UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Working Paper Series No: 2015/17

NONPARAMETRIC EULER EQUATION IDENTIFICATION AND ESTIMATION

Juan Carlos Escanciano (Indiana University)
Stefan Hoderlein (University of Cambridge)
Arthur Lewbel (Boston College)
Oliver Linton (University of Cambridge)
Sorawoot Srisuma (University of Surrey)

We consider nonparametric identification and estimation of pricing kernels, or equivalently of marginal utility functions up to scale, in consumption based asset pricing Euler equations. Ours is the first paper to prove nonparametric identification of Euler equations under low level conditions (without imposing functional restrictions or just assuming completeness). We also propose a novel nonparametric estimator based on our identification analysis, which combines standard kernel estimation with the computation of a matrix eigenvector problem. Our estimator avoids the ill-posed inverse issues associated with existing nonparametric instrumental variables based Euler equation estimators. We derive limiting distributions for our estimator and for relevant associated functionals. We provide a Monte Carlo analysis and an empirical application to US household-level consumption data.

Nonparametric Euler Equation Identification and Estimation*

Juan Carlos Escanciano, *Indiana University*

Stefan Hoderlein, *Boston College*

Arthur Lewbel, *Boston College*

Oliver Linton, *University of Cambridge*

Sorawoot Srisuma, *University of Surrey*

September 24, 2015

Abstract

We consider nonparametric identification and estimation of pricing kernels, or equivalently of marginal utility functions up to scale, in consumption based asset pricing Euler equations. Ours is the first paper to prove nonparametric identification of Euler equations under low level conditions (without imposing functional restrictions or just assuming completeness). We also propose a novel nonparametric estimator based on our identification analysis, which combines standard kernel estimation with the computation of a matrix eigenvector problem. Our estimator avoids the ill-posed inverse issues associated with existing nonparametric instrumental variables based Euler equation estimators. We derive limiting distributions for our estimator and for relevant associated functionals. We provide a Monte Carlo analysis and an empirical application to US household-level consumption data.

JEL Codes: C14, D91,E21,G12. Keywords: Euler equations, marginal utility, pricing kernel, Fredholm equations, integral equations, nonparametric identification, asset pricing.

*We thank Don Andrews, Bob Becker, Xiaohong Chen, and seminar participants at University of Miami, UC San Diego, joint MIT-Harvard, Semiparametric Methods in Economics and Finance Workshop (London, 2010), AMES (Seoul, 2011), CFE (London, 2013) and the conference in honor of Don Andrews (Konstanz, 2015) for helpful comments. All errors are our own. This paper replaces "Nonparametric Euler Equation Identification and Estimation," by Lewbel, Linton and Srisuma (2011), and "Nonparametric Identification of Euler Equations," by Escanciano and Hoderlein (2012).

1 Introduction

The optimal intertemporal decision rule of an economic agent can often be characterized by first-order condition Euler equations. These equations are fundamental objects that appear in numerous branches of economics, in particular in the literatures on consumption, on savings and asset pricing, on labor supply, and on investment. Many empirical studies of dynamic optimization behaviors rely on the estimation of Euler equations. One of the original motivations of the generalized method of moments (GMM) estimator proposed by Hansen and Singleton (1982) was estimation of rational expectations based Euler equations associated with consumption based asset pricing models. In this paper we study the nonparametric identification and estimation of such Euler equations.

To fix ideas, consider a familiar consumption based asset pricing Euler equation (e.g. Cochrane 2001)

$$bE[g(C_{t+1}, V_{t+1})R_{t+1} \mid C_t, V_t] = g(C_t, V_t), \text{ almost surely (a.s.)} \tag{1}$$

where b is the subjective discount factor, C_t is consumption at time t, V_t is a vector of other economic variables such as durables or lagged consumption (for habits) that might affect utility, R_t is the gross return of an asset, and g is the time homogeneous marginal utility function of consumption. Equation (1) is the first order condition that equates in real terms the marginal cost of an extra unit of the asset, purchased today, to the expected marginal benefit of the extra payoff received tomorrow.[1]

Our work is the first to establish nonparametric point identification of the marginal utility function g, or equivalently of the pricing kernel function M (see below), under low level assumptions.[2] We also provide a novel nonparametric estimator based on this identification analysis, which combines standard kernel estimation with the computation of a matrix eigenvector problem. Our estimator overcomes the ill-posed inverse problem that affects existing nonparametric instrumental variables based estimators.

We take the primitives of the Euler equation to be the marginal utility function g, defined up to an arbitrary sign and scale normalization, and the discount factor b. The (nonparametric) *identified set* for the Euler equation is defined to be the set of all $(g, b) \in \Theta \equiv \mathcal{G} \times (0, 1)$, for a suitable parameter space $\mathcal{G}$, that satisfy equation (1), given the true joint distribution of the data (see Tamer 2010 for a review of set identification definitions). A model is defined to be globally identified if the identified set only consists of one element.

In this paper we first show that the Euler equation is partially identified, with a finite identified

[1] For a formal derivation of this Euler equation, with internal or external habits, see the Appendix.

[2] This paper is a merged revision of two earlier working papers: Lewbel, Linton and Srisuma (2011) and Escanciano and Hoderlein (2012). Some recent papers by others that establish related identification results cite these earlier versions of our paper as prior knowledge. See the next section for details.

set for the discount factor and an identified set for marginal utilities that is the union of finite dimensional spaces. This implies that the discount factor is also locally identified (in the sense of Fisher 1966, Rothenberg 1971 and Sargan 1983), meaning that b is nonparametrically identified within a parameter space that equals a neighborhood of the true value. We then show that if the class of utility functions is restricted to be monotone, which is a natural economic restriction, then the Euler equation model is, nonparametrically, globally point identified.

Having established identification, we next propose a novel nonparametric kernel estimator for the marginal utility function and discount factor based on our identification arguments. We provide asymptotic distribution theory for the discount factor, the marginal utility function, and for semiparametric functionals of the marginal utility function such as the Mean Relative Risk Aversion ($MRRA$) parameter defined below. We illustrate the applicability of our methods with US household-level data from the Consumer Expenditure Survey (CEX).

In the empirical asset pricing literature, the Euler equation (1) is traditionally written as

$$E\left[M_{t+1}R_{t+1} \mid C_t, V_t\right] \equiv E\left[b\frac{g(C_{t+1}, V_{t+1})}{g(C_t, V_t)}R_{t+1} \mid C_t, V_t\right] = 1,$$

where $M_{t+1} = bg(C_{t+1}, V_{t+1})/g(C_t, V_t)$ is the time $t+1$ pricing kernel or Stochastic Discount Factor (SDF). Then, the pricing equation for asset R can be cast in the form of excess returns

$$E\left[M_{t+1}\left(R_{t+1} - R_{0t+1}\right) \mid C_t, V_t\right] \equiv E\left[b\frac{g(C_{t+1}, V_{t+1})}{g(C_t, V_t)}\left(R_{t+1} - R_{0t+1}\right) \mid C_t, V_t\right] = 0, \tag{2}$$

where R_{0t} denotes the return from the risk-free asset. Equation (2) is a conditional moment restriction that forms the basis of moments based estimation. In a parametric model, g (and hence M_t) is assumed known up to finite-dimensional parameters; prominent examples include Hall (1978), Hansen and Singleton (1982), Dunn and Singleton (1986), and Campbell and Cochrane (1999), among many others. Euler equations have also been specified semiparametrically, e.g., Chen and Ludvigson (2009) and Chen, Chernozhukov, Lee and Newey (2014).

Nonparametric estimators of equation (2) and similar models (taking the form of nonparametric instrumental variables models) have been proposed, by, e.g., Gallant and Tauchen (1989), Chapman (1997), Newey and Powell (2003), Ai and Chen (2003) and Darolles, Fan, Florens, and Renault (2011). However, in these applications identification is assumed rather than proved, by way of high level completeness assumptions. These models have the structure of Fredholm equations of the first kind (also called Type I equations). Solving these types of equations involves ill-posed inverse problems that can be severe, and as a result, nonparametric estimators of M based on (2) can have very slow convergence rates and possibly unstable inference.

In contrast, we start by writing the pricing kernel problem in the form of equation (1) instead of equation (2), thereby estimating g instead of M. The advantage is that equation (1) takes the form of a Fredholm linear equation of the second kind (or Type II equation). As a result, unlike equation (2), the solution of equation (1) has a well-posed generalized inverse, leading to much better asymptotic properties for inference. In particular, in solving equation (1), a candidate discount factor b and associated marginal utility function g is characterized as an eigenvalue-eigenfunction pair of a certain conditional mean operator. Under the mild assumption that this operator is compact, a classical result (see e.g. Kress (1999)) ensures that the number of eigenvalues is countable. The behavioral restriction that $b < 1$ reduces this set to a finite number of pairs, leading to our finite set identification result and hence to local identification for the discount factor. To obtain global point identification of b and g, we impose the additional behavioral restriction that utility is increasing in consumption, which implies that the function g is positive. Applying an infinite-dimensional extension of the Perron-Frobenius theorem (see Kreĭn and Rutman 1950) yields uniqueness of a positive eigenvalue-eigenfunction pair, which then provides nonparametric point identification.

Following this identification argument, we propose a new nonparametric estimator for the marginal utility function g and discount factor b. The estimator is based on standard kernel estimation of a sample analogue of (1), which with finite data replaces the problem of solving for an eigenfunction with the simpler problem of solving for a standard finite-dimensional matrix eigenvector. No numerical integration or optimization is required, making the estimator straightforward to implement (and numerically practical to bootstrap). We establish our estimator's limiting distribution under standard conditions, which are simpler than those associated with estimators that solve related ill-posed inverse problems, such as nonparametric instrumental variables. Our expansions show that, in contrast to nonparametric problems leading to Type-I equations, nonparametric inference on g in our Type-II equation is to a large extent equivalent to inference on a standard conditional mean function, and in particular has comparable rates of convergence to ordinary nonparametric regression. Although our assumptions are standard, both our identification and asymptotic theory entail machinery that is novel in the econometrics literature, applying an infinite-dimensional extension of Perron-Frobenius theory to a type II Fredholm equation (see the next section for details comparing our results to the literature).

In addition to the pricing kernel M_{t+1}, another functional of the marginal utility function g that is of interest to estimate is the Arrow-Pratt coefficient of Relative Risk Aversion, and its mean value, RRA and $MRRA$, given respectively by

$$RRA(c,v) = \frac{-c\partial g(c,v)/\partial c}{g(c,v)} \quad \text{and} \quad MRRA = E\left[RRA(C_t, V_t)\right].$$

We illustrate the applicability of our asymptotic results by establishing asymptotic normality of a nonparametric estimator of the $MRRA$. Given our estimates of $g(c,v)$, we also provide tests of whether g is independent of v, thereby testing whether lagged consumption (or any other potential covariates v such as durables consumption) affects the pricing kernel. These tests are based on semiparametric functionals of g, which are asymptotically normal under the same type of regularity conditions we use to establish asymptotics for the $MRRA$.

One of the main motivations for estimating marginal utility nonparametrically is to look for evidence on whether common parametric or semiparametric alternatives are correctly specified, or whether there is some feature of the data that parametric models may have missed. In our empirical application, we compare our nonparametric estimates to the common Constant Relative Risk Aversion ($CRRA$) specification of utility, and find evidence against the $CRRA$ specification. More generally, we find evidence that the $MRRA$ is not constant, and thereby reject semiparametric models like that of Chen and Ludvigson (2009) and Chen, Chernozhukov, Lee and Newey (2014), which assume that RRA is constant (note, though, that they estimate their model with aggregate time series data while we use individual consumer level data). We also find some, albeit weaker, evidence that habits (lagged consumption) may affect utility in more complicated ways than previous models in the literature assume.

The rest of the paper is organized as follows. After a literature review in Section 2, we provide sufficient conditions for partial identification and point identification in Section 3. We propose our kernel-type estimator in Section 4, and we investigate its asymptotic properties in Section 5. In Section 6 we describe how our asymptotic theory applies to functionals of g, and give some examples. We report the results of a Monte Carlo experiment in Section 7. In Section 8, we apply our results to US household level consumption data. Section 9 concludes. An Appendix contains the derivation of the Euler equation, as well as mathematical proofs of the main results.

2 Literature Review

The forerunners of our research are the papers by Gallant and Tauchen (1989) and Chapman (1997), who estimate nonparametrically the marginal utilities and the pricing kernel, respectively, from the Euler equation by sieves, using the moment restriction (2) (i.e. using a Type I Fredholm equation). These papers did not investigate identification, nor impose the positivity of marginal utilities, and the asymptotic properties of their nonparametric estimators were not established.

Nonparametric instrumental variables is a leading example of estimation based on a Type I Fredholm equation, yielding associated ill-posed inverse problems on estimation. Newey and Powell (2003) note that assuming statistical completeness (a high level assumption) is essentially the same as just assuming identification of this type of model. Other related examples of nonparametric and semiparametric ill-posed inverse estimation problems include Carrasco and Florens (2000), Ai and Chen (2003), Hall and Horowitz (2005), Chen and Pouzo (2009), Chen and Reiss (2010), Darolles, Fan, Florens and Renault (2011) and, more recently, Cai, Ren and Sun (2015). A particularly relevant example is Chen and Ludvigson (2009), who studied identification and estimation of a semiparametric specification of the Type-I equation (2). Their model assumes g has the semiparametric form $g(C_t, V_t) = C_t^{\eta} h(V_t)$ (here η is a constant that determines risk aversion), where h is an unknown function of current and lagged values of C_t/C_{t-1} representing habits. Virtually all parametric estimators of the asset pricing model, going back to Hansen and Singleton (1982) and including Dunn and Singleton (1986), and Campbell and Cochrane (1999), use the form of equation (2) rather than equation (1).

Many parametric rational expectations models that focus on utility or production rather than asset pricing do estimation in the form of equation (1). Early examples include Hall (1978) and Mankiw (1982) (though see Lewbel 1987 for a critique). This earlier work does not appear to recognize the theoretical integral equation advantages of casting the model in the form of equation (1). Anatolyev (1999) recognizes that this form is a Type II Fredholm equation and provides a numerical method for estimating Euler equations that makes use of this structure, but he does not consider identification or inference. We believe our paper is the first to make explicit use of this Type II Fredholm structure for identification and inference. An and Hu (2012) exploit the nature of a type II Fredholm equation to identify and estimate a measurement error rather than an Euler equation model, but they cite our working paper as prior knowledge.

Our proof of global identification makes use of extensions of the classical Perron-Frobenius theorem that positive matrices have a unique positive eigenvalue that corresponds to a unique positive eigenvector. In particular, we apply a theorem of Kreĭn and Rutman (1950), which extends Perron-Frobenius to compact operators in Banach spaces. See, e.g., Schaefer (1974) and Abramovich and Aliprantis (2002) for details regarding this theory.

Versions of Perron-Frobenius have been used before in Euler equation models, though we believe we are the first to use this machinery of infinite-dimensional Perron-Frobenius theory for nonparametric identification and inference of Euler equations. There is, however, some closely related work. Ross (2015) applies the classical finite-dimensional Perron-Frobenius theorem to identify the pricing kernel and the natural probability distribution from state prices. Starting from the ill-posed inverse form of equation (2), Hansen and Scheinkman (2009, 2012, 2013) consider a different problem of identification than ours in a continuous-time setting, using Markov theory and extensions of the classical Perron-Frobenius theorem. In our notation, they give conditions for identification of the positive eigenfunction and eigenvalue of the operator $\phi \to E[M_{t+1}\phi(C_{t+1}, V_{t+1}) \mid C_t, V_t]$, assuming that the SDF M_{t+1} is known. In contrast, we solve the also fundamental problem of showing that M_{t+1} itself is identified, by obtaining identification of b and g. Christensen (2014, 2015) applies identification results, based in part on our earlier working papers, to a discrete version of Hansen and Scheinkman (2009).

Perhaps the closest work to ours is Chen, Chernozhukov, Lee and Newey (2014). Although their paper mainly concerns local nonparametric identification, in their Euler equation application they consider a semiparametric rather than a nonparametric model like ours. Specifically, their model is the same functional form as Chen and Ludvigson (2009) described above, but allowing for a more general conditioning set. They cite the working paper versions of our paper as prior knowledge, making similar use both of well-posedness and of extended Perron-Frobenius theory. Their general theory imposes restrictions on the marginal utility. These restrictions assume a semiparametric $CRRA$ functional form, that is, their model assumes the RRA is both constant and identified, and given that assumption, they identify the role of habits. In contrast, our resuls including proving that both the role of habits and the RRA (whether constant or not) are both nonparametrically identified, and we provide inference tools to test if the RRA constant.

An alternative to our kernel based estimation would be the use of sieves. Nonparametric sieve estimation of eigenvalue-eigenvector problems for self-adjoint operators is extensively discussed in Chen, Hansen and Sheinkman (2000, 2009), Darolles, Florens and Gouriéroux (2004) and Carrasco, Florens and Renault (2007), among others. However, their results cannot be applied to our model, since in our case the associated operator is not self-adjoint. Christensen (2014) (who cites our earlier working paper version) proposes a nonparametric sieve estimator for the discrete Markov setting of Hansen and Scheinkman (2009), establishing asymptotic normality of the eigenvalue estimate and smooth functionals of it. See also Gobet, Hoffmann and Reiss (2004) for sieve estimation of eigenelements in diffusion models. As noted earlier, sieve estimation has more directly been applied to nonparametric and semiparametric versions of equation (2) going back to Gallant and Tauchen (1989). In comparison, our kernel based estimator has numerous advantages as summarized in the previous section, mainly attributable to our method of exploiting well-posedness of equation (1).

Our empirical application uses household level consumption data, and in particular considers the possible presence and role of habits, that is, lagged consumption. A large literature focuses on individual level consumption smoothing implied by equation (1), and potential sources of violations of the model, even after controlling for durables or habits. Example of possible violations include liquidity constraints and precautionary savings (see, e.g., Deaton 1992 and references therein) and the so-called consumption retirement puzzle (see, e.g., Banks, Blundell, and Tanner 1998). Also relevant is the implied impact of this model on consumption distributions. See, e.g., Deaton and Paxson (1994), Lewbel (1994), and Battistin, Blundell, and Lewbel (2009). Within these literatures, of particular relevance for our empirical application are earlier studies on individual heterogeneity of risk aversion in consumption choice, and the role of habits. For a recent summary see Gayle and Khorunzhina (2014) and references therein. Virtually all of this literature imposes parametric or strong semiparametric restrictions on g, and so, like the earlier aggregate consumption models of Hall (1978), Mankiw (1982), Hansen and Singleton (1982), or Campbell and Cochrane (1999), does not exploit the theoretical advantages of having equation (1) be type II Fredholm.

3 Identification

Since our goal is the study of Euler equations, we shall take as primitives the pair $(g, b) \in \Theta \equiv \mathcal{G} \times (0,1)$, where $\mathcal{G}$ denotes the parameter space of marginal utility functions, which satisfies some conditions below. From equation (1) it is clear that, for a given b, the Euler equation cannot distinguish between g and g' if there exists some constant $k_0 \in \mathbb{R}$ such that $g = k_0 g'$ a.s., so a scale and a sign normalization must be made. For the moment we shall assume there is just one asset, and we denote its rate of return by R_t. We later discuss how information from multiple assets can be used to aid identification. As seen in the previous section, for each period t, C_t is consumption and V_t is (possibly a vector of) other economic variable(s).

Let $S \subseteq \mathbb{R}^\ell$ denote the support of (C_t, V_t). Let (S, μ) be a σ-finite measure space, and let $\mathcal{L}^2$ denote the Hilbert space $L_2(S, \mu)$ of (equivalence classes of) square μ-integrable functions equipped with the inner product $\langle g, f \rangle = \int g f d\mu$ and the corresponding norm $\|g\|^2 = \langle g, g \rangle$ (we drop the domain of integration for simplicity of exposition). Our identification and estimation results are valid for a generic μ, as long as some conditions below are satisfied, but for concreteness and simplicity of implementation, we choose as μ the probability measure of (C_t, V_t) for estimation purposes.

Let $\mathcal{M}$ be a linear subspace of $\mathcal{L}^2$, and define the linear operator $A : (\mathcal{M}, \|\cdot\|) \to (\mathcal{M}, \|\cdot\|)$ given by

$$Ag(c, v) = E[g(C_{t+1}, V_{t+1}) R_{t+1} \mid C_t = c, V_t = v]. \tag{3}$$

The space $\mathcal{M}$ is chosen so that Ag is well-defined and $Ag \in \mathcal{M}$ for $g \in \mathcal{M}$. The requirement $\mathcal{M} \subset \mathcal{L}^2$ can be relaxed (see Escanciano and Hoderlein, 2012, Section 4) but it is made here for simplicity, and is unlikely to be violated in empirical application. We provide below an example of $\mathcal{M}$ for which our conditions are easily verifiable. With our notation, (1) can be written in a compact form as $bAg = g$. The parameter space for g, $\mathcal{G}$, will be a subset of $\mathcal{M}$ incorporating normalization restrictions. We introduce the assumption of correct specification and a formal definition of identification.

Assumption S. *There exists $(g, b) \in \Theta \equiv \mathcal{G} \times (0,1)$, $g \neq 0$, satisfying equation (1).*

Definition 1. *Given the joint distribution of $(R_{t+1}, C_{t+1}, V_{t+1}, C_t, V_t)$, the Euler equation is **nonparametrically identified** if there is a unique $(g, b) \in \Theta$ that satisfies equation (1). When the solution is unique we denoted it by $\theta_0 \equiv (g_0, b_0)$.*

Definition 2. *Given the joint distribution of* $(R_{t+1}, C_{t+1}, V_{t+1}, C_t, V_t)$, *the* ***identified set****, denoted by* Θ_0, *consists of elements in* Θ *where each* $(g, b) \in \Theta_0$ *satisfies equation (1). The sets* $B_0 = \{b \in (0,1) : \text{there is } g \in \mathcal{G} \text{ such that } (g,b) \in \Theta_0\}$ *and* $\mathcal{G}_0 = \{g \in \mathcal{G} : \text{there is } b \in (0,1) \text{ such that } (g,b) \in \Theta_0\}$ *are, respectively, the identified sets for* b *and* g.

Therefore the Euler equation is point identified, if Θ_0 is a singleton. To provide some insights on our identification and estimation strategies we consider first the case where A in (3) has a finite-dimensional range. In what follows let $\mathcal{R}(\cdot)$ denote the range of an operator, so that $\mathcal{R}(A) = \{f \in \mathcal{M} : \exists g \in \mathcal{M},\ Ag = f\}$. In this case, we can write

$$Ag\,(\cdot) = \sum_{i=1}^{I} L_i(g)\phi_i(\cdot), \tag{4}$$

for a set of functions $\{\phi_i\}$ that span $\mathcal{R}(A)$ and linear operators $L_i(g)$, $i = 1, \ldots, I$. This case arises, for example, when the support S is discrete and finite. Under (4), any potential solution of (1) has to have necessarily the form $g\,(\cdot) = \sum_{i=1}^{I} \beta_i\phi_i(\cdot)$ for a vector $\beta = (\beta_1, \ldots, \beta_I)$ satisfying the Euler equation

$$\sum_{i=1}^{I}\sum_{j=1}^{I} L_i(\phi_j)\beta_j\phi_i(c,v) = b^{-1}\sum_{i=1}^{I}\beta_i\phi_i(c,v).$$

In turn, this is the case for the solution, provided it exists, of

$$\sum_{j=1}^{I}\beta_j L_i(\phi_j) = b^{-1}\beta_i \qquad 1 \le i \le I.$$

Therefore, β, i.e. g, and b^{-1} are identified as any eigenelement of the $I \times I$ matrix $(L_i(\phi_j))_{i,j}$, with $b \in (0,1)$. In general, we may have more than one such eigenelement, i.e., we may have partial identification. In any case, the number of eigenvectors β and eigenvalues is bounded by I, so we have a finite identified set.

The previous arguments extend to the general case replacing the finite-dimensionality of $\mathcal{R}(A)$ by the compactness of A. A linear operator A is compact if it transforms bounded sets into relatively compact sets (relatively compact sets in $\mathcal{M}$ are those whose closure its compact). The compactness assumption is not needed just for identification, but is useful for obtaining asymptotics of continuous functionals of g. Note, however, that compactness rules out the case $\mathcal{M} = \mathcal{L}^2$ if there are overlapping elements in (C_{t+1}, V_{t+1}) and (C_t, V_t); see Carrasco, Florens and Renault (2007, Example 2.5, pg. 22). We could deal with the lack of compactness of A on the whole $\mathcal{L}^2$ by conditioning on (i.e. fixing) the overlapping components (see e.g. Blundell, Chen and Kristensen, 2007, pg. 1629). From the identification point of view there is little loss of generality by following this "conditioning" approach, however, for deriving asymptotics compactness is very convenient, since it guarantees that inference will be based on well-posed generalized inverses (see the discussion at the end of this section). Lemma 1 in Section 5 below provides sufficient lower level conditions for compactness of A, but for now we maintain compactness as a high level assumption.

Assumption C. *$A : (\mathcal{M}, \|\cdot\|) \to (\mathcal{M}, \|\cdot\|)$ is a compact operator.*

Let $\mathcal{G} = \{g \in \mathcal{M} : \|g\| = 1,\ g(c_0, v_0) > 0,\ (c_0, v_0) \in S\}$ be the parameter space for g.

Theorem 1. *Suppose that Assumptions S and C hold. Then, B_0 is a finite set and $\mathcal{G}_0$ is the union of finite dimensional subsets.*

Theorem 1 shows that the Euler equation is partially identified, with b identified up to a finite set corresponding to eigenvalues, and g is identified up to a corresponding set of eigenfunctions. The discount factor b is also *locally identified*, meaning that for any $b \in B_0$ there is an open neighborhood of b that does not contain any other element in B_0. Essentially, compactness of A ensures that B_0 is at most countable, and the economic restriction that discount factors lie in $(0, 1)$ ensures that B_0 is finite.

The identified set without additional economic restrictions can be further reduced if there are multiple assets. If there are J assets, then there are J Euler equations. Applying Theorem 1 to each asset, gives an identified set for each, and the true (g, b) must lie in the intersection of these identified sets. One might further shrink the identified set by imposing the restriction that $bg(C_{t+1}, V_{t+1})R_{t+1} - g(C_t, V_t)$ is uncorrelated with all variables in the information set at time t, not just (C_t, V_t).

Assumptions S and C do not suffice for point identification in general. We consider now a shape restriction on marginal utilities, which is a common behavioral assumption that is satisfied for common parametric specifications of utility. Specifically, we impose the assumption that that marginal utilities are positive. Let

$$\mathcal{P} \equiv \{g \in \mathcal{M} : g \geq 0 \quad \mu - a.s.\} \tag{5}$$

denote the subset of nonnegative functions in $\mathcal{M}$, and let $\mathcal{P}^+ \equiv \{g \in \mathcal{M} : g > 0\ \mu\text{–a.s.}\}$ denote the subset of strictly positive functions, which is assumed to be non-empty. The assumption is then:

Assumption I. *$Ag \in \mathcal{P}^+$ when $g \in \mathcal{P}$ and $g \neq 0$.*

Assumption I is a mild condition that extends the classical assumption of a positive matrix in the Perron-Frobenius theorem to an infinite-dimensional setting, see Abramovich and Aliprantis (2002, Chapter 9) and Schaefer (1974). A sufficient and mild condition for it is that the conditional expected (gross) return is strictly positive, i.e. $E\left[R_{t+1}|C_{t+1}=\cdot,V_{t+1}=\cdot,C_t=\cdot,V_t=\cdot\right]>0$ a.s. With our shape and normalization restrictions the parameter space is $\mathcal{G}=\{g\in\mathcal{P}:\|g\|=1\}$.

THEOREM 2. *Let Assumptions S, C and I hold. Then,* $(g,b)\in\mathcal{G}\times(0,1)$ *is point identified.*

Identification can be established under weaker conditions than those of Theorem 2, however, we do not pursue these conditions here because the stronger conditions of Theorem 2 will facilitate our later asymptotic inference results. These weaker conditions are evident from our proof of Theorem 2, which also shows that $b=1/\rho(A)$, where $\rho(A)$ is the spectral radius of A (see the Appendix for a definition of the spectral radius of a linear bounded operator). Following Escanciano and Hoderlein (2012) a key sufficient condition for identification of g is that A is irreducible; see Abramovich and Aliprantis (2002, Chapter 9) for a definition of irreducibility in a general setting. Assumption I is a sufficient but not necessary condition for irreducibility (cf. Abramovich and Aliprantis, 2002, Theorem 9.6).

We could consider other sufficient conditions that replace conditions on A by conditions on a power of A, i.e. we could require that Assumptions C and I hold for A^n, for some $n\geq 1$. It is hard to interpret these conditions, however, in a possibly non-Markovian environment, so we do not pursue them here. The identification result in Theorem 2 suggests It is also likely that the Euler Equation is overidentified under the conditions of Theorem 2, since as noted earlier we could exploit additional information coming from multiple assets, or from uncorrelatedness with other data in the information set at time t.

We close our study of identification with a discussion on the degree of ill-posedness of our nonparametric problem. Assumption S implies that the operator $L=bA-I$ is not one-to-one, as the marginal utility g satisfies $Lg=0$, and $g\neq 0$. Therefore, solving the Euler equation (1) is an ill-posed problem (see e.g. Carrasco, Florens and Renault 2007, Section 7). However, even though our problem is ill-posed, unlike in ill-posed Type-I equations, the ill-posedness in our Type-II equation is moderate, with stable solutions. Formally, the operator L, although not invertible, has a continuous (i.e. bounded) Moore-Penrose pseudoinverse, which is denoted by $L^{\dagger}$; (see Engl, Hanke and Neubauer 1996, p. 33). To see this, note that the compactness of A and the Second Riesz Theorem, see e.g. Theorem 3.2 in Kress (1999, p. 29), imply that the range of L, $\mathcal{R}(L)=\{f\in\mathcal{L}^2:\exists s\in\mathcal{L}^2,\, Ls=f\}$, is closed. This in turn implies that $L^{\dagger}$ is a continuous operator by Proposition 2.4 in Engl et al. (1996). It is in this precise sense that our problem leads to well-posed rather than ill-posed generalized inverses. This property of our nonparametric problem, which results from considering Type-II equations rather than Type-I equations, has important implications for inference. For example, in the next sections we obtain rates of convergence for estimation of g that are the same as those of ordinary nonparametric regression.

4 Estimation from Individual level-data

Our estimation strategy follows the identification strategy described above, and is also motivated from our empirical application below. For estimation we assume that we have a random sample of household-level data $\{(R_{t_i+1}, C_{t_i+1,i}, V_{t_i+1,i}, C_{t_i,i}, V_{t_i,i})\}_{i=1}^n$ for n households, with possibly overlapping non-decreasing time periods $t_1 \leq t_2 \leq \cdots \leq t_n$. To simplify notation denote $W_i = (R_i', C_i', V_i', C_i, V_i) \equiv (R_{t_i+1}, C_{t_i+1,i}, V_{t_i+1,i}, C_{t_i,i}, V_{t_i,i})$, where $V_i = (V_{1i}, \ldots, V_{\ell_1 i})$ and $V_i' = (V_{1i}', \ldots, V_{\ell_1 i}')$ with $\ell = \ell_1 + 1$. We assume the data, $\{W_i\}_{i=1}^n$, are independent and identically distributed (iid), generated with respect to an underlying parameter $\theta_0 \equiv (g_0, b_0) \in \Theta$. We shall henceforth assume that Assumptions S, C and I hold, so that θ_0 is point-identified. Particularly, we consider $g_0 \in \mathcal{G} = \{g \in \mathcal{P} : \|g\| = 1\}$.

Let the vector $W = (R', C', V', C, V)$ have the same distribution as $(R_i', C_i', V_i', C_i, V_i)$. We assume that the vector W is continuously distributed (the discrete case is simpler). We denote the Lebesgue density of (C, V) by f. We consider the setting described in the identification section where μ is the joint probability associated to f. Henceforth, g and b denote generic elements in $\mathcal{G}$ and $(0,1)$, respectively.

Define the Nadaraya-Watson (NW) kernel estimator of the operator A at g as follows,

$$\widehat{A}g\,(c,v) = \frac{1}{n}\sum_{i=1}^n g_i' R_i' \phi_i(c,v),$$

where, for $i = 1, \ldots, n$, $g_i' \equiv g\,(C_i', V_i')$, $\phi_i(c,v) = K_{hi}\,(c,v)\,/\widehat{f}\,(c,v)$, while for $v = (v_1, \ldots, v_{\ell_1})$,

$$\widehat{f}\,(c,v) = \frac{1}{n}\sum_{i=1}^n K_{hi}\,(c,v)\,,$$

and

$$K_{hi}\,(c,v) = h^{-\ell}\; K\left(\frac{c - C_i}{h}\right)\prod_{j=1}^{\ell_1} K\left(\frac{v_j - V_{ji}}{h}\right).$$

Here, K is a univariate kernel function and $h \equiv h_n$ is a possibly stochastic bandwidth. Note that contrary to A, the operator $\widehat{A}$ has a finite-dimensional closed range (that is spanned by the functions $\phi_i(c,v)$, $i = 1, \ldots, n$). Therefore, similar to our discussion of identification in Section 3, the number of eigenvalues and eigenfunctions of $\widehat{A}$ is finite and bounded by n, and they can be computed by solving a linear system. Indeed, any eigenfunction $\widehat{g}(c,v)$ of $\widehat{A}$ necessarily has the form $n^{-1}\sum_{i=1}^n \widehat{\beta}_i \phi_i(c,v)$, for some coefficients $\widehat{\beta}_i$, $i = 1, \ldots, n$, satisfying for its corresponding eigenvalue $\widehat{\lambda}$ the equation

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \widehat{\beta}_j \phi_j(C_i', V_i') R_i' \phi_i(c,v) = \widehat{\lambda}\frac{1}{n}\sum_{i=1}^n \widehat{\beta}_i \phi_i(c,v).$$

A solution to this eigenvalue problem exists if, for all $i = 1, \ldots, n$,

$$\frac{1}{n} \sum_{j=1}^{n} \widehat{\beta}_j \phi_j(C_i', V_i') R_i' = \widehat{\lambda} \widehat{\beta}_i,$$

which in matrix notation can be written as

$$\widehat{A}_n \widehat{\beta} = \widehat{\lambda} \widehat{\beta},$$

where $\widehat{A}_n$ is an $n \times n$ matrix with ij-th element $a_{ij} = \phi_j(C_i', V_i') R_i'/n$, and $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^\intercal$ (henceforth, $v^\intercal$ denotes the transpose of v). Thus, let $\widehat{\lambda}$ denote the largest eigenvalue in modulus of $\widehat{A}_n$ and $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^\intercal$ its corresponding eigenvector. The eigenvector $\widehat{\beta}$ is normalized so that $\widehat{\beta}^\intercal \widehat{\Omega} \widehat{\beta} = 1$, where $\widehat{\Omega}$ is the $n \times n$ matrix with entries

$$\omega_{ij} = \frac{1}{n^3} \sum_{l=1}^{n} \phi_i(C_l, V_l) \phi_j(C_l, V_l),$$

and $n^{-1} \sum_{i=1}^{n} \widehat{\beta}_i \phi_i(c_0, v_0) > 0$, for some $(c_0, v_0) \in S$. We define the estimators for b_0 and g_0 respectively as follows,

$$\hat{b} = 1/\widehat{\lambda} \qquad \text{and} \qquad \widehat{g}\,(c, v) = n^{-1} \sum_{i=1}^{n} \widehat{\beta}_i \phi_i(c, v), \tag{6}$$

where $\widehat{g}$ satisfies $\|\widehat{g}\|_n = 1$ by the normalization of $\widehat{\beta}$ above, with $\|g\|_n$ denoting the empirical norm of g, i.e. $\|g\|_n^2 = \sum_{i=1}^{n} g^2(C_i, V_i)/n$. The estimator $(\widehat{g}, \hat{b})$ can be easily obtained with any statistical package that computes eigenvalues and eigenvectors of matrices. There are also efficient algorithms for the computation of the so-called Perron-Frobenius root $\widehat{\lambda}$, see e.g. Chanchana (2007).

Notice that under very mild conditions the matrix $\widehat{A}_n$ itself satisfies the classic conditions of the Perron-Frobenius theorem, which guarantees that $\hat{b} = \rho^{-1}(\widehat{A}_n)$ and $\widehat{\beta}$ is the only eigenvector of $\widehat{A}_n$ with positive entries. That is, in this case we also have identification in finite samples. For example, for strictly positive kernels and strictly positive gross returns, $\widehat{A}_n$ has strictly positive entries, which then implies a positive estimator $\widehat{g}\,(c, v) > 0$ and a positive discount factor $\hat{b}$ with probability one for a fixed $n \geq 1$.

5 Asymptotic Theory

In this section we provide conditions for the consistency and limiting distribution theory of our estimators as defined in the previous section, under a random sampling framework.[3] We need to

[3]We consider the random sampling iid framework to be a good approximation for our household-level data. The proofs in the Appendix could be straightforwardly adapted to allow for weakly dependent data using the uniform rate results of Andrews (1995).

introduce some notation from empirical processes theory. To measure the complexity of the class $\mathcal{G}$, we can employ covering or bracketing numbers. Here, for simplicity, we focus on bracketing numbers. Given two functions l, u, a bracket $[l, u]$ is the set of functions $f \in \mathcal{G}$ such that $l \leq f \leq u$. An ε-bracket with respect to $\|\cdot\|$ is a bracket $[l, u]$ with $\|l - u\| \leq \varepsilon$, $\|l\| < \infty$ and $\|u\| < \infty$ (note that u and l not need to be in $\mathcal{G}$). The *covering number with bracketing* $N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|)$ is the minimal number of ε-brackets with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. An envelope for $\mathcal{G}$ is a function G, such that $G(c, v) \geq \sup_{g \in \mathcal{G}} |g(c, v)|$ for all (c, v). To simplify notation, we use the following definition. Denote by $\mathcal{K}(r)$ the class of r-order kernels K that are Lipschitz continuous on the support $[-1, 1]$, symmetric, integrate to one, and such that for some $r \geq 2$: $\int u^l K(u)\, du = \delta_{l0}$ for $l = 0, \ldots, r-1$, where $\delta_{ll'}$ denotes Kronecker's delta, and $\int u^r K(u)\, du > 0$.

Assumption A1:

1. *$P(\langle \widehat{g}, g_0 \rangle > 0) \to 1$ as $n \to \infty$.*

2. *For each $\varepsilon > 0$, $\log N_{[\cdot]}(\varepsilon, \mathcal{M}, \|\cdot\|) \leq C\varepsilon^{-v}$ for some $v < 2$. The class $\mathcal{G}$ is such that $g_0 \in \mathcal{G}$ and has an envelope G such that $\sup_{(c,v) \in S} E[|G(C', V')R'|^{\delta} \, |C = c, V = v] < \infty$ for some $\delta > 2$. Functions in $\mathcal{R}(A)$ are uniformly equicontinuous on S.*[4]

3. *There exists a convex and compact subset T contained in the interior of S, such that $P((C', V') \in T | (C, V) \in T) = 1$. The density function $f(\cdot)$ is bounded away from zero on T and is continuous on S.*

4. *$K \in \mathcal{K}(2)$.*

5. *As $n \to \infty$, the possibly stochastic bandwidth $h \equiv h_n$ satisfies $P(l_n \leq h_n \leq u_n) \to 1$ for deterministic sequences of positive numbers l_n and u_n such that: $u_n \to 0$ and $l_n^{\delta\ell/(\delta-2)} n / \log n \to \infty$.*

Condition A1.1 is a suitable sign normalization condition in our $\mathcal{L}^2$-setting. This is a mild condition which is guaranteed to hold if, for instance, the kernel and the gross returns are strictly positive, since then $\widehat{g}$ and g_0 are strictly positive.

Condition A1.2 requires existence of certain moments. Marginal utilities may not have finite moments around zero (where they may diverge). To overcome this problem, by suitable redefinition

[4] That is,

$$\lim_{\delta \to 0} \sup_{|(c,v)-(c',v')|<\delta} \sup_{g \in \mathcal{M}} \|\mathbb{E}[g(C', V')R'|C = c, V = v] - \mathbb{E}[g(C', V')R'|C = c', V = v']\| = 0.$$

of g we can rewrite equation (1) in the form

$$bE[C'g(C',V')\,(C/C')\,R' \mid C,V] = Cg(C,V). \tag{7}$$

This reparameterizes the problem in terms of $Cg(C,V)$, which under natural economic assumptions is bounded; see Lucas (1978). This identity also gives an alternative way to estimate the marginal utility function and other objects of interest, which we shall discuss further below. Examples of classes satisfying A1.2 abound in the literature; see van der Vaart and Wellner (1996). For example, we could take $\mathcal{M}$ as the following smooth class: For any vector a of ℓ integers define the differential operator $\partial_x^a \equiv \partial^{|a|_1}/\partial x_1^{a_1} \ldots \partial x_\ell^{a_\ell}$, where $|a|_1 \equiv \sum_{i=1}^{\ell} a_i$. For any smooth function $h: T \subset \mathbb{R}^\ell \to \mathbb{R}$ and some $\eta > 0$, let $\underline{\eta}$ be the largest integer smaller or equal than η, and

$$\|h\|_{\infty,\eta} \equiv \max_{|a|_1 \leq \underline{\eta}} \sup_{x \in T} |\partial_x^a h(x)| + \max_{|a|_1 = \underline{\eta}} \sup_{x \neq x'} \frac{|\partial_x^a h(x) - \partial_x^a h(x')|}{|x - x'|^{\eta - \underline{\eta}}}.$$

Further, let $\mathcal{C}_M^\eta(T)$ be the set of all continuous functions $h: T \subset \mathbb{R}^\ell \to \mathbb{R}$ with $\|h\|_{\infty,\eta} \leq M$ (for an integer η, the η-th derivative is assumed to be continuous). Since the constant M is irrelevant for our results, we drop the dependence on M and denote $\mathcal{C}^\eta(T)$. Then, it is known that $\log N_{[\cdot]}(\varepsilon, \mathcal{C}^\eta(T), \|\cdot\|) \leq C\varepsilon^{-v_s}$, $v_s = \ell/\eta$, so if $\mathcal{M} \subset \mathcal{C}^\eta(T)$, then $\ell < 2\eta$ suffices for the bracketing condition in A1.2. We also have that $\mathcal{M} \subset \mathcal{L}^2$ here. With some smoothness conditions on the density of W, $\mathcal{R}(A) \subset \mathcal{M}$ holds with $\mathcal{M} = \mathcal{C}^\eta(T)$. Condition A1.2 is used here to control the term $\sup_{g \in \mathcal{G}} ||\widehat{A}g - Ag||$ and also to guarantee that A is compact, as the following result shows.

Lemma 1. *Suppose that Assumption A1.2 holds. Then A is compact.*

Under Assumption A1 we can write (a.s.)

$$bE[g(C',V')1((C',V') \in T)1((C,V) \in T)R' \mid C,V] = g(C,V)1((C,V) \in T),$$

and hence, we can restrict the domain of g to T. We therefore, hereafter restrict the support of μ to T (and thus, of the associated norm $\|\cdot\|$). The assumption of densities bounded away from zero is standard in the nonparametric and semiparametric literatures, though it could be relaxed here at the cost of longer proofs by introducing a vanishing random trimming parameter. See e.g. Escanciano, Jacho-Chavez and Lewbel (2014).

The remaining conditions in Assumption A1 are self-explanatory. For A1.4 we can also use kernels with unbounded support that satisfy some smoothness and integrability conditions. Finally, note that A1.5 allows for data-driven bandwidth choices, which are common in applied work. Our next result shows the $\mathcal{L}^2$-consistency of our estimators.

Theorem 3. *Let Assumptions S, C, I and A1 hold. Then, $\hat{b} \to_p b_0$ and $\|\widehat{g} - g_0\| \to_p 0$.*

To obtain asymptotic distribution theory for our estimators, we impose the following additional assumptions and notation. Simple algebra shows that the adjoint operator of A, that is, the linear compact operator A^* such that $\langle Ag_1, g_2\rangle = \langle g_1, A^* g_2\rangle$ for all $g_1, g_2 \in \mathcal{M}$, is given by $A^*\varphi(c', v') = E\left[\varphi\left(C, V\right) R' | C' = c', V' = v'\right] \times f'(c', v')/f(c', v')$, where $f'(c', v')$ denotes the Lebesgue density of (C_i', V_i'). To see this, note that by the Law of Iterated Expectations, for any $g_1, g_2 \in \mathcal{M}$,

$$\begin{aligned} \langle Ag_1, g_2\rangle &= E\left[E\left[g_1\left(C_i', V_i'\right) R_i' | C_i, V_i\right] g_2\left(C_i, V_i\right)\right] \\ &= E\left[g_1\left(C_i', V_i'\right) g_2\left(C_i, V_i\right) R_i'\right] \\ &= E\left[g_1\left(C_i', V_i'\right) E\left[g_2\left(C_i, V_i\right) R_i' | C_i', V_i'\right]\right] \\ &= \langle g_1, A^* g_2\rangle . \end{aligned}$$

Note that b_0^{-1} is also an eigenvalue for A^*; eigenvalues of A^* are complex conjugates of those of A. Similarly as we did for g_0, it can be shown that under Assumption A1 there exists a unique (up to scale) strictly positive eigenfunction of A^* associated to b_0^{-1}.

Definition 3. *Let s be the unique strictly positive eigenfunction of A^* with eigenvalue b_0^{-1} and satisfying the normalization $\langle g_0, s\rangle = 1$.*

The function s plays an important role in the asymptotics for $\hat{b}$ and $\widehat{g}$, as does the error term

$$\varepsilon_i = g_0\left(C_i', V_i'\right) R_i' - b_0^{-1} g_0\left(C_i, V_i\right), \qquad i = 1, \ldots, n. \tag{8}$$

Henceforth, to simplify notation, define $\varphi_i = \varphi\left(C_i, V_i\right)$ for any $\varphi \in \mathcal{L}^2$. For asymptotic normality of our estimators we require the following assumption.

Assumption A2.

1. *$f \in \mathcal{C}^r(T)$, where r as in A2.4 below.*

2. *Functions in $\mathcal{R}\left(A\right)$ are in $\mathcal{C}^r(T)$ with uniformly equicontinuous $r - th$ derivative on T.*

3. *$s \in \mathcal{C}^r(T)$ and $\Sigma_s \equiv E\left[s_i^2 \varepsilon_i^2\right] < \infty$.*

4. *$K \in \mathcal{K}(r)$, for $r \geq 2$.*

5. *For l_n and u_n satisfying A1.5, it also holds that $l_n^{2\ell} n/\log n \to \infty$ and $n u_n^{2r} \to 0$ as $n \to \infty$.*

Theorem 4. *Let Assumptions S, C, I and A1-A2 hold. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{b} - b_0\right) \xrightarrow{d} N\left(0, b_0^4 \Sigma_s\right).$$

We can estimate the asymptotic variance of $\widehat{b}$ by using the sample variance of the sequence $\{\widehat{s}_i\widehat{\varepsilon}_i\}_{i=1}^n$ where $\widehat{\varepsilon}_i = \widehat{g}\left(C_i', V_i'\right) R_i' - \widehat{b}^{-1}\widehat{g}\left(C_i, V_i\right)$, and $\widehat{s}$ is obtained as our estimator $\widehat{g}$, with the normalization

$$\frac{1}{n}\sum_{l=1}^{n}\widehat{g}(C_l, V_l)\widehat{s}(C_l, V_l) = 1.$$

An alternative is to use bootstrap. Since the eigenvalue b_0^{-1} is simple, and isolated from other eigenvalues, we expect the standard bootstrap sampling with replacement to provide a consistent estimation for the asymptotic distribution of $\sqrt{n}(\widehat{b} - b_0)$, and in particular, for confidence intervals. See Hall, Lee, Park and Paul (2009) and references therein.

Our next result establishes an asymptotic expansion for $\widehat{g} - g_0$. This expansion can be used to obtain rates for $\widehat{g} - g_0$ and to establish asymptotic normality of (semiparametric) functionals of $\widehat{g}$. Define the process $\Delta_n\left(c, v\right) \equiv n^{-1}\sum_{i=1}^n \varepsilon_i\phi_i(c, v)$, where recall that $\phi_i(c, v) = K_{hi}\left(c, v\right)/\widehat{f}\left(c, v\right)$. Note that a standard result in kernel estimation is that for all (c, v) in the interior of S, under suitable conditions,

$$\sqrt{nh_n^\ell}\Delta_n(c, v) \xrightarrow{d} N\left(0, \Sigma_\Delta\left(c, v\right)\right),$$

with $\Sigma_\Delta\left(c, v\right) = f^{-1}(c, v)\sigma^2\left(c, v\right)\kappa_2$, $\kappa_2 = \int K^2(u)du$ and $\sigma^2\left(c, v\right) = E\left[\varepsilon_i^2|C_i = c, V_i = v\right]$.

Recall $L^\dagger$ denotes the Moore-Penrose pseudoinverse of $L = b_0A - I$, which under our conditions is continuous (cf. Section 3.1).

Theorem 5. *Let Assumptions S, C, I and A1-A2 hold. Then, in $\mathcal{L}^2$, as $n \to \infty$,*

$$\sqrt{nh_n^\ell}\left(\widehat{g} - g_0\right) = b_0L^\dagger\sqrt{nh_n^\ell}\Delta_n + o_P\left(1\right).$$

This result implies that the rates of convergence of $\widehat{g} - g_0$ in $\mathcal{L}^2$ are the same as those of the NW kernel estimator of $E\left[\varepsilon_i|C_i = c, V_i = v\right]$. Combined with standard kernel regression results, this also implies asymptotic normality for $\sqrt{nh_n^\ell}L\left(\widehat{g} - g_0\right)$, which can be used for inference on g. For example, we could use the expansion of Theorem 5 to test parametric hypotheses about g, i.e., $H_0 : g_0(c, v) = g_{\eta_0}(c, v)$, against nonparametric alternatives, where the function $g_{\eta_0}(c, v)$ is known up to a finite-dimensional unknown parameter η_0 (e.g. power utility). A test can be based on the discrepancy

$$T_n = \left\|\sqrt{nh_n^\ell}\widehat{L}\left(\widehat{g} - \widetilde{g}\right)\right\|^2,$$

where $\widehat{L} = \widehat{b}\widehat{A} - I$ and $\widetilde{g} = g_{\widehat{\eta}}(c, v)$ is a parametric fit, with $\widehat{\eta}$ denoting a consistent estimator for η_0 under the null (e.g. a GMM estimator). Noting that $\widehat{L}\widehat{g} = 0$, T_n further simplifies to $T_n = ||\sqrt{nh_n^\ell}\widehat{L}\widetilde{g}||^2$. Similar test statistics have been suggested by Härdle and Mammen (1993) in a different context. More generally, we could test nonparametric hypotheses such as the significance of certain variables, for example $H_0 : g_0(c, v) = g_0(c, v')$ for all v, v', against nonparametric alternatives. The same T_n can be used, where now $\widetilde{g}$ denotes a restricted estimator of g_0 under the null (e.g. our marginal utility estimator depending only on c). In each case, the expansion in Theorem 5 is instrumental in analyzing the asymptotic limiting distribution of T_n, which can be readily obtained combining Theorem 5 here with the results of Härdle and Mammen (1993).

6 Summary Measures

We now consider some summary measures of the model, specifically, functionals of $\widehat{g}$. These are either behavioral parameters of interest such as the mean value of relative risk aversion ($MRRA$), or parameters having values that are relevant for testing. We first apply the results of the previous section to establish asymptotic normality of the estimated $MRRA$. We then list some other functionals of interest that can, in the same way, be shown to be asymptotically normal.

Define the $MRRA$ functional by

$$\gamma\left(g\right) \equiv E\left[\frac{-C\partial g(C,V)/\partial c}{g(C,V)}\right]. \tag{9}$$

The natural estimator of $\gamma\left(g_0\right)$ is the sample analog based on our estimator $\widehat{g}$, i.e.

$$\gamma_n\left(\widehat{g}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{-C_i\partial\widehat{g}(C_i,V_i)/\partial c}{\widehat{g}(C_i,V_i)}.$$

Under the assumptions for Theorem 5 above, $\widehat{g}$ is differentiable and bounded away from zero with probability tending to one, so $\gamma_n\left(\widehat{g}\right)$ is well-defined for large n. Define the class of functions

$$\mathcal{D} = \left\{(c,v) \to -c\frac{\partial\log(g(c,v))}{\partial c} : g \in \mathcal{G}\right\},$$

and the functions

$$d(c,v) \equiv \frac{\partial\left(c \times f(c,v)\right)}{\partial c}\frac{1}{f(c,v)} \qquad \text{and} \qquad \chi(c,v) \equiv \frac{d(c,v)}{g_0(c,v)}. \tag{10}$$

Also, we need to introduce some notation to be used in the asymptotic normality of $\gamma_n\left(\widehat{g}\right)$. Assuming $\chi \in \mathcal{L}^2$, define

$$\chi_s = \chi - \left\langle g_0, \chi\right\rangle\left\langle g_0, s\right\rangle^{-1} s. \tag{11}$$

The function χ_s has a geometrical interpretation as the value of χ projected parallel to s on a subspace of functions orthogonal to g_0. Let L^* denote the adjoint operator of L, and let χ_s^* denote the minimum norm solution of $\chi_s = L^*r$ in r, i.e. $\chi_s^* = \arg\min\{\|r\| : \chi_s = L^*r\}$, which is well defined because $\chi_s \in \mathcal{N}^{\perp}(L) = \mathcal{R}(L^*)$; see Luenberger (1997, Theorem 3, p. 157) for the latter equality.

Here $\mathcal{N}^{\perp}(L)$ denotes the orthogonal complement of the null space of L, see Luenberger (1997, p. 52) for a definition.

The $MRRA$ estimator behaves asymptotically as a sample average, with an influence function given by

$$\xi_i = (\zeta_i - E[\zeta_i]) - b_0\chi_s^*(C_i, V_i)\varepsilon_i, \tag{12}$$

where $\zeta_i = -C_i\left(\partial g_0(C_i, V_i)/\partial c\right)/g_0(C_i, V_i)$. The second term in ξ_i accounts for the estimation effect due to estimating g_0.

Assumption A3.

1. *$P(\widehat{g} \in \mathcal{G}) \to 1$ as $n \to \infty$ and the class $\mathcal{D}$ is P-Donsker*[5].

2. *$S = [l_c, u_c] \times S_V$, $\lim_{c \to l_c} cf(c, v) = 0 = \lim_{c \to u_c} cf(c, v)$ for all $v \in S_V$ and $P(\min\{g_0, \widehat{g}\} > \varepsilon) \to 1$ for some $\varepsilon > 0$.*

3. *$d \in \mathcal{L}^2$, $E[|\xi_i|^2] < \infty$ and $\chi_s^* \in \mathcal{C}^r(T)$.*

Assumption A3.1 is standard in the semiparametric literature, see, e.g. Chen, Linton and Van Keilegom (2003). The following Lemma provides sufficient conditions for an example of $\mathcal{D}$ satisfying the P-Donsker property of the second part of Assumption A3.1. Its proof is a standard exercise in empirical processes theory, and hence it is omitted.

Lemma 2. *Suppose that $\mathcal{G}$ is a subset of $\mathcal{C}^{\eta}(T)$ of functions bounded away from zero, where $\eta > (2 + \ell)/2$, and that $E[C_i^2] < \infty$. Then, $\mathcal{D}$ is P-Donsker.*

Assumption A3.2 is similar to other assumptions required in estimation of average derivatives, see Powell, Stock and Stoker (1989). This assumption guarantees that $\gamma_n(\widehat{g})$ is well defined and regular. Assumption A3.3 implies that the asymptotic variance of $\gamma_n(\widehat{g})$ is finite.

Theorem 6. *Let Assumptions S, C, I and A1-A3 hold. Then,*

$$\sqrt{n}\left(\gamma_n(\widehat{g}) - \gamma(g_0)\right) \xrightarrow{d} N\left(0, E\left[\xi_i^2\right]\right),$$

where ξ_i is defined in (12).

[5] Let P_n be the empirical measure with respect to P. Using a standard empirical process notation, define $\mathbb{G}_n g = \sqrt{n}(P_n - P)$. Then $\mathcal{D}$ is P-Donsker if $\mathbb{G}_n$ converges weakly to $\mathbb{G}$ in the space of uniformly bounded functions on $\mathcal{D}$, $l^{\infty}(\mathcal{D})$, where $\mathbb{G}$ is a mean-zero P-Brownian bridge process with uniformly continuous sample paths with respect to the semi-metric $\rho(d, d')$ defined by $\rho(d, d') = \sqrt{Var\left(d(C, V) - d'((C, V))\right)}$. For further details we refer the reader to van der Vaart and Wellner (1996).

Estimating the asymptotic variance of $\gamma_n(\widehat{g})$ by plug-in methods would be possible but complicated. In our application we use the bootstrap, which can be justified along the lines of Chen, Linton and Van Keilegom (2003).

Now consider some other functionals of interest. The asymptotic normality of each can be established using the same methods as Theorem 6. As with $\gamma_n(\widehat{g})$, in our applications we will use the bootstrap to estimate their limiting distributions. In our empirical work we consider a model allowing for habits, where $C_i = C_{t+1,i}$ and $V_i = C_{t,i}$ for two time periods t and $t+1$ (these time periods may vary across individuals). For the remainder of this section we drop the i subscript for clarity. Closely related to the $MRRA$ are local averages defined by

$$\rho(q,s) = E\left[\frac{-C_{t+1}\partial g_0(C_{t+1},C_t)/\partial C_{t+1}}{g_0(C_{t+1},C_t)}|C_{t+1}\in Q_q, C_t\in S_s\right], \tag{13}$$

where Q_q denotes the interval between the $q-1$ and q quartile of C_{t+1}, and S_s denotes the interval between the $s-1$ and s quartile of C_t for $q,s=1,2,3,4$. We refer to each of these local averages of the RRA between different quartiles as a $QRRA$ (quartile relative risk aversion).

We can use our results to construct tests of heterogeneity in risk aversion measures as follows. The sample analogs of the $QRRA$ parameters $\rho(q,s)$ can be shown to be asymptotically normal under the same conditions above used for the $MRRA$. That is, with the simplified notation $\rho(q)\equiv\rho(q,q)$ for the parameter and $\rho_n(q)\equiv\rho_n(q,q)$ for the plug-in estimator, it can be shown

$$\sqrt{n}\left(\rho_n(q)-\rho(q)\right)\xrightarrow{d} N\left(0,\sigma^2(q)\right),$$

for a suitable asymptotic variance $\sigma^2(q)$, $q=1,2,3$ and 4. Moreover, by definition, $\sqrt{n}\left(\rho_n(q)-\rho(q)\right)$ and $\sqrt{n}\left(\rho_n(s)-\rho(s)\right)$ are asymptotically independent for $q\neq s$. This suggests a simple strategy for testing heterogeneity in risk aversion by means of simple pairwise t-tests for the hypotheses, for $q\neq s$,

$$H_{0qs}:\rho(q)=\rho(s)\qquad vs\qquad H_{1qs}:\rho(q)\neq\rho(s).$$

The t-statistics are constructed as

$$t_{qs}=\frac{\sqrt{n}\left(\rho_n(q)-\rho_n(s)\right)}{\sqrt{\sigma_n^2(q)+\sigma_n^2(s)}},$$

for suitable consistent estimates $\sigma_n^2(q)$ of the asymptotic variances $\sigma^2(q)$, for $q=1,2,3$ and 4. We then reject H_{0qs} when t_{qs} is large in absolute value, using that t_{qs} converges to a standard normal under H_{0qs}. We use these tests of heterogeneity in our application below.

We also construct some tests for the absence of habits, i.e.

$$\frac{\partial g_0(C_{t+1},C_t)}{\partial C_t}=0.$$

Our tests are based on the functional

$$\delta\left(g\right)=E\left[\frac{\partial g(C_{t+1},C_t)}{\partial C_t}\tau(C_{t+1},C_t)\right],$$

for various positive functions $\tau(\cdot)$. When there is no habit effect $\delta\left(g_0\right)=0$ for any choice of τ. As with $\gamma\left(g_0\right)$, for each choice of function τ we estimate $\delta\left(g_0\right)$ by plugging in $\widehat{g}$ for g_0 and replacing the expectation with a sample average. The asymptotic normality of this estimator and its bootstrap approximation is then used for inference, analogous to our analysis of $\gamma\left(g_0\right)$.

7 Monte Carlo Experiment

In this section we illustrate the finite-sample performance of our estimator described in the previous sections based on a $CRRA$ utility function so that $g_0\left(c,v\right)=c^{-\eta_0}$, where η_0 in this case equals the $MRRA$. The model is then given by the Euler equation

$$b_0E\left[C_{t+1}^{-\eta_0}R_{t+1}|C_t\right]=C_t^{-\eta_0}.$$

We set $b_0=0.95$ and $\eta_0=0.5$. We draw a random sample of (C_t,C_{t+1}) from the distribution

$$(\log C_t,\log C_{t+1})\sim N\left(0,\left(\begin{array}{cc}0.25 & 0.1\\ 0.1 & 0.25\end{array}\right)\right),$$

and construct $R_{t+1}=b_0^{-1}\left(1+\epsilon_t\right)\left(C_{t+1}/C_t\right)^{\eta_0}$, where ϵ_t is distributed uniformly on $[-0.5,0.5]$ and drawn independently of (C_t,C_{t+1}). This design was chosen to generate data that satisfies the Euler equation model, has realistic parameter values and consumption distribution, and avoids the approximation and other numerical errors that would result from solving each individual's dynamic optimization problem numerically.

To save space we only report simulation results for two experiments, each with sample sizes $n=500$ and $n=2000$. The number of bootstrap replications used in each simulation is 200, and we repeat each simulation 1000 times. We compute our proposed nonparametric estimators and compare them to the method of moments estimator defined using the correctly specified $CRRA$ utility function with a constant and C_t as instruments. So while our estimator attempts to recover the constant b_0 and the entire function g_0, this alternative just estimates the two constants b_0 and η_0, using two moments of the data. In our tables estimates from this correctly specified parametric functional form are labeled $CRRA$.

We consider two nonparametric estimators. The first one, which we label $NP-1$, correctly conditions on just C_t (since our choice of $g_0\left(c,v\right)$ does not depend on v), and so only entails estimation of a one-dimensional marginal utility function. In anticipation of our empirical application in the next section, the second nonparametric estimator, denoted $NP-2$, uses both C_t and V_t as conditioning variables, where $V_t = C_{t-1}$ is in this case an irrelevant habit variable. We simulate C_{t-1} by drawing from a $N(1,1)$ distribution that is independent of (C_t, C_{t+1}).

We compute our estimates using the procedure described in Section 4 that incorporates the transformation suggested in equation (7). While not necessary in theory, we find that estimates of g_0 fit better in the tails using this transformation than not, though the differences in overall integrated mean square errors and other measures of fit are small. In order to apply the transformation, note that equation (7) can be re-written as

$$bE[g^*(C_{t+1}, V_{t+1})R^*_{t+1} \mid C_t, V_t] = g^*(C_t, V_t),$$

where $g^*(C_{t+1}, V_{t+1}) \equiv C_{t+1}g(C_{t+1}, V_{t+1}), g^*(C_t, V_t) \equiv C_t g(C_t, V_t)$ and $R^*_{t+1} \equiv (C_t/C_{t+1}) R_{t+1}$. With these definitions the procedure remains as described in Section 4 after redefining the return variable, from R_{t+1} to R^*_{t+1}. The procedure then yields an estimate of g^*, from which the marginal utility function g is then recovered using the relation $g(c,v) = g^*(c,v)/c$. Throughout we set the bandwidth to be $1.06sn^{-1/3.5}$, where s is the sample standard deviation of C_t. This is essentially Silverman's rule applied to the rate $n^{-1/3.5}$. All of our estimators for g_0 are normalized to have a unit norm with respect to the empirical L^2-norm.

For each finite-dimensional parameter and summary measure we consider, we report the mean, standard deviation, 2.5th percentile, 97.5th percentile, 95% coverage probability based on normal distribution, their bootstrap counterparts and the root mean square error.[6] Table 1 reports estimates of the discount factor from our three estimators, $CRRA$, $NP-1$, and $NP-2$. Table 2 reports estimates of the $MRRA$, which for the $CRRA$ model is just the estimated constant η_0, while for the nonparametric estimators the $MRRA$ is $\gamma(g_0)$ defined by equation (9). Table 1 shows that all of the estimators succeed in estimating the discount factor b very accurately. This is in contrast to many macro models, which often calibrate the discount factor due to the difficulty in estimating it accurately. Table 2 shows somewhat more difficulty in estimating the $MRRA$, but the relative accuracy of our nonparametric estimates to the parametric alternative is similar. In both tables the root mean squared errors of our nonparametric estimates are seen to shrink with sample size and increase with dimensionality at rates that are generally consistent with asymptotic theory.

Figures 1 and 2 show plots of the one-dimensional nonparametric (i.e., $NP-1$) estimated marginal utility function g_0 as a function of C_t. Figure 1 is $n = 500$ while Figure 2 is $n = 2000$. For each figure, the solid line denotes the mean, the dotted line denotes the 95% confidence interval, and the dashed line is the true. One can see from these figures that $NP-1$ quite accurately tracks the true

[6]The normal coverage probability is constructed ex-post using the true (simulated) standard deviation.

function. The precision of these fits can also be summarized by their integrated mean square error (weighted with respect to the true density), which is 0.0014 for $n = 500$ and 0.0005 for $n = 2000$.

Not surprisingly, estimates of the two-dimensional $NP-2$ are noisier, since by design the second conditioning variable V_t is irrelevant. The results for $NP-2$ can be summarized by their implied quartile averages $QRRA$. Table 3 reports estimates of each $QRRA$, $\rho(q,s)$ for all quartiles q and s having $|q-s| \leq 1$.[7] Table 3 shows that estimates of $QRRA$ have generally about an order of magnitude larger root mean squared error than $MRRA$, which is not surprising since each $\rho(q,s)$ is obtained by averaging over 1/16 as much data (one quartile of current consumption and one quartile of lagged consumption observations) as $MRRA$.

One unexpected finding is that estimates of $\rho(q,s)$ display substantially larger biases and root mean squared errors for larger values of q and s than for smaller values, suggesting that our $NP-2$ estimates of the marginal utility function tend to be less accurate at higher consumption levels. This can also be seen for $NP-1$ in Figure 1, where the standard error bands widen at higher consumption levels.

In Table 4 we report estimates of $\delta(g_0)$ that can be used to test for the presence of habits in g_0. In our experiments estimates of $\delta(g_0)$ do not differ significantly from zero as expected, since our specification of g_0 does not have any habit effect. Generally, all of our parameter estimates and test statistics appear to have distributions across simulations that are reasonably well approximated by the bootstrap, e.g., biases are relatively small, bootstrap standard errors are generally close to the standard deviations across simulations, and bootstrap confidence intervals are generally close to the true. Both coverage probabilities based on the normal approximation and the bootstrap generally are relatively close to the nominal.

[7] We only report pairs of quartiles i and j where $|q-w| \leq 1$, because a value that violates this inequality, like $\rho(4,1)$, corresponds to individuals who's consumption jumps from the fourth to the first quartile, and in real data the number of such individuals who make this jump would be too small to reliably estimate their $QRRA$.

	b_0	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n=500$	$CRRA$	0.000	0.012	0.926	0.975	0.946	0.012	0.926	0.974	0.940	0.012
	$NP-1$	0.006	0.027	0.917	0.971	0.984	0.018	0.915	0.980	0.929	0.028
	$NP-2$	0.009	0.041	0.808	0.983	0.963	0.031	0.895	1.012	0.932	0.042
$n=2000$	$CRRA$	0.000	0.006	0.938	0.961	0.960	0.006	0.938	0.962	0.950	0.006
	$NP-1$	0.004	0.020	0.936	0.960	0.992	0.009	0.932	0.965	0.924	0.020
	$NP-2$	0.005	0.028	0.862	0.965	0.974	0.021	0.922	0.994	0.946	0.028

Table 1: Summary statistics of Monte Carlo estimates of the discount factor b_0. The true is $b_0 = 0.95$. $CRRA$, $NP-1$ and $NP-2$ refer respectively to the parametric, one-dimensional nonparametric, and two-dimensional nonparametric estimators.

	$MRRA$	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n=500$	$CRRA$	0.000	0.046	0.420	0.590	0.956	0.046	0.411	0.592	0.944	0.046
	$NP-1$	-0.058	0.107	0.431	0.714	0.961	0.101	0.359	0.751	0.906	0.122
	$NP-2$	-0.096	0.194	0.277	0.888	0.952	0.194	0.209	0.986	0.930	0.217
$n=2000$	$CRRA$	0.001	0.023	0.456	0.545	0.950	0.023	0.454	0.544	0.952	0.023
	$NP-1$	-0.032	0.077	0.470	0.610	0.988	0.052	0.430	0.628	0.914	0.083
	$NP-2$	-0.067	0.092	0.412	0.716	0.934	0.109	0.355	0.782	0.906	0.114

Table 2: Summary statistics of Monte Carlo estimates of the $MRRA$, which is η_0 for the parametric and $\gamma(g_0)$ for the nonparametric estimators. The true is $MRRA = 0.5$. $CRRA$, $NP-1$ and $NP-2$ refer respectively to the parametric, one-dimensional nonparametric, and two-dimensional nonparametric estimators.

	$QRRA$	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n = 500$	$\rho(1,1)$	-0.158	0.205	0.273	1.068	0.910	0.242	0.115	1.068	0.878	0.259
	$\rho(1,2)$	-0.068	0.366	-0.049	1.167	0.969	0.358	-0.137	1.287	0.969	0.372
	$\rho(2,1)$	-0.149	0.222	0.242	1.060	0.932	0.246	0.145	1.118	0.904	0.267
	$\rho(2,2)$	-0.055	0.327	0.000	1.151	0.961	0.355	-0.137	1.274	0.965	0.331
	$\rho(2,3)$	-0.010	0.450	-0.240	1.187	0.973	0.480	-0.433	1.477	0.973	0.450
	$\rho(3,2)$	-0.053	0.326	-0.014	1.081	0.969	0.351	-0.121	1.275	0.966	0.330
	$\rho(3,3)$	0.009	0.457	-0.279	1.180	0.972	0.460	-0.408	1.428	0.966	0.457
	$\rho(3,4)$	-0.102	0.785	-0.850	1.972	0.963	0.933	-1.320	2.452	0.972	0.792
	$\rho(4,3)$	-0.029	0.400	-0.137	1.181	0.969	0.470	-0.345	1.515	0.978	0.401
	$\rho(4,4)$	-0.281	0.980	-0.957	2.378	0.954	1.079	-1.486	2.876	0.955	1.019
$n = 2000$	$\rho(1,1)$	-0.104	0.179	0.350	0.825	0.978	0.158	0.280	0.889	0.888	0.206
	$\rho(1,2)$	-0.023	0.272	0.125	0.903	0.984	0.249	0.048	1.027	0.954	0.273
	$\rho(2,1)$	-0.087	0.146	0.330	0.859	0.938	0.171	0.245	0.910	0.912	0.170
	$\rho(2,2)$	-0.018	0.214	0.151	0.882	0.964	0.251	0.031	1.030	0.968	0.214
	$\rho(2,3)$	-0.007	0.319	0.004	1.019	0.988	0.314	-0.104	1.133	0.956	0.319
	$\rho(3,2)$	-0.009	0.274	0.078	0.871	0.980	0.254	0.024	1.013	0.954	0.274
	$\rho(3,3)$	-0.016	0.376	0.095	0.956	0.986	0.310	-0.067	1.153	0.962	0.377
	$\rho(3,4)$	-0.078	0.388	-0.136	1.322	0.952	0.573	-0.583	1.722	0.970	0.396
	$\rho(4,3)$	-0.002	0.385	0.129	0.913	0.980	0.302	-0.054	1.123	0.964	0.385
	$\rho(4,4)$	-0.244	0.476	0.053	1.641	0.940	0.624	-0.571	1.948	0.958	0.535

Table 3: Summary statistics of Monte Carlo estimates of $QRRA$, which is $\rho(q,s)$ from $NP-2$. The true is $\rho(q,s) = 0.5$ for all q and s.

	$\tau(C_{t+1}, C_t)$	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n = 500$	C_{t+1}	-0.002	0.111	-0.111	0.132	0.975	0.118	-0.255	0.200	0.975	0.111
	C_t	-0.006	0.097	-0.128	0.125	0.975	0.118	-0.245	0.209	0.980	0.097
	C_{t+1}^2	-0.010	0.289	-0.249	0.252	0.977	0.262	-0.567	0.438	0.965	0.290
	C_t^2	-0.030	0.237	-0.331	0.270	0.967	0.269	-0.531	0.502	0.977	0.238
	$C_{t+1}C_t$	-0.015	0.229	-0.209	0.190	0.972	0.220	-0.463	0.370	0.973	0.230
$n = 2000$	C_{t+1}	-0.005	0.078	-0.070	0.072	0.978	0.077	-0.154	0.131	0.978	0.079
	C_t	-0.009	0.080	-0.084	0.072	0.982	0.077	-0.154	0.132	0.978	0.081
	C_{t+1}^2	-0.013	0.229	-0.176	0.149	0.986	0.188	-0.374	0.319	0.968	0.229
	C_t^2	-0.036	0.244	-0.270	0.150	0.986	0.195	-0.382	0.344	0.966	0.247
	$C_{t+1}C_t$	-0.016	0.222	-0.146	0.107	0.984	0.160	-0.313	0.268	0.970	0.223

Table 4: Summary statistics of Monte Carlo estimates of $\delta(g_0)$, used to test for the presence of habit effects. The true value of each $\delta(g_0)$ is zero. The $\tau(C_{t+1}, C_t)$ column lists the functions that are used to define $\delta(g_0)$.

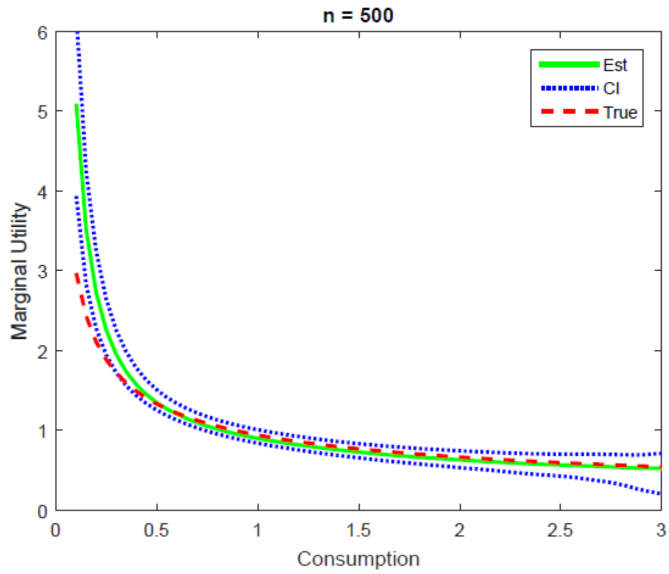

Figure 1: Estimates of the marginal utility function g_0 using simulated data with $n = 500$. Est, CI, and $True$ represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true.

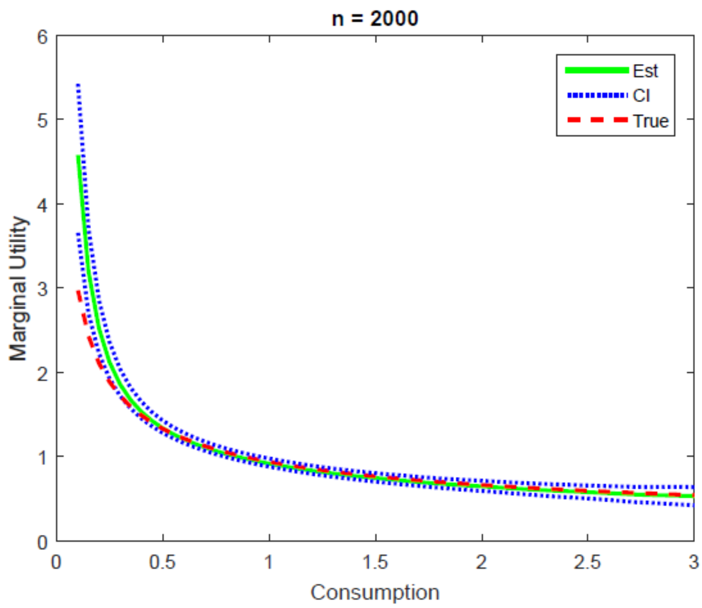

Figure 2: Estimates of the marginal utility function g_0 using simulated data with $n = 2000$. Est, CI, and $True$ represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true.

8 Empirical Application

In this section, we apply our framework to a real world consumption data set. Specifically, we use quarterly US Consumer Expenditure Survey (CEX) household-level data for households sampled between 1980Q1 and 2012Q4. Our consumption data $C_{t_i,i}$ (for household i in period t_i) is total expenditures on nondurables that we convert from nominal to real by deflating using the US consumer price index (with year 2000 as base). We also deflate by household size to get real expenditures per capita within the household. To avoid including additional demographic regressors we focus on a relatively homogenous sample by including only urban households, with each head of household being between 30 and 50 years of age and an education level of high school diploma or higher. We only consider households that report four consecutive quarters of consumption, and removed as outliers households that displayed extreme variation in consumption, defined as a greater than 50% change in consumption from one quarter to the next. The resulting dataset contains 18912 households. We construct two types of asset returns R_t, one risk free and the other is risky. The risk free return is based on 1-month US treasury bills. The risky return is based on the Wilshire 5000 stock index, with dividends reinvested. Both asset returns are converted into real terms computed on a quarterly basis. We provide some summary statistics of the data in Table 5.

		Mean	Std	10th	25th	50th	75th	90th
Consumption	$C_{t_i-1,i}$	3048.128	1438.924	1565.728	2066.940	2765.843	3712.934	4827.351
	$C_{t_i,i}$	2991.451	1419.682	1529.328	2025.249	2715.915	3631.024	4765.924
	$C_{t_i+1,i}$	2938.243	1401.901	1503.810	1989.552	2664.001	3574.610	4655.104
Risk free	R_{t+1}	1.040	0.031	0.999	1.015	1.040	1.055	1.080
Risky	R_{t+1}	1.016	0.068	0.938	0.986	1.024	1.065	1.091

Table 5: Summary statistics of the quarterly CEX and return data in real terms (year 2000 as base), containing the sample mean, standard deviation and various percentiles of the variables.

Using this CEX data, we apply the same estimators as in the Monte Carlo study, that is, the parametric $CRRA$, the one-dimensional $(NP-1)$ nonparametric estimator that assumes no habit is present, and the two-dimensional $(NP-2)$ nonparametric estimator. These three estimators are each implemented twice; once using the riskless returns, and a second time using the risky returns. Note that if the model is correctly specified, both assets should result in roughly the same estimates of b_0 and g_0. We employ the bandwidth $h = 1.06sn^{-1/4}$, where s is the sample standard deviation of consumption.[8] Standard errors and confidence intervals are computed using

[8] This is a slightly larger rate for Silverman's rule than we used in the Monte Carlo. We chose this rate by an informal comparison of a few alternatives, choosing the one that by eye appeared least erratic. We speculate that

nonparametric bootstrap, in the same way as with the simulated data.

The estimates for the discount factor b_0 and $MRRA$ are reported in Tables 6 and 7 respectively, and the $QRRA$'s are in Table 8. Table 9 reports p-values from the t-statistics constructed from the normalized pairwise differences between estimates of $\rho(q)$ and $\rho(s)$, as suggested at the end of Section 6, which can be used to detect heterogeneity of risk aversions in different parts of the population. The tests for habits can be found in Table 10. Using the risk free asset, Figure 3 plots the $NP-1$ estimate of g_0, while figures 4, 5 and 6 plot the $NP-2$ estimates of g_0 conditioning on the lag consumption level at the first, second and third quartiles respectively. Figures 7 to 10 are analogous plots using the risky asset.

As in the simulations, we find the estimates of the discount factor b_0 to be quite similar across all estimators, though their estimated standard errors seem surprisingly low even with a large sample. Likewise, the nonparametric model error bands in Figures 4 to 10 seem very tight, given some of the peculiar shapes seen at higher consumption levels, and given the modest differences seen in the two assets. The estimates of the $MRRA$ are rather low compared to the literature, however, the $QRRA$ show larger values for at least some ranges of consumption. For the nonparametric models we generally find similar estimates for the riskless and risky asset, which provides evidence that the pricing model is appropriate.

One motivation for estimating marginal utility nonparametrically is to look for evidence on whether standard parametric alternatives are correctly specified, or whether there is some feature of the data that parametric models may have missed. Looking across these estimates, one can see evidence that the popular $CRRA$ parametric model is misspecified. The $CRRA$ estimate of $MRRA$ is essentially zero, and indeed changes sign across the riskless and risky asset. As the name implies, $CRRA$ assumes relative risk aversion is constant across consumption levels. In contrast, the $QRRA$ estimates show variation in risk aversion, depending both on current and on last period's consumption level. Generally, the estimates show levels of risk aversion that decrease as individual's consumption levels increase. Formal testing based on pairwise t-statistics also confirms that some variation exists. Moreover, the shapes seen in the Figures 4 to 6 and 8 to 10 suggest that utility may depend in more complicated ways on past consumption than typical habit models permit, including even semiparametric habit models like Chen and Ludvigson (2009) or Chen, Chernozhukov, Lee and Newey (2014). Figures 3 and 7 show that, if one ignores or averages over past consumption, the departures from $CRRA$ become smaller, which suggests that standard models may to some extent obscure the complexity of habit affects by averaging. The overall estimated average values of risk aversion (the $MRRA$) in the nonparametric models are still rather low (see Table 7), but are not

measurement error in C_{i,t_i} may be causing increased noisiness in the estimates, requiring greater smoothing than in our simulated data.

nearly as implausibly close to zero as the $CRRA$ model.

The test results for habits in Tables 9 are mixed. On one hand, some of the point estimates of $\delta(g_0)$ are very far from zero, suggesting that utility may well possess habits. However, the standard errors and confidence bands for these statistics are also very wide, so most of these departures, particularly with the risk-free rate, while numerically large, are not statistically significant. However, for the risky asset almost all specifications of $\delta(g_0)$ do significantly reject the assumption of no habits.

We end this section with some caveats regarding our estimates, and our model in general. First, CEX data are known to be quite noisy, often varying substantially from quarter to quarter. Indeed, for this reason most applications of CEX data aggregate up to the annual level, thereby removing the short panel component of the data that we exploit. However, we require data in which households are observed for a few periods in a row (to construct a $C_{t_i+1,i}$, $C_{t_i,i}$, and $C_{t_i-1,i}$ for each household i), and we also require data that covers a long span of time (in this case 129 quarters) to observe significant variation in asset returns. This greatly limits our choices for possible data sets. Still, interpretation of our results should recognize that our data may suffer from rather substantial amounts of measurement error. See Gayle and Khorunzhina (2014) for evidence on the potential effects of measurement error in consumption Euler equations with habits.

Another limitation of our results is that we do not model unobserved preference heterogeneity. The vector $V_{t_i,i}$ can in theory include observable characteristics of consumers that affects preferences, such as demographic characteristics, stocks of previously purchased durables, past consumption, etc. For simplicity, rather than including such variables (other than past consumption), we focused on a relatively homogeneous subset of households. It should be noted, however, than an offsetting advantage of our model is that we do not impose the restrictions on preferences that are generally needed to estimate asset pricing models. In particular, pricing models are generally estimated using aggregate consumption data, and so impose strong homogeneity restrictions on preferences, and hence on the functional form of g, to allow aggregation of marginal utility functions across consumers. An alternative approach that allows for unobserved heterogeneity in parametric Euler equation models is explored in Hoderlein, Nesheim and Simoni (2012), but this approach is very different from ours and cannot be readily extended to our nonparametric framework.

A more subtle issue is the potential role of aggregate shocks. To illustrate, suppose all of our consumers had been observed in the same two time periods, and a large negative macro shock had occurred in the second of these periods. Then second period consumption would on average have been lower than expected for most consumers, and as a result the observed joint distribution of consumption across the two periods would not equal the joint distribution that first period consumption was based upon. In our model, this potential source of estimation bias is mitigated by our choice of data. Each household is observed for at most four periods, but we draw data over 129 time periods (quarters), so some households are observed in the 1980's and others as late as 2012. As a result, the impacts on our estimates of potential bias due to negative aggregate shocks in some periods is should be largely offset by positive aggregate shocks in other periods. However, although our point estimates should therefore be largely unaffected by aggregate shocks, our asymptotic theory assumes independence across households, and aggregate shocks could cause dependence across consumers that happen to be observed in the same time period. Our asymptotic theory could be modified to allow for some dependence using uniform rate results from Andrews (1995).

	b_0	Est	Ste	Lpc	Upc
Risk free	$CRRA$	0.966	0.000	0.966	0.967
	$NP-1$	0.961	0.001	0.960	0.963
	$NP-2$	0.961	0.001	0.960	0.963
Risky	$CRRA$	0.986	0.001	0.985	0.987
	$NP-1$	0.979	0.001	0.978	0.981
	$NP-2$	0.978	0.001	0.976	0.982

Table 6: Summary statistics of CEX data estimates of the discount factor b_0. $CRRA$, $NP-1$ and $NP-2$ refer respectively to the parametric, one-dimensional nonparametric, and two-dimensional nonparametric estimators.

	$MRRA$	Est	Ste	Lpc	Upc
Risk free	$CRRA$	-0.004	0.003	-0.010	0.002
	$NP-1$	0.133	0.052	0.096	0.168
	$NP-2$	0.194	0.026	0.133	0.237
Risky	$CRRA$	0.006	0.007	-0.009	0.018
	$NP-1$	0.196	0.020	0.150	0.231
	$NP-2$	0.281	0.032	0.202	0.325

Table 7: Summary statistics of CEX data estimates of the $MRRA$, which is η_0 for the parametric and $\gamma(g_0)$ for the nonparametric estimators. $CRRA$, $NP-1$ and $NP-2$ refer respectively to the parametric, one-dimensional nonparametric, and two-dimensional nonparametric estimators.

	$QRRA$	Est	Ste	Lpc	Upc
Risk free	$\rho(1,1)$	0.342	0.047	0.219	0.417
	$\rho(1,2)$	0.154	0.028	0.082	0.201
	$\rho(2,1)$	0.253	0.035	0.169	0.311
	$\rho(2,2)$	0.139	0.022	0.086	0.175
	$\rho(2,3)$	0.076	0.018	0.038	0.106
	$\rho(3,2)$	0.147	0.023	0.093	0.189
	$\rho(3,3)$	0.063	0.016	0.024	0.091
	$\rho(3,4)$	0.098	0.018	0.049	0.125
	$\rho(4,3)$	0.050	0.026	-0.003	0.108
	$\rho(4,4)$	0.296	0.084	0.119	0.467
Risky	$\rho(1,1)$	0.436	0.059	0.311	0.540
	$\rho(1,2)$	0.257	0.040	0.165	0.316
	$\rho(2,1)$	0.358	0.049	0.249	0.436
	$\rho(2,2)$	0.237	0.032	0.157	0.285
	$\rho(2,3)$	0.184	0.028	0.125	0.229
	$\rho(3,2)$	0.242	0.034	0.156	0.297
	$\rho(3,3)$	0.145	0.027	0.082	0.187
	$\rho(3,4)$	0.190	0.032	0.118	0.248
	$\rho(4,3)$	0.177	0.039	0.086	0.252
	$\rho(4,4)$	0.336	0.084	0.170	0.495

Table 8: Summary statistics of CEX data estimates of $QRRA$, which is $\rho(q,s)$ from $NP-2$.

p-values	q/s	1	2	3	4
Risk free	1	—	0.000	0.000	0.632
	2	—	—	0.005	0.070
	3	—	—	—	0.006
Risky	1	—	0.003	0.000	0.325
	2	—	—	0.028	0.271
	3	—	—	—	0.030

Table 9: Summary statistics of CEX data for the p-values of a pairwise t-statistics base on $QRRA$ to test the null hypothesis that estimates of $\rho(q) = \rho(s)$ for $q \neq s$.

	$\tau(C_{t+1}, C_t)$	Est	Ste	Lpc	Upc
Risk free	$C_{t_i+1,i}$	-0.012	0.011	-0.028	0.017
	$C_{t_i,i}$	-0.018	0.012	-0.034	0.014
	$C^2_{t_i+1,i}$	-95.52	63.55	-189.2	76.19
	$C^2_{t_i,i}$	-162.5	80.65	-273.0	60.12
	$C_{t_i+1,i}C_{t_i,i}$	-118.4	65.43	-209.6	62.80
Risky	$C_{t_i+1,i}$	-0.041	0.013	-0.060	-0.006
	$C_{t_i,i}$	-0.046	0.014	-0.064	-0.007
	$C^2_{t_i+1,i}$	-163.7	63.45	-261.4	-2.332
	$C^2_{t_i,i}$	-217.2	78.51	-328.1	0.956
	$C_{t_i+1,i}C_{t_i,i}$	-178.7	64.29	-266.5	-4.279

Table 10: Summary statistics of CEX data estimates of $\delta(g_0)$, used to test for the presence of habit effects. The $\tau(C_{t+1}, C_t)$ column lists the functions that are used to define $\delta(g_0)$.

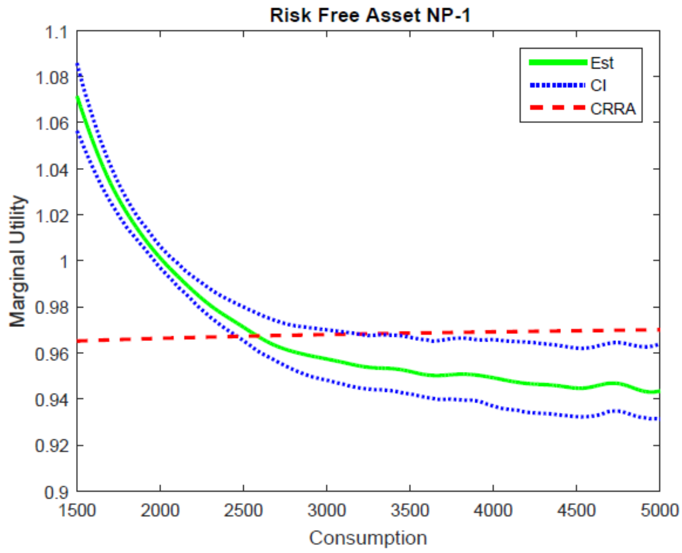

Figure 3: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI and $CRRA$ represent respectively the one-dimensional nonparametric estimate, its 95% confidence interval, and the parametric estimate.

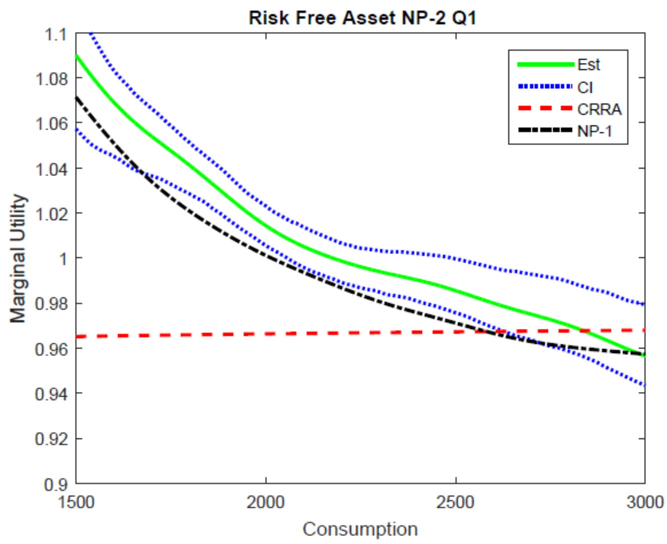

Figure 4: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the first quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

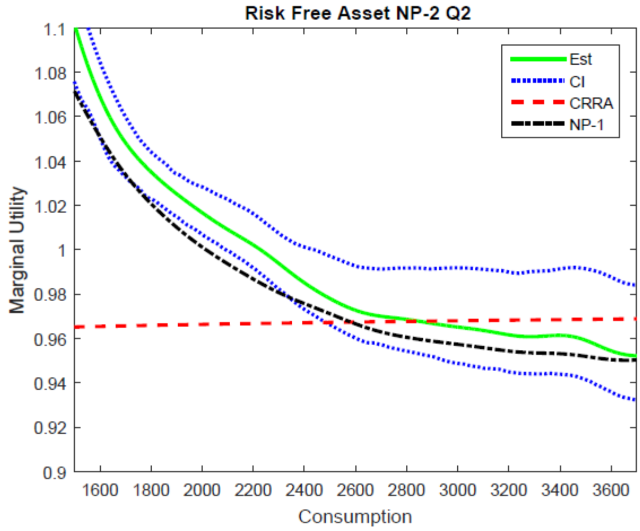

Figure 5: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the second quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

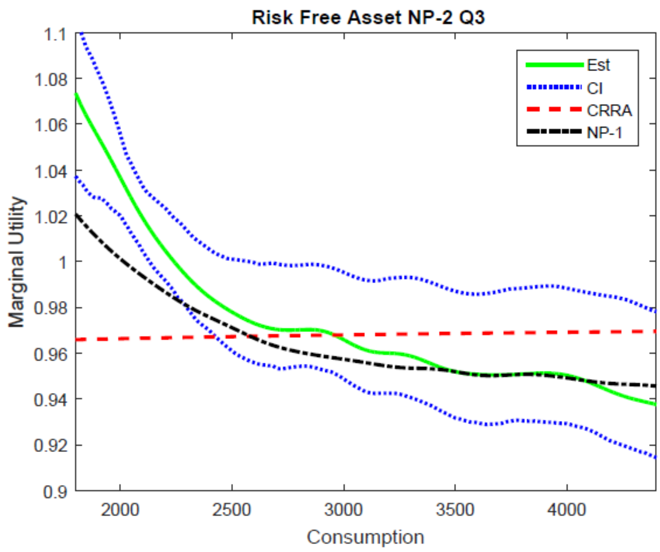

Figure 6: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the third quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

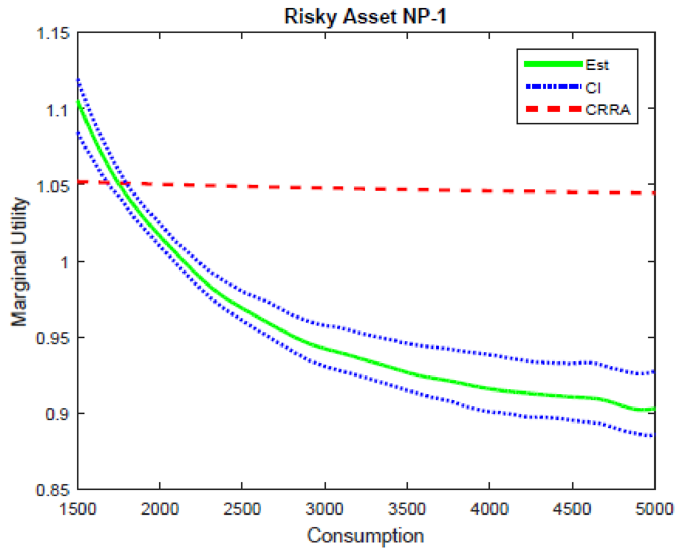

Figure 7: Estimates of the marginal utility function g_0 using CEX data with the risky returns. Est, CI and $CRRA$ represent respectively the one-dimensional nonparametric estimate, its 95% confidence interval, and the parametric estimate.

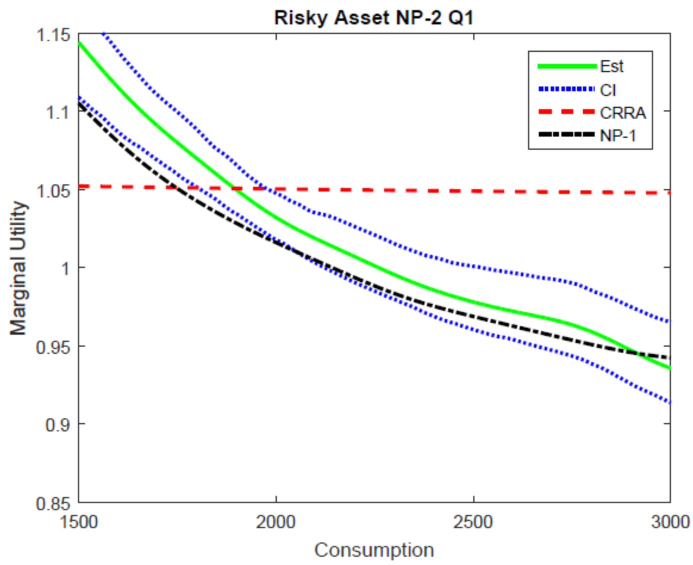

Figure 8: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the first quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

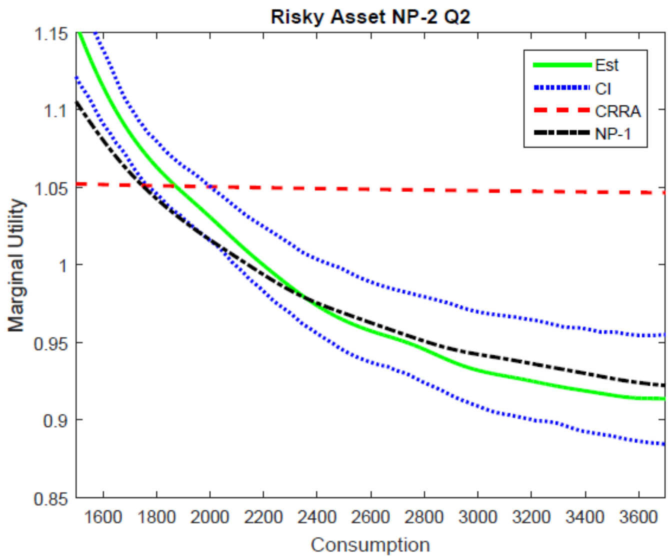

Figure 9: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the second quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

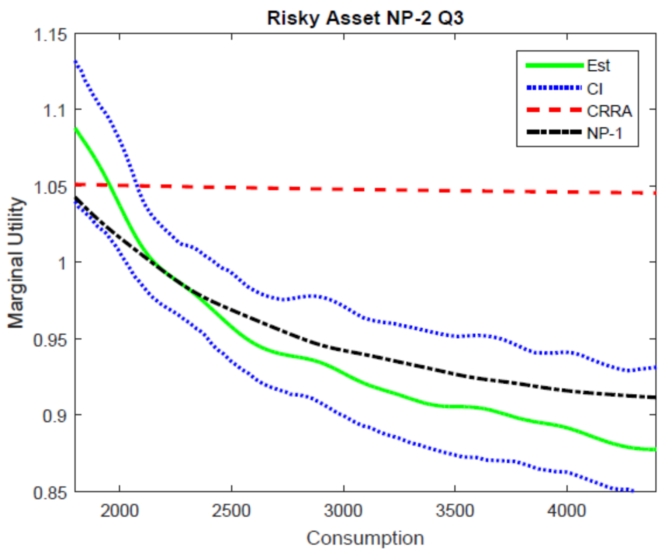

Figure 10: Estimates of the marginal utility function g_0 using CEX data with the risk free returns. Est, CI, $CRRA$ and $NP-1$ represent respectively the two-dimensional nonparametric estimate conditioning on the lag consumption level at the third quartile, its 95% confidence interval, the parametric estimate, and the one-dimensional nonparametric estimate.

9 Conclusions

We investigate nonparametric identification and estimation of marginal utilities and discount factors in consumption-based asset pricing Euler equations. The main features of our nonparametric identification results are: (i) the decomposition of the pricing kernel into its marginal utility and discount factor components, cast in the form equation (1), and (ii) the use of shape restrictions (positive marginal utilities). Together, these allow us to establish nonparametric global point identification of the model. Based on our identification arguments, we propose a new nonparametric estimator for marginal utilities and the discount factor that combines standard kernel estimation with the computation of a (finite-dimensional) matrix eigenvalue-eigenvector problem. No numerical integration or optimization is involved. The estimator is based on a sample analogue of (1) and is easy to implement, since no numerical searches are required. We establish a useful expansion for the marginal utility (suitably normalized), and limiting distribution theory for the discount factor and associated functionals of the marginal utility like the mean level of relative risk aversion. Due to the well posedness of equation (1), our estimator converges at comparable rates to ordinary nonparametric regression and does not suffer from issues associated with nonparametric instrumental variables estimation.

We apply our nonparametric methods to household-level CEX data and find evidence against the common assumption of constant relative risk aversion across consumers. Our estimates are fairly insensitive to the choice of asset used (risk-free vs risky), which supports our nonparametric model. We find empirical evidence for the presence of habits, and evidence that risk aversion varies across current and lagged consumption levels in ways that are not fully captured by standard parametric or even semiparametric specifications of habits in asset pricing models.

10 Appendix

10.1 Euler Equation Derivation

To encompass a large class of existing Euler equation and asset pricing models, consider utility functions that in addition to ordinary consumption, may include both durables and habit effects. Let U be a time homogeneous period utility function, b is the one period subjective discount factor, C_t is expenditures on consumption, D_t is a stock of durables, and Z_t is a vector of other variables that affect utility and are known at time t. Let V_t denote the vector of all variables other than C_t that affect utility in time t. In particular, V_t contains Z_t, V_t contains D_t if durables matter, and V_t contains lagged consumption C_{t-1}, C_{t-2} and so on if habits matter.

The consumer's time separable utility function is

$$\max_{\{C_t,D_t\}_{t=1}^{\infty}} E\left[\sum_{t=0}^{\infty} b^t U(C_t,V_t)\right].$$

The consumer saves by owning durables and by owning quantities of risky assets A_{jt}, $j=1,\ldots,J$. Letting C_t be the numeraire, let P_t be the price of durables D_t at time t and let R_{jt} be the gross return in time period t of owning one unit of asset j in period $t-1$. Assume the depreciation rate of durables is δ. Then without frictions the consumer's budget constraint can be written as, for each period t,

$$C_t + (D_t - \delta D_{t-1})\,P_t + \sum_{j=1}^{J} A_{jt} \leq \sum_{j=1}^{J} A_{jt-1}R_{jt}$$

We may interpret this model either as a representative consumer model, or a model of individual agents which may vary by their initial endowments of durables and assets and by $\{Z_t\}_{t=0}^{\infty}$. The Lagrangean is

$$E\left[\sum_{t=0}^{T} b^t U(C_t,V_t) - \left(C_t + (D_t-\delta D_{t-1})\,P_t + \sum_{j=1}^{J}(A_{jt}-A_{jt-1}R_{jt})\right)\lambda_t\right] \tag{14}$$

with Lagrange multipliers $\{\lambda_t\}_{t=0}^{\infty}$.

Consider the roles of durables and habits. For durables, define

$$g_d(C_t,V_t) = \frac{\partial U(C_t,V_t)}{\partial D_t}$$

which will be nonzero only if V_t contains D_t. For habits, we must handle the possibility of both internal or external habits. Habits are defined to be internal (or internalized) if the consumer considers both the direct effects of current consumption on future utility through habit as well as through the budget constraint. In the above notation, habits are internal if the consumer takes into account the fact that, due to habits, changing C_t will directly change V_{t+1}, V_{t+2} etc. Otherwise, if the consumer ignores this effect when maximizing, then habits called external.

If habits are external or if there are no habit effects at all, then define the marginal utility function g by

$$g(C_t,V_t) = \frac{\partial U(C_t,V_t)}{\partial C_t}$$

If habits exist and are internal then define the function $\widetilde{g}$ by

$$\widetilde{g}(I_t) = \sum_{\ell=0}^{L} b^{\ell} E\left[\frac{\partial U(C_{t+\ell},V_{t+\ell})}{\partial C_t} \mid I_t\right].$$

where L is such that V_t contains $C_{t-1}, C_{t-2}, \ldots, C_{t-L}$, and I_t is all information known or determined by the consumer at time t (including C_t and V_t). For external habits, we can write $\widetilde{g}(I_t) = g(C_t, V_t)$, while for internal habits define

$$g(C_t, V_t) = E\left[\widetilde{g}(I_t) \mid C_t, V_t\right].$$

With this notation, regardless of whether habits are internal or external, we may write the first order conditions associated with the Lagrangean (14) as

$$\begin{aligned} \lambda_t &= b^t \widetilde{g}(I_t) \\ \lambda_t &= E\left[\lambda_{t+1} R_{jt+1} \mid I_t\right] \qquad j = 1, \ldots, J \\ \lambda_t P_t &= b^t g_d(C_t, V_t) - \delta E\left[\lambda_{t+1} P_{t+1} \mid I_t\right] \end{aligned}$$

Using the consumption equation $\lambda_t = b^t \widetilde{g}(I_t)$ to remove the Lagrangeans in the assets and durables first order conditions gives

$$\begin{aligned} b^t \widetilde{g}(I_t) &= E\left[b^{t+1} \widetilde{g}(I_{t+1}) R_{jt+1} \mid I_t\right] \qquad j = 1, \ldots, J \\ b^t \widetilde{g}(I_t) P_t &= b^t g_d(C_t, V_t) - \delta E\left[b^{t+1} \widetilde{g}(I_{t+1}) P_{t+1} \mid I_t\right]. \end{aligned}$$

Taking the conditional expectation of the asset equations, conditioning on C_t, V_t, yields the Euler equations for asset j

$$g(C_t, V_t) = bE\left[g(C_{t+1}, V_{t+1}) R_{jt+1} \mid C_t, V_t\right] \qquad j = 1, \ldots, J, \tag{15}$$

for all t. Therefore, given the pair (U, b) of utility function and discounting factor the optimal decision satisfies the Euler equations for all asset j.

10.2 Preliminary Lemmas

The following lemma draws heavily on Einmahl and Mason (2005). We denote by $\psi \equiv (\varphi, c, v)$ a generic element of the set $\Psi \equiv \mathcal{G} \times T$. Let $f(c, v)$ denote the density of (C, V) evaluated at (c, v). Define the regression function $m(\psi) \equiv E[\varphi(C', V')R' | C = c, V = v]$. Then, an estimator for $m(\psi)$ is given by

$$\widehat{m}_h(\psi) = \frac{1}{nh^\ell \widehat{f}(c, v)} \sum_{i=1}^{n} \varphi(C_i', V_i') R_i' K\left(\frac{c - C_i}{h}\right) \prod_{j=1}^{\ell_1} K\left(\frac{v_j - V_{ji}}{h}\right) \equiv \frac{\widehat{T}_h(\psi)}{\widehat{f}(c, v)}.$$

Henceforth, we abstract from measurability issues that may arise in $\sup_{g \in \mathcal{G}: \|g\| \leq 1} \left\| \widehat{A}g - Ag \right\|$ (see van der Vaart and Wellner (1996) for ways to deal with lack of measurability).

Lemma B1. *Suppose that Assumption A1 holds. Then,*

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} |\widehat{m}_h(\psi) - m(\psi)| = o_P(1). \tag{16}$$

If, in addition, A2 holds, then

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} |\widehat{m}_h(\psi) - m(\psi)| = O_P\left(\sqrt{\frac{\ln n}{n l_n^{\ell}}} + u_n^r\right). \tag{17}$$

Proof. By the Triangle inequality

$$\begin{aligned}
& |\widehat{m}_h(\psi) - m(\psi)| \\
\le\ & \left|\widehat{m}_h(\psi) - \frac{E[\widehat{T}_h(\psi)]}{E[\widehat{f}\,(c,v)]}\right| + \left|\frac{E[\widehat{T}_h(\psi)]}{E[\widehat{f}\,(c,v)]} - m(\psi)\right| \\
\le\ & \frac{1}{\left|\widehat{f}\,(c,v)\right|}\left|\widehat{T}_h(\psi) - E[\widehat{T}_h(\psi)]\right| + \frac{\left|E[\widehat{T}_h(\psi)]\right|}{\left|\widehat{f}\,(c,v)\, E[\widehat{f}\,(c,v)]\right|}\left|\widehat{f}\,(c,v) - E[\widehat{f}\,(c,v)]\right| \\
& + \frac{1}{\left|E[\widehat{f}\,(c,v)]\right|}\left|E[\widehat{T}_h(\psi)] - T(\psi)\right| + \frac{|T(\psi)|}{\left|E[\widehat{f}\,(c,v)] f\,(c,v)\right|}\left|E[\widehat{f}\,(c,v)] - f\,(c,v)\right|.
\end{aligned}$$

We shall apply a variation of Theorem 4 in Einmahl and Mason (2005) to obtain uniform rates for $\widehat{T}_h(\psi) - E[\widehat{T}_h(\psi)]$; the rates for $\widehat{f}\,(c,v) - E[\widehat{f}\,(c,v)]$ follow analogously and are simpler to obtain (see their Theorem 1, 1.3). Our conditions A1.2 and A1.4 imply the assumptions needed for their Theorem 4, where the bracketing conditions replace their covering conditions (see their Remark 3 and Lemma B.4 in Escanciano, Jacho-Chávez and Lewbel (2014)). Then, we conclude

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} \left|\widehat{T}_h(\psi) - E\left[\widehat{T}_h(\psi)\right]\right| = O_P\left(\sqrt{\frac{\ln n}{n l_n^{\ell}}}\right).$$

On the other hand, Lemma 2 in Einmahl and Mason (2005) and the uniform equicontinuity of $\mathcal{M}$ in Assumption A2.2 yield

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} \left|E\left[\widehat{T}_h(\psi)\right] - T(\psi)\right| = o(1),$$

where $T(\psi) \equiv m(\psi) f\,(c,v)$, and likewise for the density bias term. This together with the above expansion for $\widehat{m}_h - m$ completes the proof of (16).

To obtain rates for the bias terms we need the smoothness conditions of Assumption A2. A standard Taylor expansion argument, the higher-order property of the kernel and the uniform equicontinuity of the $r-th$ derivative of the class $\mathcal{M}$ imply that

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} \left|E\left[\widehat{T}_h(\psi)\right] - T(\psi)\right| = O\left(u_n^r\right),$$

and similarly for the density bias term. The proof is completed by standard arguments using the boundedness away from zero of $f(c,v)$ over the domain. ■

Lemma B2. *Suppose that Assumption A1 holds. Then, as $n \to \infty$:*

$$\left\| \widehat{A} - A \right\| = \sup_{g \in \mathcal{G}: \|g\| \leq 1} \left\| \widehat{A}g - Ag \right\| = o_P(1).$$

Proof. Follows from the definition of $\widehat{A}$ and the first part of Lemma B1. ■

We introduce a useful class of functions:

DEFINITION 4. *Let $\mathcal{L}^2(r)$ be the class of functions $\varphi \in \mathcal{L}^2$ such that $\Sigma_\varphi \equiv E[\varphi_i^2 \varepsilon_i^2] < \infty$ and φ is $r-$times continuously differentiable.*

Lemma B3. *Suppose that Assumptions A1 and A2 hold. Then, for any $\varphi \in \mathcal{L}^2(r)$, it holds that*

$$\sqrt{n} \left\langle \left(\widehat{A} - A \right) g_0, \varphi \right\rangle \xrightarrow{d} N(0, \Sigma_\varphi).$$

Proof. Define

$$\widehat{T} g_0(c,v) = \frac{1}{n} \sum_{i=1}^{n} g'_{0i} R'_i K_{hi}(c,v),$$

with $g'_{0i} \equiv g_0(C'_i, V'_i)$ and note that $\widehat{A} g_0(c,v) = \widehat{T} g_0(c,v) / \widehat{f}(c,v)$. Using standard arguments, we write

$$\left(\widehat{A} - A \right) g_0(c,v) = a_n(c,v) + r_n(c,v),$$

where

$$a_n(c,v) = f^{-1}(c,v) \left(\widehat{T} g_0(c,v) - T g_0(c,v) - A g_0(c,v) \left(\widehat{f}(c,v) - f(c,v) \right) \right),$$

$T g_0(c,v) \equiv f(c,v) A g_0(c,v), \widehat{T} g_0(c,v) \equiv \widehat{f}(c,v) \widehat{A} g_0(c,v)$ and

$$r_n(c,v) \equiv -\frac{\widehat{f}(c,v) - f(c,v)}{\widehat{f}(c,v)} a_n(c,v).$$

Lemma B1 and our conditions on the bandwidth imply $\|r_n\| = o_P(n^{-1/2})$. It then follows that $\left\langle \left(\widehat{A} - A \right) g_0, \varphi \right\rangle$ has the following expansion

$$\int \varphi(c,v)[\widehat{T} g_0(c,v) - T g_0(c,v)] dc dv \tag{18}$$

$$- \int \varphi(c,v) A g_0(c,v) [\widehat{f}(c,v) - f(c,v)] dc dv \tag{19}$$

$$+ o_P(n^{-1/2}).$$

We now look at terms (18)-(19). Firstly, it follows from standard arguments and A2.5 that the difference between $Tg_0(c,v)$ and $E[\widehat{T}g_0(c,v)]$ is $O_P(u_n^r) = o_P(n^{-1/2})$ by the condition $nu_n^{2r} \to 0$. Hence,

$$\int \varphi(c,v)[\widehat{T}g_0(c,v) - Tg_0(c,v)]dcdv = \int \varphi(c,v)[\widehat{T}g_0(c,v) - E(\widehat{T}g_0(c,v))]dcdv + o_P(n^{-1/2})$$
$$= \frac{1}{n}\sum_{i=1}^{n} g_{0i}' R_i' \int \varphi(c,v) K_{hi}(c,v)\, dcdv - \int \varphi(c,v) E(g_0' R_i' K_{hi}(c,v)) dcdv + o_P(n^{-1/2}),$$
$$= \frac{1}{n}\sum_{i=1}^{n} \varphi(C_i,V_i) g_{0i}' R_i' - E[\varphi(C_i,V_i)\, Ag_0(C_i,V_i)] + o_P(n^{-1/2}),$$

where the last equality follows from the standard change of variables argument and our Assumption A2. Likewise, the term (19) becomes $n^{-1/2}\sum_{i=1}^{n} \varphi(C_i,V_i) Ag_0(C_i,V_i) - E[\varphi(C_i,V_i)\, Ag_0(C_i,V_i)] + o_P(n^{-1/2})$. In conclusion, we have

$$\sqrt{n}\left\langle \left(\widehat{A} - A\right) g_0, \varphi \right\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varphi(C_i,V_i)\varepsilon_i + o_P(n^{-1/2}).$$

Then, the result follows from a standard central limit theorem, since $\{\varphi(C_i,V_i)\varepsilon_i\}_{i=1}^{n}$ is iid with zero mean and finite variance. ■

For a generic function $r \in \mathcal{L}^2$, define

$$r_s = r - \langle g_0, r\rangle \langle g_0, s\rangle^{-1} s.$$

Also for $r \in \mathcal{N}^{\perp}(L) = \mathcal{R}(L^*)$ denote by r^* the unique minimum norm solution of $r = L^* r^*$. Note that for $r \in \mathcal{R}(L^*)$, r_s^* does not depend on the solution r^* considered of $r = L^* r^*$ (whether or not is minimum norm). This follows because under our conditions $\mathcal{N}(L^*)$ is the linear span generated by s.

Lemma B4. *Let Assumptions S, C, I and A1-A2 hold. If $\varphi \in \mathcal{N}^{\perp}(L)$, so $\varphi = L^*\varphi^*$ for some φ^*, and if $\varphi_s^* \in \mathcal{L}^2(r)$, then*

$$\sqrt{n}\langle \widehat{g} - g_0, \varphi\rangle \xrightarrow{d} N\left(0, b_0^2 \Sigma_{\varphi_s^*}\right).$$

Proof. Note that by (20) below and the adjoint property

$$\begin{aligned}
\sqrt{n}\langle \widehat{g} - g_0, \varphi\rangle &= \sqrt{n}\langle \widehat{g} - g_0, L^*\varphi^*\rangle \\
&= \sqrt{n}\langle L(\widehat{g} - g_0), \varphi^*\rangle \\
&= -\sqrt{n}\left(\widehat{b} - b_0\right) b_0^{-1}\langle g_0, \varphi^*\rangle - b_0\sqrt{n}\left\langle (\widehat{A} - A) g_0, \varphi^* \right\rangle + o_P(1).
\end{aligned}$$

Then, by the proof of Theorem 4, this can be further simplified to

$$b_0\sqrt{n}\left\langle \left(\widehat{A}-A\right) g_0, s\left\langle g_0,\varphi^*\right\rangle-\varphi^*\right\rangle = -b_0\sqrt{n}\left\langle \left(\widehat{A}-A\right) g_0, \varphi_s^*\right\rangle + o_P(1).$$

Then, the result follows from the last display and Lemma B3. ■

10.3 Main Proofs

With some abuse of notation, denote by $\|\cdot\|$ the usual norm for linear bounded operators,

$$\|B\| = \sup_{g\in\mathcal{G}:\|g\|\leq 1} \|Bg\|.$$

The spectral radius $\rho(T)$ of a linear continuous operator T on a Banach space $\mathcal{X}$ is defined as $\sup_{\lambda\in\sigma(T)}|\lambda|$, where $\sigma(T)\subset\mathbb{C}$ denotes the spectrum of T. Any compact operator T has a discrete spectrum, so that $\sigma(T)$ is simply the set of eigenvalues of T. For more definitions and further details see Kress (1999, Chapter 3.2). The operator B is called positive if $Bg\in\mathcal{P}$ when $g\in\mathcal{P}$.

Proof of Theorem 1. By Assumption C the set of countable eigenvalues of A has zero as a limit point, and thus, the set of eigenvalues λ with $\lambda^{-1}\in(0,1)$ is a finite set. By Theorem 3.1 in Kress (1999) for each such eigenvalue there is a finite-dimensional eigenvector space. ■

Proof of Theorem 2. Let A^* denote the adjoint of A, which is also compact and positive by well known results in functional analysis. Assumption S implies that $\rho(A)>0$. Also notice that the eigenvalues of A^* are complex conjugates of those of A (in particular, $\rho(A)=\rho(A^*)$). Then, by the Krein-Rutman's theorem (see Theorem 7.10 in Abramovich and Aliprantis, 2002) the spectral radius $\rho(A)$ is an eigenvalue of A^* having a strictly positive eigenfunction $s(\cdot)$. But $\langle g,s\rangle = b\langle Ag,s\rangle = b\langle g,A^*s\rangle = b\rho(A)\langle g,s\rangle$. Hence, since g is nonnegative and s strictly positive, $\langle g,s\rangle\neq 0$, and then $b=\rho^{-1}(A)$. Assumption I implies that A is strongly expanding, using the terminology of Abramovich and Aliprantis (2002, Chapter 9)), and hence irreducible by Theorem 9.6 in the latter reference. Now, identification of g follows from Theorem V.5.2(i) in Schaefer (1974, p. 329) applied to $T=bA$. ■

Proof of Lemma 1. It is well known that in a complete metric space a set is relatively compact if and only if is totally bounded. Then, the compactness of A follows if we show that $\mathcal{R}(A)$ is totally bounded. Let $[l_j,u_j]$ be ε-brackets, $j=1,\ldots,N_\varepsilon\equiv N_{[\cdot]}(\varepsilon,\mathcal{G},\|\cdot\|)$, covering $\mathcal{G}$ with respect to $\|\cdot\|$. Assume without loss of generality that the kernel $k\geq 0$. Then, $[Al_j,Au_j]$, $j=1,\ldots,N_\varepsilon$, forms a set of $\|A\|$ ε-brackets covering $\mathcal{R}(A)$. Since $\|A\|<\infty$ it follows that $\mathcal{R}(A)$ is totally bounded. ■

Proof of Theorem 3. From well known inequalities (see e.g. Bosq, 2000, p. 103-104) we obtain:

$$\left|\widehat{b}^{-1}-b_0^{-1}\right| \leq \left\|\widehat{A}-A\right\|$$

$$\|\widehat{g}-\widetilde{g}\| \leq C\left\|\widehat{A}-A\right\|,$$

where C is a real positive number that depends only on b_0, $\widetilde{g} = sgn\left(\langle\widehat{g}, g_0\rangle\right) g_0/\left\|g_0\right\|_n$ (sgn is the sign function, i.e., $sgn(x) = 1(x > 0) - 1(x < 0)$). By Lemma B2, $\left\|\widehat{A}-A\right\| = o_P(1)$. Then, by the continuous mapping theorem $|\widehat{b} - b_0| = o_P(1)$. By Assumption A1.1, for large n, $\widetilde{g} = g_0/\left\|g_0\right\|_n$, and by the Law of Large Numbers and the normalization $\|g_0\| = 1$, it holds $\|\widetilde{g} - g_0\| = o_P(1)$. Hence, by the triangle inequality, $\|\widehat{g} - g_0\| = o_P(1)$. ∎

Proof of Theorem 4. By definition

$$\widehat{b}\widehat{A}\widehat{g} - b_0 A g_0 = \widehat{g} - g_0.$$

Write the left hand side of the last display as

$$\left(\widehat{b} - b_0\right) A\widehat{g} + b_0\left(\widehat{A} - A\right) g_0 + b_0 A(\widehat{g} - g_0) + \widehat{R},$$

where $\widehat{R} = \left(\widehat{b} - b_0\right)\left(\widehat{A} - A_o\right)\widehat{g} + b_0\left(\widehat{A} - A\right)(\widehat{g} - g_0)$. Then, after noticing that (by definition of s),

$$\langle b_0 A(\widehat{g} - g_0), s\rangle = \langle\widehat{g} - g_0, s\rangle,$$

we obtain

$$\left(\widehat{b} - b_0\right) b_0^{-1}\langle\widehat{g}, s\rangle + b_0\left\langle\left(\widehat{A} - A\right) g_0, s\right\rangle + \left\langle\widehat{R}, s\right\rangle = 0.$$

Assumption A2.5, Lemma B1, and Cauchy-Schwarz inequality yield

$$\begin{aligned}\left|\left\langle\widehat{R}, s\right\rangle\right| &\leq \left\|\widehat{R}\right\|\|s\| \\ &= O_P\left(\left\|\widehat{A} - A\right\|^2\right) \\ &= o_P(n^{-1/2}).\end{aligned}$$

Then, by continuity of the inner product, $\langle\widehat{g}, s\rangle \to_p \langle g_0, s\rangle \equiv 1$, and by Slutzky Theorem

$$\sqrt{n}\left(\widehat{b} - b_0\right) = -\sqrt{n}b_0^2\left\langle\left(\widehat{A} - A\right) g_0, s\right\rangle + o_P(1).$$

Hence, the result follows from Lemma B3. ∎

Proof of Theorem 5. Define the operators $L = b_0 A - I$, and its estimator $\widehat{L} = \widehat{b}\widehat{A} - I$. Then, by definition

$$\begin{aligned}0 &= \widehat{L}\widehat{g} - L g_0 \\ &= L(\widehat{g} - g_0) + (\widehat{L} - L) g_0 + (\widehat{L} - L)(\widehat{g} - g_0). \qquad (20)\end{aligned}$$

First, from previous results it is straightforward to show that

$$\left\|(\widehat{L}-L)(\widehat{g}-g_0)\right\| = o_P\left(\sqrt{nh_n^{\ell}}\right)$$

and

$$\left\|(\widehat{L}-L)g_0 - b_0(\widehat{A}-A)g_0\right\| = o_P\left(\sqrt{nh_n^{\ell}}\right).$$

Hence, in $\mathcal{L}^2$,

$$\begin{aligned}\sqrt{nh_n^{\ell}}L(\widehat{g}-g_0) &= -\sqrt{nh_n^{\ell}}b_0(\widehat{A}-A)g_0 + o_P(1) \\ &= -\sqrt{nh_n^{\ell}}b_0\Delta_n + o_P(1).\end{aligned}$$

■

Proof of Theorem 6. Set $\widehat{\zeta}(C_i,V_i) = -C_i\partial\widehat{g}(C_i,V_i)/\partial c/\widehat{g}(C_i,V_i)$, which estimates consistently $\zeta(C_i,V_i) = -C_i\left(\partial g_0(C_i,V_i)/\partial c\right)/g_0(C_i,V_i)$. Then, using standard empirical processes notation, write

$$\sqrt{n}\left(\gamma_n\left(\widehat{g}\right)-\gamma\left(g_0\right)\right) = \sqrt{n}\left(P_n\widehat{\zeta}-P\widehat{\zeta}\right) + \sqrt{n}\left(P\widehat{\zeta}-P\zeta\right).$$

By the P-Donsker property of $\mathcal{D}$, $P\left(\widehat{g}\in\mathcal{G}\right)\to 1$ and the consistency of $\widehat{g}$,

$$\sqrt{n}\left(P_n\widehat{\zeta}-P\widehat{\zeta}\right) = \sqrt{n}\left(P_n\zeta - P\zeta\right) + o_P(1).$$

Since $\widehat{g}-g_0$ is bounded with probability tending to one, we can apply integration by parts and use Assumption A3 to write

$$\begin{aligned}\sqrt{n}\left(P\widehat{\zeta}-P\zeta\right) &= \sqrt{n}\left\langle \log(\widehat{g})-\log(g_0), d\right\rangle + o_P(1) \\ &= \sqrt{n}\left\langle \widehat{g}-g_0, \chi\right\rangle + o_P(1),\end{aligned}$$

where the last equality follows from the Mean Value Theorem and the lower bounds on g and $\widehat{g}$. Note that $\chi\in\mathcal{N}^{\perp}(L)$, since $\langle g_0,\chi\rangle = E[d(C,V)] = 0$. Then, by Lemma B4

$$\sqrt{n}\left(P\widehat{\zeta}-P\zeta\right) = \frac{-b_0}{\sqrt{n}}\sum_{i=1}^{n}\chi_s^*(C_i,V_i)\varepsilon_i + o_P(1),$$

and therefore

$$\sqrt{n}\left(\gamma_n\left(\widehat{g}\right)-\gamma\left(g_0\right)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\zeta(C_i,V_i)-P\zeta\right) - b_0\chi_s^*(C_i,V_i)\varepsilon_i + o_P(1).$$

The result then follows from the Lindeberg-Levy central limit theorem and $E[\varepsilon_i|C_i,V_i]=0$. ■

References

[1] Ai, C. and X. Chen (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.

[2] Abramovich, Y. A. and Aliprantis, C. D. (2002). *An Invitation to Operator Theory*. Graduate Studies in Mathematics 50. American Mathematical Society.

[3] An, Y. and Y. Hu (2012), "Well-posedness of measurement error models for self-reported data," Journal of Econometrics, 168, 259–269.

[4] Anatolyev, S. (1999), "Nonparametric Estimation of Nonlinear Rational Expectation Models," *Economics Letters*, 62, 1-6.

[5] Andrews, D. W. K. (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory,* 11, 560–596.

[6] Banks, J., R. Blundell, and S. Tanner (1998), "Is There a Retirement-Savings Puzzle?" *The American Economic Review*, 88, 769-788.

[7] Battistin, E., R. Blundell, and A. Lewbel, (2009), "Why is consumption more log normal than income? Gibrat's law revisited," *Journal of Political Economy*, 117, 1140-1154.

[8] Bosq, D. (2000), *Linear Processes in Function Spaces*. Springer, New York.

[9] Cai, Z., Ren, Y. and L. Sun, (2015), "Pricing Kernel Estimation: A Local Estimating Equation Approach," *Econometric Theory*, 31, 560-580.

[10] Campbell, J. Y., and J. Cochrane, (1999), "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205-251.

[11] Carrasco, M. and J. P. Florens (2000), "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797-834.

[12] Carrasco, M., J.P. Florens and E. Renault (2007): "Linear Inverse Problems and Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, vol. 6, eds. J. Heckman and E. Leamer. North-Holland.

[13] Chanchana, P. (2007), "An Algorithm for Computing the Perron Root of a Nonnegative Irreducible Matrix" Ph.D. Dissertation, North Carolina State University, Raleigh.

[14] Chapman, D. A. (1997), "Approximating the Asset Pricing Kernel," *The Journal of Finance*, 52, 1383–1410.

[15] Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014), "Identification in Semiparametric and Nonparametric Conditional Moment Models," *Econometrica*, 82, 785-809.

[16] Chen, X., Hansen, L. P. and J. Scheinkman (2000), "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions," unpublished manuscript.

[17] Chen, X., Hansen, L. P. and J. Scheinkman (2009), "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions," *Annals of Statistics*, 37, 4279–4312.

[18] Chen, X. and S. C. Ludvigson (2009), "Land of addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, 24, 1057-1093.

[19] Chen, X. and D. Pouzo (2009), "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46-60.

[20] Chen, X. and M. Reiss (2010), "On Rate Optimality For Ill-Posed Inverse Problems In Econometrics," *Econometric Theory*, 27, 497-521.

[21] Christensen, T.M. (2014), "Nonparametric Stochastic Discount Factor Decomposition", unpublished manuscript.

[22] Christensen, T.M. (2015), "Nonparametric Identification of Positive Eigenfunctions", forthcoming, *Econometric Theory*.

[23] Cochrane, J. (2001). *Asset Pricing.* Princeton University Press.

[24] Darolles, S., J. P. Florens and C. Gouriéroux (2004): "Kernel-based Nonlinear Canonical Analysis and Time Reversibility," *Journal of Econometrics*, 119, 323-353.

[25] Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011), "Nonparametric Instrumental Regression," *Econometrica,* 79, 1541-1565.

[26] Deaton, A. (1992), Understanding Consumption Oxford: Oxford University Press

[27] Deaton, A. and C. Paxson, (1994). "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102, 437-467.

[28] Dunn, K. B. and K. J. Singleton, (1986) "Modeling the Term Structure of Interest Rates Under Non-Separable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27-55.

[29] Einmahl, J. H. J., and D. M. Mason (2005): "Uniform in Bandwidth Consistency of Kernel-Type Function Estimators," *Annals of Statistics*, 33, 1380–1403.

[30] Engl. H.W., M. Hanke, and A. Neubauer (1996), *Regularization of Inverse Problems*, Kluwer Academic Publishers.

[31] Escanciano, J. C. and S. Hoderlein (2012), "Nonparametric Identification of Euler Equations," unpublished manuscript.

[32] Escanciano, J. C., D. T. Jacho-Chávez and A. Lewbel (2014), "Uniform Convergence of Weighted Sums of Non and Semiparametric Residuals for Estimation and Testing," *Journal of Econometrics*, 178, 426-443.

[33] Fisher, F. (1966), The Identification Problem in Econometrics, New York: McGraw-Hill.

[34] Fleissig, A. R., A. R. Gallant, and J. J. Seater (2000), "Separability, Aggregation, and Euler Equation Estimation," *Macroeconomic Dynamics*, 4, 547-572.

[35] Gallant, A. R. and G. Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091-1120.

[36] Gayle, W.-R. and N. Khorunzhina (2014), "Micro-Level Estimation of Optimal Consumption Choice with Intertemporal Nonseparability in Preferences and Measurement Errors," Unpublished manuscript.

[37] Gobet, E., Hoffmann, M. and Reiss, M. (2004), "Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data," *Annals of Statistics*, 26, 2223-2253.

[38] Hall, R. E. (1978), 'Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,' *Journal of Political Economy*, 86, 971-987.

[39] Hall, P., and , J. L. Horowitz, (2005), "Nonparametric Methods for Inference in the Presence of Instrumental Variables" *Annals of Statistics*, 33, 2904-2929.

[40] Hall, P., Lee, Y. K., Park, B. U., and Paul, D. (2009), "Tie-respecting Bootstrap Methods for Estimating Distributions of Sets and Functions of Eigenvalues," *Bernoulli*, 15, 380-401.

[41] Hansen, L. P., (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[42] Hansen, L. P. and J. A. Scheinkman (2009): "Long-Term Risk: An Operator Approach," *Econometrica*, 77, 177-234.

[43] Hansen, L. P. and J. A. Scheinkman (2012): "Recursive Utility in a Markov Environment with Stochastic Growth," *Proceedings of the National Academy of Sciences*, 109, 11967-11972.

[44] Hansen, L. P. and J. A. Scheinkman (2013): "Stochastic Compounding and Uncertain Valuation," Working paper, University of Chicago.

[45] Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

[46] Härdle, W. and Mammen, E. (1993), "Comparing Nonparametric Versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926-1947.

[47] Hoderlein, S., Nesheim, L., and A. Simoni (2012), "Semiparametric Estimation of Random Coefficients in Structural Economic Models,", cemmap Working Papers, CWP09/12.

[48] Kreĭn, M. G. and M. A. Rutman (1950), *Linear Operators Leaving Invariant a Cone in a Banach Space*, American Mathematical Society, New York.

[49] Kubler, F. and K. Schmedders (2010): "Non-Parametric Counterfactual Analysis in Dynamic General Equilibrium," *Economic Theory*, 45, 181-200.

[50] Kress, R. (1999). *Linear Integral Equations.* Springer.

[51] Lawrance, E. C., (1991), "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, 99, 54-77.

[52] Lewbel, A. (1987), "Bliss Levels That Aren't," *Journal of Political Economy,* 95, 211-215.

[53] Lewbel, A. (1994), "Aggregation and Simple Dynamics," *American Economic Review,* 84, 905-918.

[54] Lucas, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429-1445.

[55] Luenberger, D. G. (1997). Optimization by Vector Space Methods. New York: John Wiley & Sons.

[56] Mankiw, N. G., (1982), "Hall's Consumption Hypothesis and Durable Goods," *Journal of Monetary Economics*, 10, 417-425.

[57] Newey, W. and J. Powell (2003), "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1557-1569.

[58] Ross, S. A. (2015): "The Recovery Theorem," *Journal of Finance*, 70, 615-648.

[59] Rothenberg, T. J. (1971). "Identification in parametric models," Econometrica, 39, 577-591.

[60] Sargan, J. D. (1983). "Identification and lack of identification." Econometrica, 51, 1605-1633.

[61] Schaefer, H.H. (1974). *Banach Lattices and Positive Operators*, Springer-Verlag, New York, Heidelberg, Berlin.

[62] Stock, J., M. Yogo and J. Wright (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518-529.

[63] Tamer, E. (2010). "Partial identification in econometrics." Annual Review of Economics, 2(1), 167-195.

[64] van der Vaart, A. W., and J. A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer Series in Statistics. Springer-Verlag, New York, 1 edn.