound this in probability, apply Markov's inequality with the first moment

$$
\mathbb{E}\Big|\sum_{i<j}\mathbb{E}^* h_n(W_i^*, W_j^*)^2\Big| = \sum_{i<j}\mathbb{E}\, h_n(W_i^*, W_j^*)^2 = n^{-2}h^{-(d_X+d_Z)}2\Sigma_{\mathcal{K}}^2(1+o(1))
$$

from which it follows that

$$
Var^*\left[I_{\mathcal{K}n,1}\right] \xrightarrow{P} Var\left[I_{\mathcal{K}n,1}\right].
$$

This is now used to show the first condition. Together with (7.11) and (7.18) we obtain

$$
\begin{aligned}
&\frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n \mathbb{E}^* h_n(W_i^*, W_j^*)^2}{Var\left[I_{\mathcal{K}n,1}\right]} \\
&\quad= \frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n h_n(W_i, W_j)^2 + O_P(n^{-3}(\tilde h^r + (\log n/(n\tilde h^{d_X}))^{1/2}))}{Var\left[I_{\mathcal{K}n,1}\right]} \\
&\quad= \frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n h_n(W_i, W_j)^2}{Var\left[I_{\mathcal{K}n,1}\right]} + O_P(n^{-1}(\tilde h^r + (\log n/(n\tilde h^{d_X}))^{1/2})) \\
&\quad= o_P(1)
\end{aligned}
$$

For the second condition we again use the convergence of the denominator. Then using the first moment to bound the probability leads to similar calculations as done in the proof of equation (7.8).. *Q.E.D.*

Proof of (7.16) The proof of equation (7.16) consists of using iterated expectations and use there the same calculations as to proof equation (7.15).. *Q.E.D.*

Proof of (7.17) As $\mathbb{E}^*\varepsilon_j^{k,*} = 0$ the same arguments as for $\Delta_{\mathcal{K}n}$ remain to hold for $\Delta_{\mathcal{K}n}^*$.. *Q.E.D.*