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

May 12, 2015

Abstract

We consider nonparametric identification and estimation of consumption based asset pricing Euler equations. This entails estimation of pricing kernels or equivalently marginal utility functions up to scale. The standard way of writing these Euler pricing equations yields Fredholm integral equations of the first kind, resulting in an ill posed inverse problem. We show that these equations can be written in a form that resembles Fredholm integral equations of the second kind, having well posed rather than ill posed inverses. We allow durables, habits, or both to affect utility. With few low level assumptions, we show that marginal utility functions and pricing kernels are locally nonparametrically identified, and we give conditions for point identification of these functions using shape restrictions. We propose a nonparametric estimator for the marginal utility and the discount factor that combines standard kernel estimation with the computation of a matrix eigenvector problem. The estimator is very easy to implement. We also investigate the associated limiting distributions for nonparametric and semiparametric functionals, and consider an empirical application to US household-level 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 and Xiaohong Chen, and seminar participants at University of Miami, UC San Diego and joint MIT-Harvard for helpful comments. All errors are our own. This version subsumes parts of the previous versions in "Nonparametric Euler Equation Identification and Estimation," by Lewbel, Linton and Srisuma (2011), and "Nonparametric Identification of Euler Equations," by Escanciano and Hoderlein (2010, 2012).

1 Introduction

The optimal intertemporal decision rule of an economic agent can often be characterized by a first-order condition known as Euler equation. 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. Indeed, one of the original motivations of the well-known generalized method of moments (GMM) estimator proposed by Hansen and Singleton (1982) was estimation of the Euler equations associated with rational expectations in consumption and associated consumption based asset pricing models. In this paper we study the nonparametric identification and estimation of Euler equations.

Nonparametric identification asks the fundamental question whether it is possible to learn the primitives of the model from an ideal data set. If nonparametric identification is not possible, then any empirical results have to depend on specific functional form assumptions. 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 possibly a vector of other economic variables such as durables or lagged consumption (for habits), R_t is the gross return of a risky or risk-free asset and g is the time homogeneous marginal utility 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] We take the primitives for the Euler equation to be the marginal utility function and the discount factor. The Euler equation is (nonparametrically) *identified* if: given the true joint distribution of the data, there is a unique (up to sign and scale normalization for g) solution to the equation (1) in $(g, b) \in \Theta := \mathcal{G} \times \mathcal{B}$, for a suitable parameter space Θ. We say the Euler equation is *partially identified* if the set of solutions is a proper subset of Θ, the so-called *identified set*, with at least two distinct elements (g_1, b_1) and (g_2, b_2) such that g_1 is not proportional to g_2 or $b_1 \neq b_2$.

In this paper we first show that, without further restrictions on utilities, Euler equations are generally not identified. Instead, they are partially locally identified under certain conditions, where partial identification is local in the sense that the identified set is countable so that each point in such set is isolated. We then show that Euler equations are identified if the class of utility functions are restricted to be strictly monotone, so that marginal utility is positive, which is a natural economic restriction. We propose a nonparametric kernel estimator for the marginal utility function and

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

discount factor using our identification arguments. We provide asymptotic distribution theory for the discount factor, marginal utility and for semiparametric functionals of the marginal utility such as the Average Relative Risk Aversion ($ARRA$) parameter. We illustrate the applicability of our methods with US household-level data from the Consumer Expenditure Survey (CEX).

For inference, from the empirical asset pricing literature, the Euler equation (1) is traditionally written in the following way:

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

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 the risky asset can be cast in the form of excess returns

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

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 (hence M_t) is 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 can also be specified semiparametrically, for instance Chen and Ludvigson (2009) assumes a functional form of g that depends on unknown habit functions of current and lagged consumption growth. More recently, these types of Euler equation models have been used as leading examples of nonparametric instrumental variables estimators, for instance see Newey and Powell (2003), which have the structure of Fredholm equations of the first kind (or Type I equation). Identification of a general nonlinear non/semi-parametric conditional moment restrictions is non-trivial. For instance, see Komunjer (2011), in an even simpler setting, where she considers identification of a parametric unconditional moment model, and see Chen, Chernozhukov, Lee and Newey (2014) for local identification conditions of semiparametric and nonparametric moment models.

The insight to our nonparametric identification results is the decomposition of the pricing kernel into the primitives, as in equation (1), that constitute a Fredholm linear equation of the second kind (or Type II equation). Then, the candidate discounting factor and marginal utility are characterized as an eigenvalue-eigenfunction pair of certain linear operator. Under the mild assumption that the operator is compact, a classical result (see e.g. Kress (1999)) ensures that the number of eigenvalue-eigenfunction pairs is countable, leading to our local identification result. To obtain point identification we restrict ourselves to positive marginal utilities, where we apply Kreĭn-Rutman's theorem, which extends the classical Perron-Frobenius theorem (that positive matrices have a unique positive eigenvalue that corresponds to a unique positive eigenvector) to compact operators in Banach spaces (see Schaefer (1974) and Abramovich and Aliprantis (2002) for a comprenhensive exposition of this theory). An attractive feature of our identification approach is that the solution to equation (1) generally has well-posed inverses, as opposed to solving an ill-posed inverse problem commonly associated with a Type I Fredholm equation.

Based on our identification arguments, we propose a new nonparametric estimator for marginal utilities and 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 very easy to implement. We establish its limiting distribution theory under weak and standard conditions, which are simpler than those associated with estimators that solve other structural ill-posed inverse problems.[2]

Given our estimates of g and hence of the pricing kernel M, we may test whether g is independent of durables consumption, lagged consumption, or both, thereby testing whether durable consumption or habit formation plays a role in determining the pricing kernel. We will also want to test various popular functional restrictions on utility such as parametric or homogeneity restrictions. In addition to the pricing kernel M, other functionals of the marginal utility function g that are of interest to estimate are the Arrow Pratt coefficients of Relative and of Absolute Risk Aversion (RRA and ARA, respectively),

$$RRA(c,v) = \frac{-c\partial g(c,v)/\partial c}{g(c,v)} \qquad \text{and} \qquad ARA(c,v) = \frac{-\partial g(c,v)/\partial c}{g(c,v)},$$

and marginal rates of substitution (MRS),

$$MRS(c,v) = \frac{\partial g(c,v)/\partial v}{\partial g(c,v)/\partial c}.$$

These measures are all independent of the scale of g. One might also be interested in the averages of these functions, corresponding to the Average RRA, ARA and MRS ($ARRA$, $AARA$ and $AMRS$, respectively). We illustrate below the applicability of our asymptotic results by establishing the asymptotic normality for a nonparametric estimator of the $ARRA$. The asymptotic theory of other semiparametric functional follows analogously from our asymptotic results.

The rest of the paper is organized as follows. After a literature review in Section 2, we provide sufficient conditions for partial and point identification in Section 3. We propose our kernel-type estimator in Section 4, and we investigate its asymptotic properties in Section 5. We report the results of a Monte Carlo experiment in Section 6. In Section 7, we apply our results to US household

[2] Examples of estimation of ill-posed inverse problems can be found in Carrasco and Florens (2000), Newey and Powell (2003), Ai and Chen (2003), Hall and Horowitz (2005), Chen and Ludvigson (2009), Chen and Pouzo (2009), Chen and Reiss (2010), and Darolles, Fan, Florens and Renault (2011), to mention but a few.

data. Finally, Section 8 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 SDF, 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 monotonicity of marginal utilities, and the asymptotic properties of their nonparametric estimators were not established.

Hansen and Scheinkman (2009, 2012, 2013) consider the related but different problem of identification of positive eigenfunctions and eigenvalues in a general continuous-time setting, using Markov theory and extensions of the classical Perron-Frobenius theorem.[3] Anatolyev (1999) provided a numerical method to estimate Euler equations that also makes use of the structure of the Type II Fredholm equations, but he did not discuss inference nor identification in his paper. Chen and Ludvigson (2009) studied identification and estimation of a semiparametric asset pricing model with habits, using an identification strategy based on Type-I equations. The present paper subsumes and extends parts of the previous papers by Escanciano and Hoderlein (2010, 2012) and Lewbel, Linton and Srisuma (2011), both using a Type II equation for identification. Escanciano and Hoderlein (2010, 2012) provided a theoretical analysis of point identification of positive marginal utilities in a model without habits, using the Kreĭn-Rutman's theorem. Here, we extend the nonparametric identification results of Escanciano and Hoderlein (2012) to the case of habits. After the previous versions of this paper were written, Chen, Chernozhukov, Lee and Newey (2014) used a Type-II equation and the Perron-Frobenius theorem to prove identification of the semiparametric asset pricing model of Chen and Ludvigson (2009). Ross (2015) has also recently applied the classical finite-dimensional Perron-Frobenius theorem to identify the pricing kernel and the natural probability distribution from state prices.

On the nonparametric estimation problem, the previous version of our paper, Lewbel, Linton and Srisuma (2011), proposed a combination of kernel and sieve nonparametric estimation, establishing rates and asymptotic distribution theory for nonparametric estimates of marginal utility and discount factor. Chen, Hansen and Sheinkman (2000, 2009), Darolles, Florens and Gouriéroux (2004) and

[3]Using our notation, their problem is to show existence and uniqueness of the positive eigenvalue of the operator $\phi \to E[M_{t+1}\phi(C_{t+1}, V_{t+1}) \mid C_t, V_t]$, taken as known the stochastic discount factor M_{t+1}. Here we consider identification of g and b (and hence of M_{t+1}). Christensen (2014, 2015) has recently applied identification results similar to those first obtained by Escanciano and Hoderlein (2010, 2012) to a discrete version of Hansen and Scheinkman (2009).

Carrasco, Florens and Renault (2007) discussed sieve estimation of related eigenvalue-eigenvector problems for self-adjoint operators. These results do not apply to our problem, since the associated operator is not self-adjoint. Christensen (2014) has recently proposed a nonparametric sieve estimator for the discrete 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. Our kernel estimator is very easy to implement and to analyze, as it does not require numerical integration and the estimator is amenable to standard kernel-based theory. In addition to the simplicity of the procedure, we establish new distribution theory for the nonparametric estimate of the marginal utility (positive eigenfunction), providing an expansion that is shown to be useful in semiparametric inference. This expansion allows us to show that to a large extent nonparametric inference on g is mathematically equivalent to inference on a standard conditional mean.

3 Identification

Since our goal is the study of Euler equations, we shall take as primitives the pair $(g, b) \in \Theta := \mathcal{G} \times \mathcal{B}$, where $\mathcal{G}$ denotes the parameter space of marginal utility functions, which satisfies some conditions below, and $\mathcal{B} = (0, 1]$. 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 the scale and sign normalization must be made (henceforth, we use the convention that for such pairs $g = g'$ unless otherwise stated). For the moment we shall assume there is just one asset, and we denote its rate of return by R_t. We shall discuss further below 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) squared 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 $\mathcal{G}$ be a linear subspace of $\mathcal{L}^2$, and define the linear operator $A : (\mathcal{G}, \|\cdot\|) \to (\mathcal{G}, \|\cdot\|)$ 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 $\mathcal{G}$ is choosen so that Ag is well-defined and $Ag \in \mathcal{G}$ for $g \in \mathcal{G}$. The requirement $\mathcal{G} \subset \mathcal{L}^2$ is not necessary for identification (see Escanciano and Hoderlein, 2012, Section 4) but it is made here for simplicity of presentation. We provide below an example of $\mathcal{G}$ satisfying our conditions. With our notation, (1) can be written in a compact form as $bAg = g$. We introduce the assumption of correct specification and a formal definition of identification.

ASSUMPTION S: There exists $(g, b) \in \Theta$ 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).

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).

Therefore the Euler equation is identified if Θ_0 is a singleton. Otherwise the Euler equation is not identified. To provide some insights on our identification and estimation strategies we consider first the case where A in (3) has a finite-dimensional range $\mathcal{R}(A) = \{f \in \mathcal{G} : \exists g \in \mathcal{G},\ Ag = f\}$. In that case, we can write

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

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

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

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 $b < 1$. 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.

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 $\mathcal{G}$ are those whose closure its compact). The compactness assumption although not necessary for identification (cf. Escanciano and Hoderlein, 2012), it is useful for the asymptotic theory of continuous functionals of g. It, however, rules out the case $\mathcal{G} = \mathcal{L}^2$ 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 no much loss of generality by following this "conditioning" approach, but for estimation of aggregated quantities, such as the $ARRA$, it is more convenient to use a "global" approach. Lemma 1 below provides sufficient conditions for compactness.

Assumption A1: $A : (\mathcal{G}, \|\cdot\|) \to (\mathcal{G}, \|\cdot\|)$ is a compact operator.

Theorem 1: *Under S and A1,* Θ_0 *is a finite set.*

Without further restrictions Theorem 1 claims that the Euler equation may generally be partially identified. Specifically, if the Euler equation is not identified, the Euler equation is *locally identified* in the sense that for any $\theta \in \Theta_0$ there is an open neighborhood of θ that does not contain any other element in Θ_0. Essentially, compactness of A ensures that Θ_0 is at most countable and the economic restriction that suggests the discounting factor to lie in $(0, 1]$ ensures Θ_0 is finite. Notice that the partial identification problem of the Euler equation can be reduced if there are multiple assets. Suppose there are J risky assets, then there are J Euler equations. If the linear operator for each asset is compact then the identified set for each asset is finite, the true (g, b) must therefore lie in the intersection of all identified sets. Likewise, the restriction that $bg(C_{t+1}, V_{t+1})R_{t+1} - g(C_t, V_t)$ is uncorrelated with any variable in the information set at time t may aid the identification and reduce the identified set.

Assumptions S and A1 do not suffice for point identification in general. We consider now shape restrictions on marginal utilities. Specifically, we shall impose that marginal utilities are positive. Let $\mathcal{P} := \{g \in \mathcal{G} : g \geq 0 \ \mu\text{−a.s.}\}$ denote the subspace of nonnegative functions in $\mathcal{G}$, and let $\mathcal{P}^+ := \{g \in \mathcal{G} : g > 0 \ \mu\text{−a.s.}\}$ denote the subspace of strictly positive functions, which is assumed to be non-empty. We introduce the following assumption.

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.

Theorem 2. *Let Assumptions S, A1 and I hold. Then,* $(b, g) \in (0, 1) \times \mathcal{P}$ *is identified.*

Identification of b follows under weaker conditions than those of Theorem 2, however we do not pursue these conditions here for simplicity. Moreover, from the proof we obtain $b = 1/\rho(A)$, 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).

Remark 1. *Theorem 2 holds under different sets of conditions. We could consider other sufficient conditions that replace conditions on A by conditions on a power of A, i.e. require Assumptions A1 and I for A^n, for some $n \geq 1$. It is hard to interprete these conditions, however, in a possibly non-Markovian environment.*

Remark 2. *The identification result in Theorem 2 suggests that the model is overidentified under the conditions of the theorem, as we are not exploiting the restriction that $bg(C_{t+1}, V_{t+1})R_{t+1} - g(C_t, V_t)$ is uncorrelated with any variable in the information set at time t different from (C_t, V_t).*

4 Estimation

4.1 Individual level-data

Our estimation strategy follows the identification strategy described above and it is also motivated from our application below. For estimation we assume that we have a 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 increasing time periods $t_1 \leq t_2 \leq \cdots \leq t_n$. To simplify notation denote $W_i = (R_i', C_i', V_i', C_i, V_i) \equiv (R_{t_i+1}, C_{t_i+1,i}, V_{t_i+1,i}, C_{t_i,i}, V_{t_i,i})$. Let the vector $W = (R', C', V', C, V)$ have the same distribution as $(R_i', C_i', V_i', C_i, V_i)$. For estimation we assume that the vector W is continuously distributed (the discrete case is simpler). Let $f(c, v)$ denote the Lebesgue density of (C, V), which might be different from the density of (C', V') (i.e. we allow for non-stationary consumption). We consider the setting described in the identification section where μ is the joint probability associated to f.

Consider the standard Nadaraya-Watson (NW) kernel estimator for the operator A

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

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

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

K is a univariate kernel function, h a bandwidth, $\ell = \ell_1 + 1$, $V_i = (V_{1i}, ..., V_{\ell_1 i})$ and

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

Note that contrary to A, the operator $\widehat{A}$ has a finite dimensional closed range (that is spanned by the functions $\phi_i(c, v)$, $i = 1, ..., n$). Therefore, 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. Let $\widehat{\lambda}$ and $\widehat{g}$ be an eigenvalue and eigenfunction of $\widehat{A}$, respectively. Any eigenfunction $\widehat{g}$ necessarily has the form $n^{-1} \sum_{i=1}^{n} \widehat{\beta}_i \phi_i(c, v)$, for some coefficients $\widehat{\beta}_i$, $i = 1, ..., n$, and the equation to solve becomes

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

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

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

Thus, $\widehat{\lambda}$ is the largest eigenvalue in modulus and $\widehat{\beta}$ the corresponding eigenvector of the $n \times n$ matrix A_n with ij-th element $a_{ij} = \phi_j(C_i', V_i') R_i' / n$. The eigenvector $\widehat{\beta} = (\widehat{\beta}_1, ..., \widehat{\beta}_n)'$ solves $A_n \widehat{\beta} = \widehat{\lambda} \widehat{\beta}$, and it is normalized so that

$$\widehat{\beta}' \Omega \widehat{\beta} = 1$$

and

$$n^{-1} \sum_{i=1}^{n} \widehat{\beta}_i \phi_i(c_0, v_0) > 0,$$

for some $(c_0, v_0) \in S$ and an $n \times n$ matrix Ω with entries

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

This normalization guarantees that the proposed kernel-type estimator

$$\widehat{g}(c, v) = n^{-1} \sum_{i=1}^{n} \widehat{\beta}_i \phi_i(c, v),$$

satisfies $\|\widehat{g}\|_n = 1$, where $\|g\|_n$ denotes the empirical norm of g, i.e.

$$\|g\|_n^2 = \frac{1}{n} \sum_{i=1}^{n} g^2(C_i, V_i).$$

Our estimate of the discount factor is $\hat{b} = 1/\widehat{\lambda}$. The estimator $(\widehat{g}, \hat{b})$ is easily implemented with any statistical package that computes eigenvalues and eigenvectors of matrices.

Notice that under very mild conditions the matrix A_n itself satisfies the classic conditions of the Perron-Frobenius theorem, which guarantees that $\hat{b} = \rho^{-1}(A_n)$ and $\widehat{\beta}$ is the only eigenvalue of 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 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

Here we provide conditions for consistency and limiting distribution theory for our estimators. We consider the following set of assumptions for our asymptotic results. Define the class of functions

$$\mathcal{M} = \{(c, v) \to E[g(C', V')R'|C = c, v = v] : g \in \mathcal{G}\}.$$

We need to 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).

Assumption A3:

1. Assumptions S and I hold. Also, assume $\|g\|^2 = 1$, $\|\widehat{g}\|_n^2 = 1$ and $\widehat{g}(c_0, v_0) > 0$ for $(c_0, v_0) \in S$.

2. $\{W_i\}_{i=1}^n$ is independent and identically distributed (iid).

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

4. There exists a convex and bounded subset $T \subseteq S$, with non-empty interior, $(c_0, v_0) \in T$, and such that $P((C', V') \in T|(C, V) \in T) = 1$. The probability density function $f(\cdot)$ is bounded, bounded away from zero and continuous on T.

[4] I.e.

$$\lim_{\delta \to 0} \sup_{|(c,v)-(c_1,z_1)|<\delta} \sup_{g \in \mathcal{G}} \|\mathbb{E}[g(C', V')R'|C = c, V = v] - \mathbb{E}[g(C', V')R'|C = c_1, V = v_1]\| = 0.$$

5. K is Lipschitz continuous, symmetric, with support $[-1, 1]$ and integrates to 1.

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

Condition A3.1 implies correct specification and point-identification in the space of nonnegative marginal utilities. The conditions $\|g\|^2 = 1$, $\|\widehat{g}\|_n^2 = 1$ and $\widehat{g}(c_0, v_0) > 0$ are free normalizations. Assumption A3.2 is a good approximation for our household-level data. This assumption could be relaxed to weakly dependent data using, for example, the results in Andrews (1995) to obtain the uniform rates of our Lemma A1 below. Other proofs remain the same under weakly dependent data. Some assumptions, such as A3.3 require existence of certain moments. Marginal utilities may not have finite moments around zero (where they typically diverge). To overcome this problem, we may rewrite equation (1) as

$$bE[C_{t+1}g(C_{t+1}, V_{t+1})(C_t/C_{t+1}) R_{t+1} \mid C_t, V_t] = C_t g(C_t, V_t), \tag{5}$$

and reparametrized the problem in terms of $C_{t+1}g(C_{t+1}, V_{t+1})$, which under natural economic assumptions is bounded; see Lucas (1978).

Examples of classes satisfying A3.3 abound in the literature; see van der Vaart and Wellner (1996). For example, we could take $\mathcal{G}$ as the following smooth class. For any vector a of ℓ integers define the differential operator $\partial_x^a := \partial^{|a|_1}/\partial x_1^{a_1} \ldots \partial x_\ell^{a_\ell}$, where $|a|_1 := \sum_{i=1}^{\ell} a_i$. 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 than η, and

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

Further, let $C_M^\eta(T)$ be the set of all continuous functions $h : T \subset \mathbb{R}^\ell \to \mathbb{R}$ with $\|h\|_{\infty,\eta} \leq M$. Since the constant M is irrelevant for our results, we drop the dependence on M and denote $C^\eta(T)$. Then, it is known that $\log N_{[\cdot]}(\varepsilon, C^\eta(T), \|\cdot\|) \leq C\varepsilon^{-v_s}$, $v_s = \ell/\eta$, so if $\mathcal{G} \subset C^\eta(T)$, then $\ell < 2\eta$ suffices for the bracketing condition in A3.3. It is clear that $\mathcal{G} \subset \mathcal{L}^2$. Under suitable smoothness conditions on the density of W, $A\mathcal{G} \subset \mathcal{G}$ holds. A3.3 is used here to control the term $\sup_{g \in \mathcal{G}} \left\| \widehat{A}g - Ag \right\|$ and also to guarantee that A is compact, as the following result shows.

LEMMA 1. *Assumption A3.3 implies that A is compact.*

We could take T equals to S in Assumption A3.4, but this is not necessary. Under Assumption A3.4 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. The assumption of densities bounded away from zero is standard in the nonparametric and semiparametric literatures, and it can be relaxed at the cost of longer proofs by introducing a vanishing random trimming, see e.g. Escanciano, Jacho-Chavez and Lewbel (2014). In A3.5 we can also use kernels with unbounded support that satisfy some smoothness and integrability conditions. Finally, notice A3.6 allows for data-driven bandwidth choices, which are common in applied work.

THEOREM 3. *Let Assumption A3 hold. Then,* $\left|\hat{b}-b\right| \rightarrow_p 0$ *and* $\|\widehat{g}-g\| \rightarrow_p 0$.

REMARK 3. *We can actually obtain from our results the uniform convergence of* $\widehat{g}$, *i.e.* $\|\widehat{g}-g\|_\infty \rightarrow_p 0$, *where* $\|h\|_\infty = \sup_{(c,v)\in T} |h(c,v)|$.

To obtain the asymptotic distribution theory for our estimators, we impose the following additional assumptions and notation. Simple algebra shows that the adjoint operator of A 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, $g_2 \in \mathcal{G}$,

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

Note that b^{-1} is also an eigenvalue for A^* (eigenvalues of A^* are complex conjugates of those of A). Similarly as we did for g, it can be shown that under Assumption A3 there exists a unique (up to scale) strictly positive eigenfunction of A^* associated to b^{-1}.

DEFINITION 3: Let s be the unique strictly positive eigenfunction of A^* with eigenvalue b^{-1} and satisfying $\langle g,s\rangle = 1$.

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

$$\varepsilon_i = g\left(C_i',V_i'\right)R_i' - b^{-1}g\left(C_i,V_i\right), \qquad i=1,...,n. \tag{6}$$

Henceforth, to simplify notation, define $\varphi_i = \varphi\left(C_i,V_i\right)$. 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 $\Sigma_\varphi := E\left[\varphi_i^2\varepsilon_i^2\right] < \infty$ and φ is $r-$times continuous differentiable.

ASSUMPTION A4:

1. The r-th derivatives of $f(\cdot)$ are uniformly continuous and bounded on T.
2. Functions in $\mathcal{M}$ are $r-$times continuous differentiable with uniformly equicontinuous $r-th$ derivative on T.
3. K is an $r-$th order kernel that is Lipschitz continuous symmetric with support $[-1,1]$.[5]
4. It holds that $l_n^{2\ell}n/\log n \to \infty$ and $nu_n^{4r} \to 0$ as $n \to \infty$.
5. $s(\cdot) \in \mathcal{L}^2(r)$.

THEOREM 4. *Under Assumptions A3 and A4, as $n \to \infty$,*

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

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 resampling methods (e.g. bootstrap or subsampling).

Our next result establishes an asymptotic expansion for $\widehat{g} - g$. This expansion can be used to obtain rates for $\widehat{g} - g$ and to establish asymptotic normality of (semiparametric) functionals of $\widehat{g}$. For example, we use this expansion below to obtain the asymptotic normality of the $ARRA$ parameter

$$\theta_0 = E\left[\frac{-C\partial g(C,V)/\partial c}{g(C,V)}\right]. \tag{7}$$

Define the process $\Delta_n(c,v) \equiv n^{-1}\sum_{i=1}^n \varepsilon_i\phi_i(c,v)$, where recall $\phi_i(c,v) = K_{hi}(c,v)/\widehat{f}(c,v)$. Note that a standard result in kernel estimation is that for all (c,v) in the interior of S, under suitable conditions,

$$\sqrt{nh_n^{\ell}}\Delta_n(c,v) \xrightarrow{d} N\left(0, \Sigma_0(c,v)\right),$$

with $\Sigma_0(c,v) = f^{-1}(c,v)\sigma^2(c,v)\kappa_2$, $\kappa_2 = \int K^2(u)du$ and $\sigma^2(c,v) = E\left[\varepsilon_i^2 | C_i = c, V_i = v\right]$.

Define the operator $L = bA - I$ and its adjoint $L^* = bA^* - I$. Let $\mathcal{R}(L)$ and $\mathcal{N}(L)$ denote the range and the kernel of the operator L, i.e. $\mathcal{R}(L) = \{f \in \mathcal{L}^2 : \exists s \in \mathcal{L}^2,\ Ls = f\}$ and

[5] So K is an even function 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 $\int u^r K(u)\, du > 0$.

$\mathcal{N}(L) = \{f \in \mathcal{L}^2 : Lf = 0\}$, and let $L^\dagger$ denote its Moore-Penrose pseudoinverse (see Engl, Hanke and Neubauer 1996, p. 33). Notice that under identification, $\mathcal{N}(L) = \{\kappa g : \kappa \in \mathbb{R}\}$. For a subspace $V \subset \mathcal{L}^2$, $V^\perp$ and $\overline{V}$ denote, respectively, its orthogonal complement and closure, with respect to the norm topology, in $\mathcal{L}^2$. It is well-known that $\mathcal{N}^\perp(L) = \overline{\mathcal{R}(L^*)}$. By the compactness of A and the Second Riesz Theorem, see e.g. Theorem 3.2 in Kress (1999, p. 29), the subspace $\mathcal{R}(L)$ is closed. This in turn implies that $L^\dagger$ is a continuous operator (see 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 inverses.

THEOREM 5. *Let Assumptions A3 and A4 hold. Then, in* $\mathcal{L}^2$,

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

This result implies that the rates of convergence of $\widehat{g} - g$ in $\mathcal{L}^2$ are the same as the NW kernel estimator of $E\left[\varepsilon_i | C_i = c, V_i = v\right]$. Combined with standard kernel regression results, it also implies the asymptotic normality for $\sqrt{nh_n^\ell}L\left(\widehat{g} - g\right)$, which can be used to make inference on g.

We now consider estimation of the $ARRA$ parameter θ_0 and establish the asymptotic normality of the estimator. The natural estimator of θ_0 is the sample analog based on our estimator $\widehat{g}$

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

Note that under our conditions on the kernel $\widehat{g}$ is differentiable and positive with probability tending to one, so $\widehat{\theta}$ is well-defined.

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

$$d(c, v) := \frac{\partial\left(c \times f(c, v)\right)}{\partial c}\frac{1}{f(c, v)} \qquad \text{and} \qquad \varphi(c, v) := \frac{d(c, v)}{g(c, v)}. \tag{8}$$

Also, we need to introduce some generic notation to be used in the asymptotic normality of $\widehat{\theta}$. For a function $r \in \mathcal{L}^2$, define

$$r_s = r - \langle g, r\rangle \langle g, s\rangle^{-1} s. \tag{9}$$

The function r_s has a geometrical interpretation, $r_s = P_{gs}r$, where the linear transformation P_{gs} is a projection. It projects parallel to s on a subspace of functions orthogonal to g. Similarly, for $r \in \mathcal{N}^\perp(L) = \mathcal{R}(L^*)$, we denote by r^* the minimum norm solution of $r = L^* r^*$. The influence function of the $ARRA$ estimator is given by

$$\xi_i = \left(\psi_i - E\left[\psi_i\right]\right) - b\varphi_s^*(C_i, V_i)\varepsilon_i, \tag{10}$$

where $\psi_i = -C_i \partial g(C_i, V_i)/\partial c / g(C_i, V_i)$. The second term in ξ_i comes from the estimation effect due to estimating g, and depends on φ_s^* which solves $L^* \varphi_s^* = \varphi_s$, where φ_s applies (9) to the score in (8).

Assumption A5: (i) $P(\widehat{g} \in \mathcal{G}) \to 1$ as $n \to \infty$ and the class $\mathcal{D}$ is P-Donsker; (ii) $S = [l_c, u_c] \times S_V$, and $\lim_{c \to l_c} cf(c, v) = 0 = \lim_{c \to u_c} cf(c, v)$ for all $v \in S_V$; (iii) Furthermore, $d \in \mathcal{L}^2$, $\mathbb{E}[|\xi_i|^2] < \infty$ and $\varphi_s^* \in \mathcal{L}^2(r)$.

Assumption A5(i) 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 $\mathcal{G}$ satisfyng the P-Donsker property of the second part of Assumption A5(i).

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

Assumption A5(ii) is similar to other assumptions required in estimation of average derivatives, see Powell, Stock and Stoker (1989). Assumption A5(iii) implies that the asymptotic variance of $\widehat{\theta}$ is finite.

Theorem 6. *Let Assumptions A3-A5 hold. Then,*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, E\left[\xi_i^2\right]\right),$$

where ξ_i is defined in (10).

Estimating the asymptotic variance of $\widehat{\theta}$ by plug-in methods is complicated. In our application we use bootstrap. In the empirical application we discuss other semiparametric functionals of $\widehat{g}$, whose asymptotic normality can be established along the same lines as those used for $\widehat{\theta}$. These semiparametric functionals can be used to test hypotheses such as testing the significance of habits or the homogeneity of marginal utilities.

6 Monte Carlo Experiment

In this section we illustrate the finite-sample performance of our estimator described in the previous section based on a Constant RRA ($CRRA$) utility function so that $g(x) = x^{-\gamma}$, where γ represents the $ARRA$ parameter. We simulate data to satisfy the following Euler equation:

$$bE\left[R_{t+1} C_{t+1}^{-\gamma} | C_t\right] = C_t^{-\gamma},$$

by choosing a joint distribution for (C_t, C_{t+1}) and calibrate R_{t+1} for some (b, γ). Our basic design sets $(b, \gamma) = (0.95, 0.5)$. We use $(\log C_t, \log C_{t+1}) \sim N\left(0, \begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{pmatrix}\right)$ and define R_{t+1} by $b^{-1}(1+\epsilon_t)(C_{t+1}/C_t)^{\gamma}$, where ϵ_t is distributed uniformly on $[-0.5, 0.5]$ and independently of (C_t, C_{t+1}). We choose this artificial design for the Monte Carlo study to abstract away from approximation and other numerical errors that are typically associated with solving dynamic optimization problems numerically.

For each set of experiments we generate 1000 replications of such random sample with sample size $n \in \{100, 500, 1000, 2500\}$. We compute our nonparametric estimators, and compare them to the correctly specified parametric $CRRA$ fit using a just-identified method of moments estimator (using C_t as an instrument). We undersmooth our nonparametric estimator to ensure the bias is negligible when computing the bootstrap standard errors and coverage probabilities by setting the bandwidth to be $1.06sn^{-1/3}$, where s is the sample standard deviation of C_t. All of our estimators for g are normalized to have a unit norm with respect to the empirical L^2 norm as explained above.

Tables 1 and 2 give the results for the discount factor and the $ARRA$ parameter, respectively. For each parameter we report the bias, median of the bias, Monte Carlo standard deviation, (Efron) bootstrapped standard error, 95% coverage probabilities based on bootstrapping, and root-mean squared error of their estimators. The results suggest that our estimators for the discount factor and the $ARRA$ are consistent. The bootstrapped coverage probabilities perform well, especially for the $ARRA$ parameter. In Table 3, we report the integrated mean squared errors for the nonparametric estimator of the marginal utility. Our nonparametric estimator $\widehat{g}$ appears to be L^2-consistent.

	n	Bias	MdBias	Std	BStd	BCv	Rmse
$CRRA$	100	0.000	-0.001	0.029	0.029	0.948	0.029
	500	0.000	0.000	0.012	0.012	0.953	0.012
	1000	0.000	0.000	0.009	0.009	0.940	0.009
	2500	0.000	0.000	0.005	0.005	0.946	0.005
NP	100	-0.012	-0.006	0.051	0.045	0.919	0.052
	500	-0.005	-0.002	0.030	0.020	0.928	0.030
	1000	-0.004	-0.001	0.027	0.017	0.926	0.027
	2500	-0.003	-0.001	0.022	0.010	0.930	0.022

Table 1: Summary statistics of the estimator for b. 1000 MC

	n	Bias	MdBias	Std	BStd	BCv	Rmse
$CRRA$	100	0.004	0.003	0.110	0.114	0.956	0.110
	500	0.000	0.000	0.044	0.046	0.948	0.044
	1000	0.001	0.002	0.033	0.033	0.944	0.033
	2500	0.000	0.000	0.020	0.021	0.946	0.020
NP	100	-0.062	-0.069	0.283	0.291	0.933	0.289
	500	-0.031	-0.032	0.113	0.119	0.943	0.117
	1000	-0.025	-0.023	0.084	0.093	0.952	0.088
	2500	-0.018	-0.016	0.060	0.059	0.964	0.063

Table 2: Summary statistics of the estimator for the $ARRA$. 1000 MC

n	100	500	1000	2500
Imse	0.0080	0.0022	0.0015	0.0006

Table 3: L^2-Mean squared errors of the estimator for g.

To get more insights into the nonparametric estimation for the marginal utility function, Figures 1 - 4 below provide the plot of these nonparametric curves for $n = 100, 500, 1000, 2500$ respectively, with consumption on the horizontal axis and marginal utilities on the vertical; the black line is the true and the blue lines are the estimates and 95% bootstrap confidence bands.

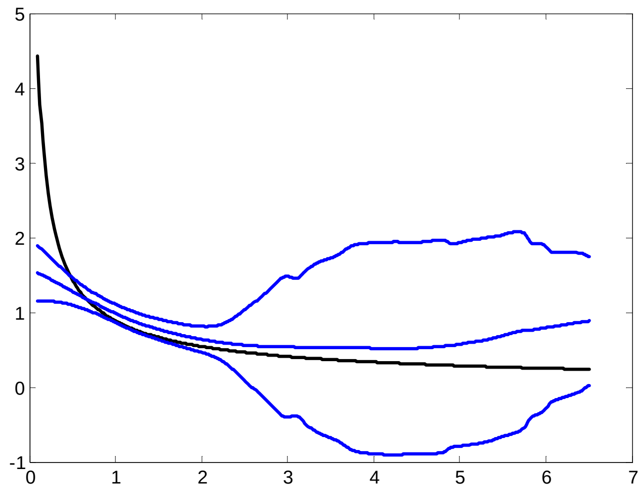

Figure 1: $n = 100$

As expected, the nonparametric fit improves with the sample size. The nonpametric estimator, however, is unable to capture the explosive behaviour of the marginal utility in the left tail. To overcome this limitation, we implement the transformation in (5). Figures 5-6 plot the nonparametric marginal utilities estimates applying the suggested reparametrization. It is evident that the fit at the boundary improves considerably. Nevertheless, unreported simulations suggest little or no benefits in the finite sample performance of semiparametric functionals such as the $ARRA$ parameter based on the transformation (i.e. corresponding to Tables 1 and 2), which might be explained by the low frequency of observations at the boundary.

7 Empirical Illustration

We use quarterly CEX household-level data for households sampled between 1980Q1 and 2012Q1. Our consumption data is based on real nondurables that we deflate using CPI from year 2000 as base, which we then scale by the inverse of the household size. In order to look at a relatively homogenous sample, we only consider households that are in urban areas for those in the age group between 30 and 50 years of age with an education level at high school diploma or higher. We only consider households that report four consecutive periods of consumption. The dataset contains 18912 households. We construct two types of asset returns, 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

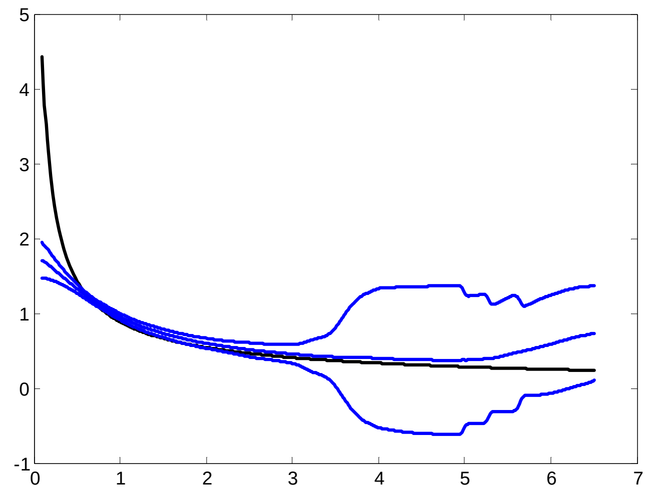

Figure 2: $n = 500$

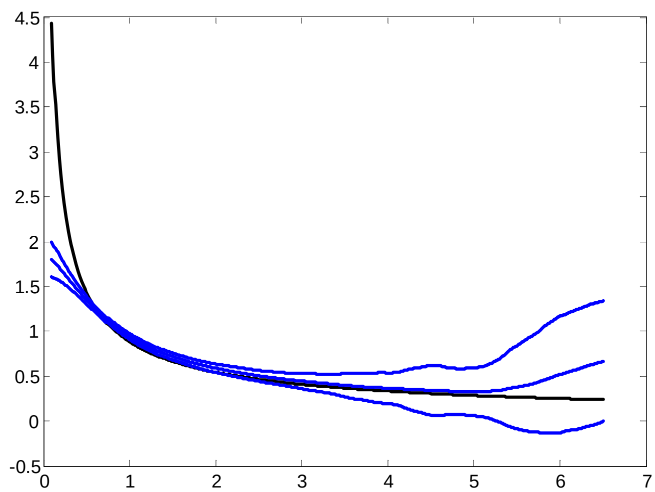

Figure 3: $n = 1000$

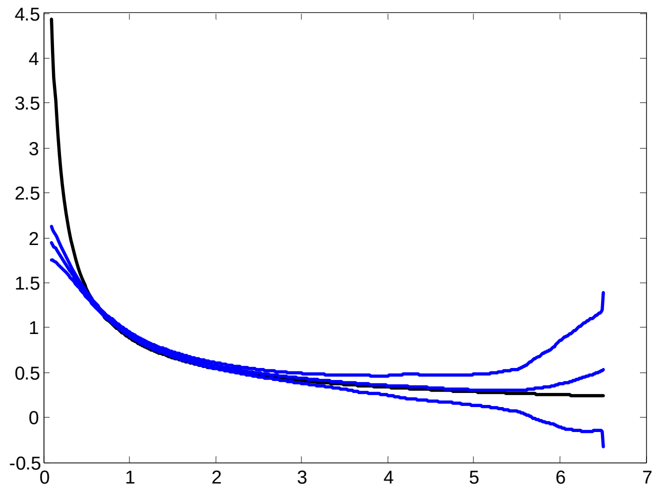

Figure 4: $n = 2500$

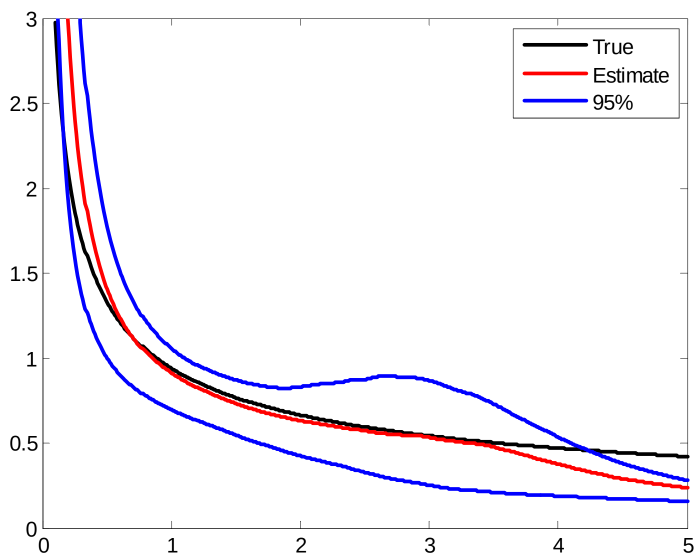

Figure 5: Nonparametric Marginal Utility (with transformation) when n=100

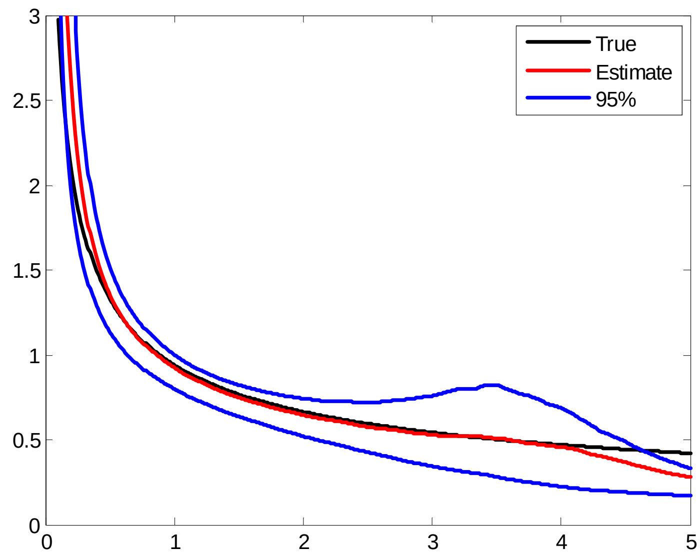

Figure 6: Nonparametric Marginal Utility (with transformation) when n=500

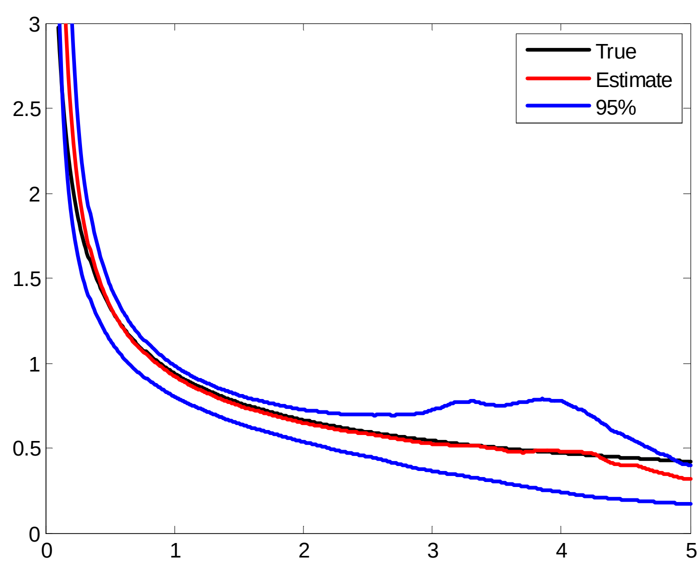

Figure 7: Nonparametric Marginal Utility (with transformation) when n=1000

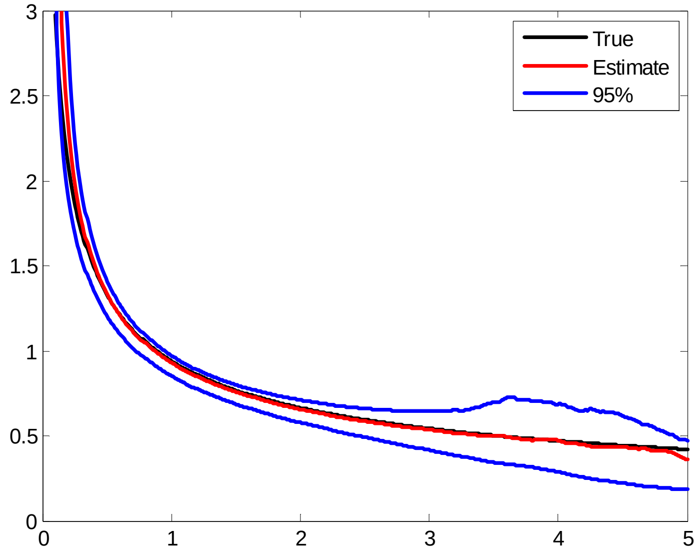

Figure 8: Nonparametric Marginal Utility (with transformation) when n=2500

index, with dividends reinvested. Both asset returns have been converted into real terms; computed on a quarterly basis.

Using each asset return we construct two set of estimates for the discount factor and various measures related to the risk aversion. We provide parametric estimates based on the $CRRA$ utility function.[6] We compare these with nonparametric estimates that allow for habit effects from the consumption level of the previous period. These can be found in Tables 4 and 5. All standard errors and confidence intervals in this section are computed using nonparametric bootstrap. We find that the estimates of the $ARRA$ for the parametric estimate to be unreasonably low, while the nonparametric ones are within the more conventionally expected range. Another surprising feature is the relative imprecision of the parametric estimators. This might be driven by the well documented identification problems of the $CRRA$ specification (this is a leading example in the literature of weak instruments, see for example Stock, Yogo and Wright (2002)). It also might be driven by misspecification of the parametric model. For example, we show below that there are clear indications that habit affects consumption behavior in the data. The nonparametric estimates are similar for risk-free and risky assets, which suggests that identification is achieved already with one asset. Having similar estimates over risky and risk-free is strong support for the model. The rate of time preference is known for being difficult to estimate precisely. Our estimates are somewhat low, relative to values commonly

[6] We use $(1, \frac{C_t}{C_{t-1}})$ as instruments. Including other lagged consumption ratios does not change the implications of our findings.

used in the literature, but they are reasonable. Note that others have found evidence that the rate of time preference can vary over the lifecycle and across income levels, e.g., Lawrence (1991).

		est	SE	LB	UB
Discount Factor	$CRRA$	0.9856	0.0005	0.9845	0.9866
	NP	0.9433	0.0017	0.9399	0.9465
$ARRA$	$CRRA$	0.0060	0.0072	-0.0070	0.0210
	NP	1.0248	0.0031	1.0185	1.0306

Table 4: Estimates of the discount factor and average relative risk aversion using risk free asset returns.

		est	SE	LB	UB
Discount Factor	$CRRA$	0.9662	0.0002	0.9658	0.9666
	NP	0.9243	0.0016	0.9213	0.9273
$ARRA$	$CRRA$	-0.0042	0.0028	-0.0096	0.0012
	NP	1.0242	0.0031	1.0178	1.0301

Table 5: Estimates of the discount factor and average relative risk aversion using risky asset returns.

In order to study whether the relative risk aversion for different parts of the population differ we compute local averages of relative risk aversion conditioning on whether the consumption levels at time t and time $t+1$ lie within particular quartiles of the observed distributions. That is, we estimate the parameters

$$E\left[RRA\left(C_{t+1}, C_t\right)|C_{t+1} \in I_i, C_t \in I_j\right],$$

where I_j denotes the interval between the $(j-1)-th$ and $j-th$ empirical quartile of consumption's distribution. These estimates can be found in Tables 6 and 7. The nonparametric estimates of RRA for different portions of the population show variation in risk attitude. For example we find that households that consume more in one period and less in the next tend to be associated with higher RRA value. All the RRA estimates lie within a relatively narrow range, roughly .9 to 1.1. The vertical variations in the estimated RRA imply that habit matters. If we focus on households whose consumption remained relatively stable over time (i.e. the diagonal in Tables 6 and 7), we observe that their RRA's appear to increase over the first three quartiles, and flatten out or possibly decrease in the last. The overall conclusion is that the standard assumption of constant RRA is good for this subpopulation, but there is evidence of some systematic departures from constant RRA overall.

	$C_{t+1} \in I_1$	$C_{t+1} \in I_2$	$C_{t+1} \in I_3$	$C_{t+1} \in I_4$
$C_t \in I_1$	0.9872	0.9064	-	-
	(0.0068)	(0.0111)	-	-
$C_t \in I_2$	1.1168	1.0337	0.9578	-
	(0.0118)	(0.0096)	(0.0076)	-
$C_t \in I_3$	-	0.9578	1.0421	0.9922
	-	(0.0076)	(0.0071)	(0.0042)
$C_t \in I_4$	-	-	1.1135	1.0174
	-	-	(0.0077)	(0.0019)

Table 6: Estimates of $ARRA$ using risk-free asset returns for different subsamples. Standard error in parentheses.

	$C_{t+1} \in I_1$	$C_{t+1} \in I_2$	$C_{t+1} \in I_3$	$C_{t+1} \in I_4$
$C_t \in I_1$	0.9873	0.9056	-	-
	(0.0068)	(0.0111)	-	-
$C_t \in I_2$	1.1160	1.0323	0.9573	-
	(0.0117)	(0.0096)	(0.0076)	-
$C_t \in I_3$	-	0.9573	1.0415	0.9919
	-	(0.0076)	(0.0071)	(0.0042)
$C_t \in I_4$	-	-	1.1130	1.0173
	-	-	(0.0077)	(0.0019)

Table 7: Estimates of $ARRA$ using risky asset returns for different subsamples. Standard error in parentheses..

We can also see the presence of habit effects visually and perform formal tests. Using risk free returns, Figures 5 - 7 plot the nonparametric estimates of $g(\cdot, c_0)$ and their confidence bands for c_0 taking the 1st, 2nd and 3rd quartiles of the observed consumption and Figure 8 plots all three curves together. Figures 9 - 12 are the risky returns counterparts. Figures 13 - 16 and Figures 17 - 20 are analogous plots for various functions of $RRA(\cdot, c_0)$ from risk-free and risky returns, respectively.

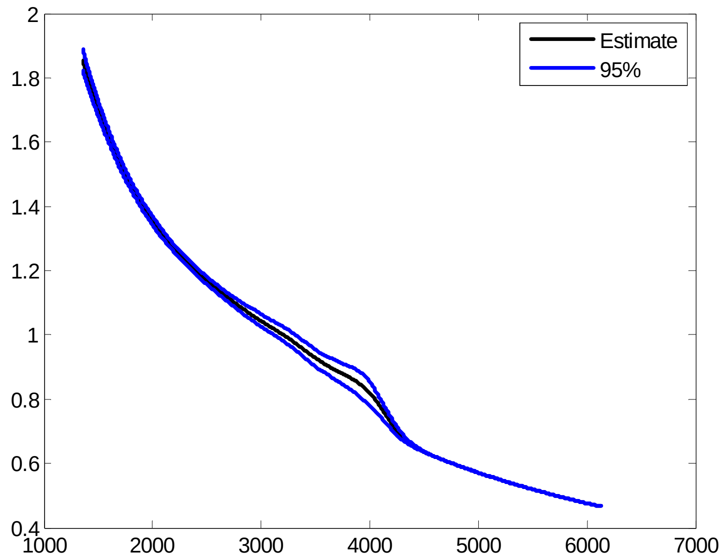

Figure 9: Marginal utility with risk free returns, previous level of consumption at the first quartile.

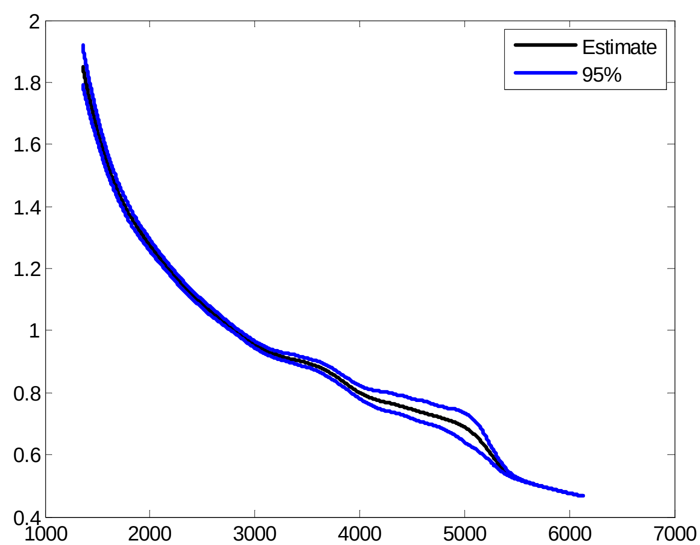

Figure 10: Marginal utility with risk free returns, previous level of consumption at the second quartile.

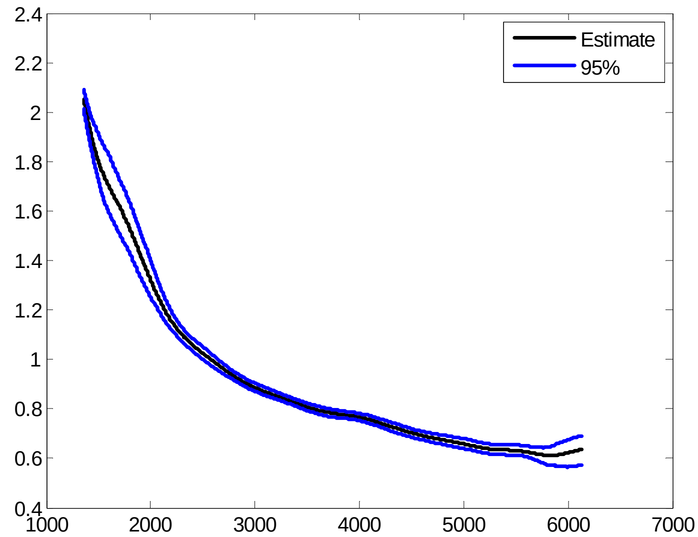

Figure 11: Marginal utility with risk free returns, previous level of consumption at the third quartile.

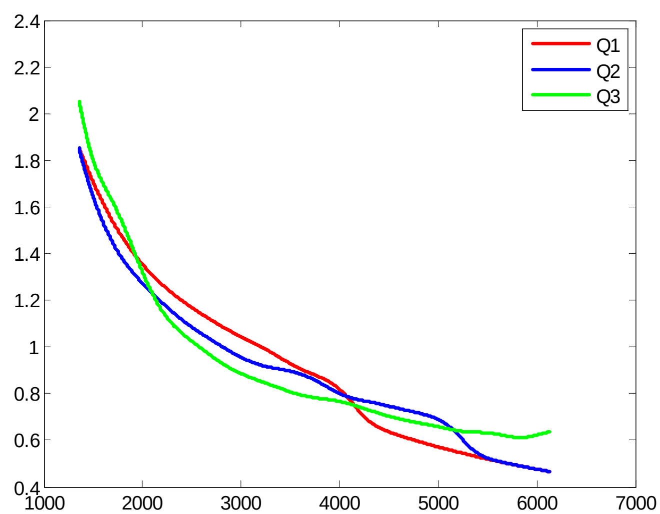

Figure 12: Marginal utility with risk free returns, previous level of consumption all three quartiles.

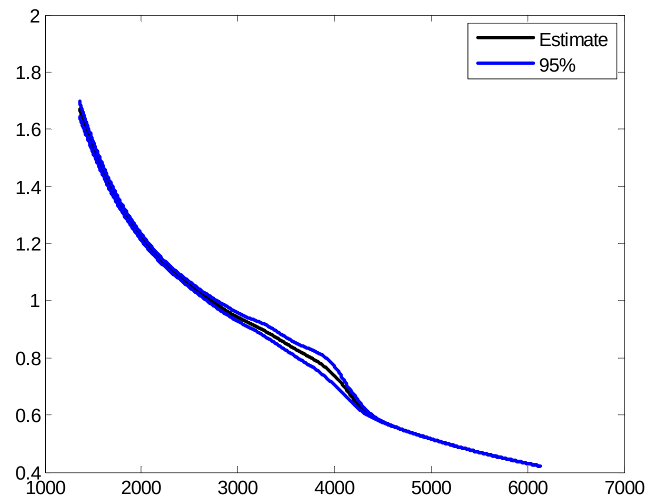

Figure 13: Marginal utility with risky returns, previous level of consumption at the first quartile.

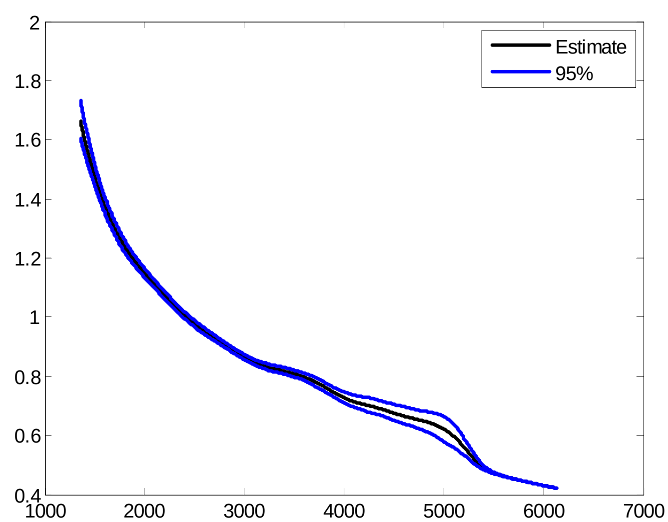

Figure 14: Marginal utility with risky returns, previous level of consumption at the second quartile.

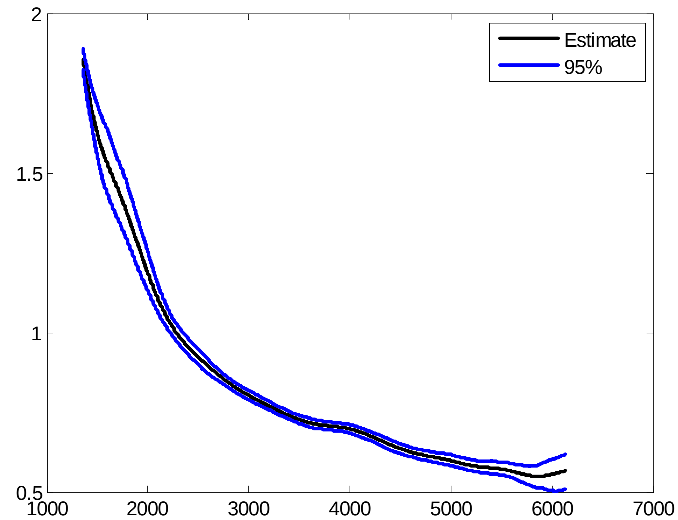

Figure 15: Marginal utility with risk free returns, previous level of consumption at the third quartile.

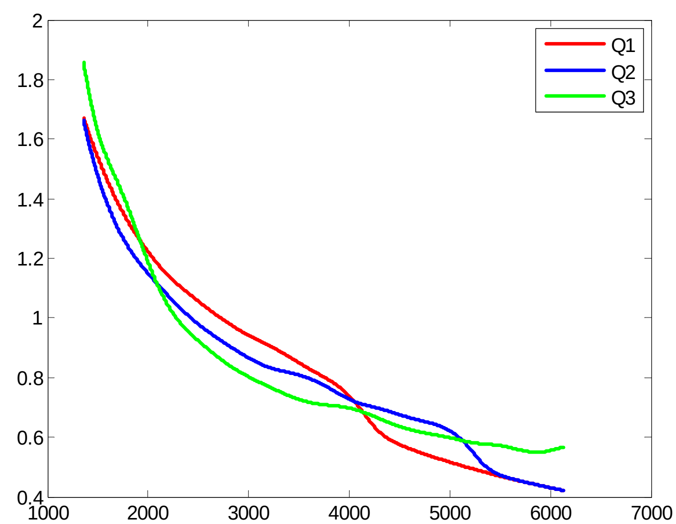

Figure 16: Marginal utility with risky returns, previous level of consumption all three quartiles.

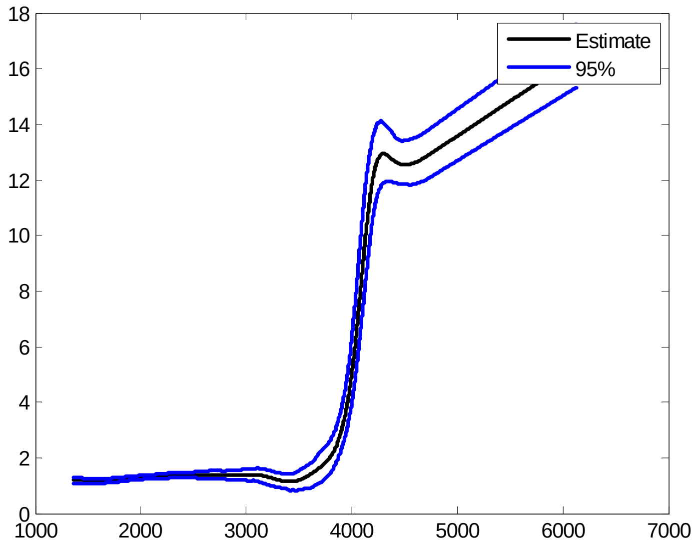

Figure 17: Average relative risk aversion with risk free returns, previous level of consumption at the first quartiles.

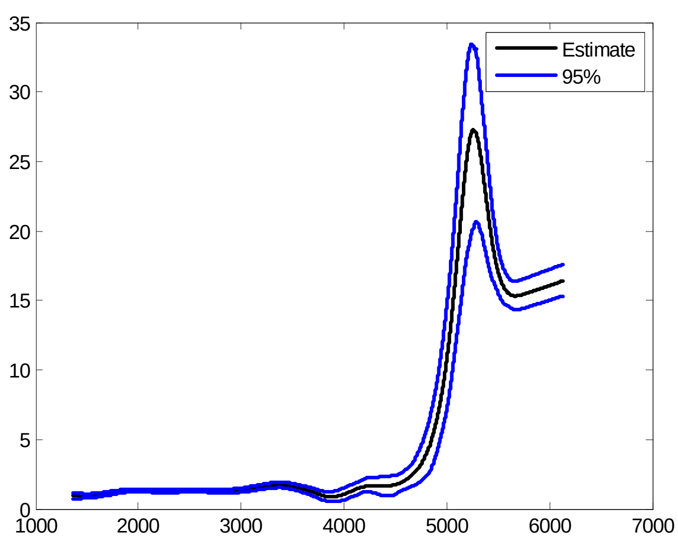

Figure 18: Average relative risk aversion with risk free returns, previous level of consumption at the second quartiles.

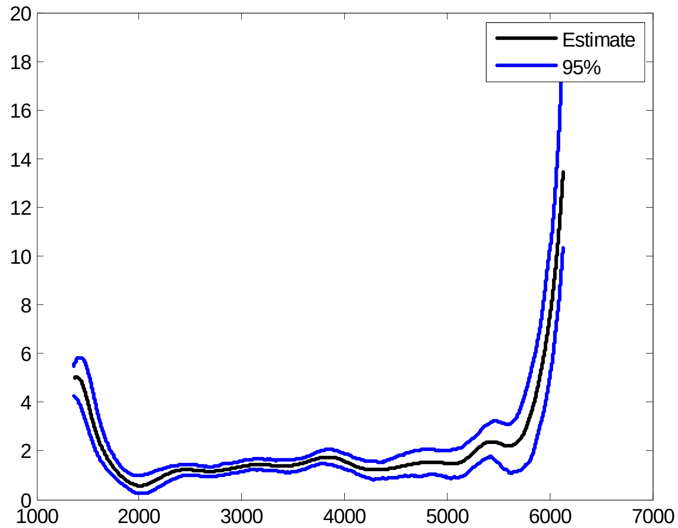

Figure 19: Average relative risk aversion with risk free returns, previous level of consumption at the third quartiles.

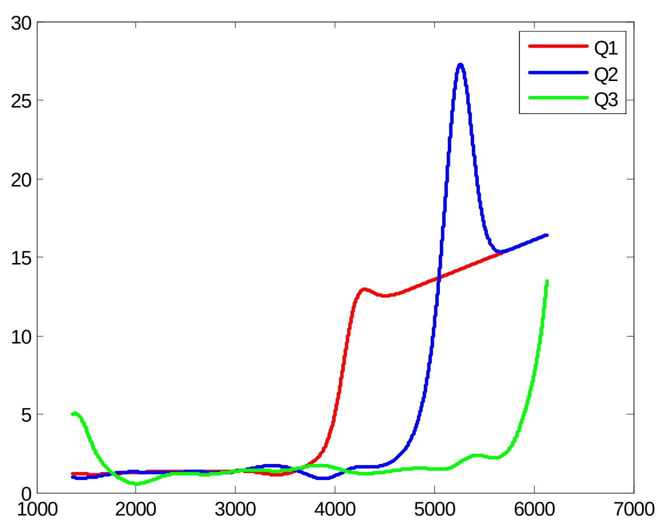

Figure 20: Average relative risk aversion with risk free returns, previous level of consumption across all quartiles.

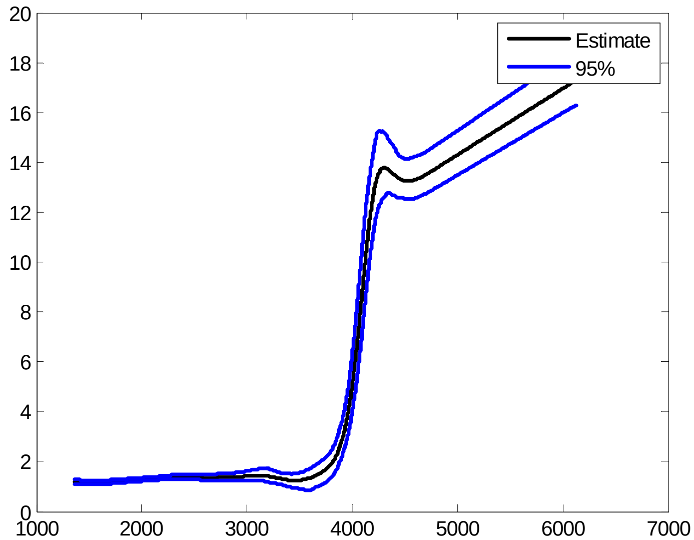

Figure 21: Average relative risk aversion with risky returns, previous level of consumption at the first quartiles.

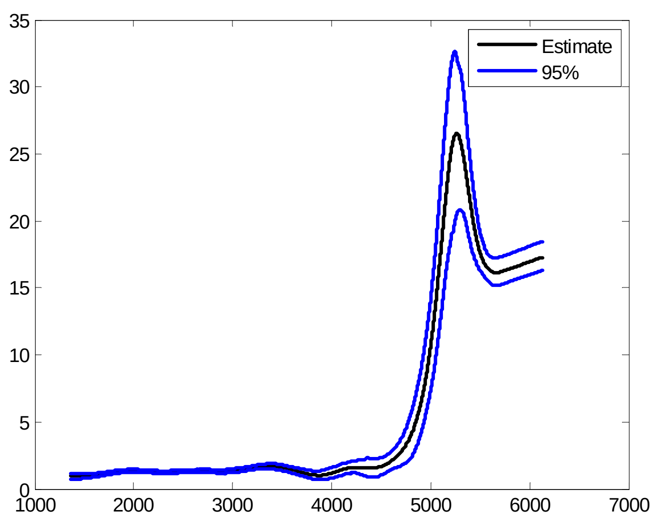

Figure 22: Average relative risk aversion with risky returns, previous level of consumption at the second quartiles.

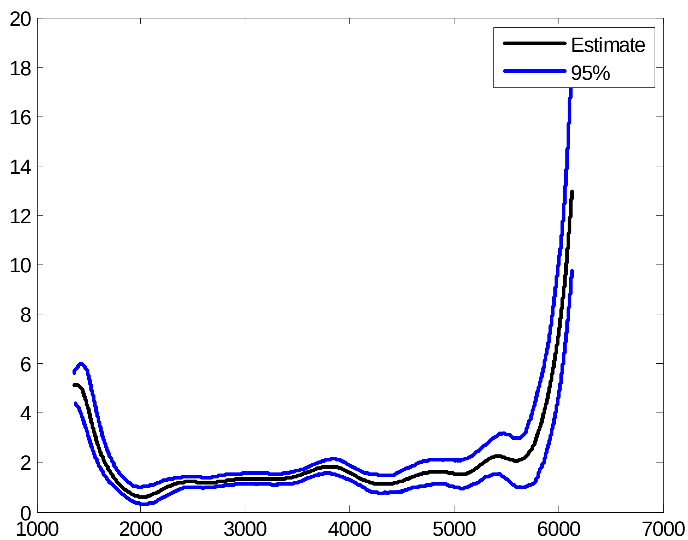

Figure 23: Average relative risk aversion with risky returns, previous level of consumption at the third quartiles.

The first observation from these plots is that the estimated nonparametric marginal utility is positive and decreasing (i.e. utility is concave). There are small differences in the estimated marginal utility with the risk-free and the risky asset, which supports our nonparametric model. RRA is increasing in most cases. For poor households (those in the first quartile of the consumption's distribution), RRA significantly increases with abrupt changes around the third quartile of the next period consumption's distribution. That is, poorer households are willing to take a significant amount of risk for small incremental values of consumption in that region. Richer households are uniformly less risk averse.

There are a number of tests that we could run with our methods. For example, we could formally test for the significance of habits, i.e.

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

Inference on this can be done using a moment based test based on semiparametric estimators of

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

for a fixed positive function φ, for example, $\varphi(C_{t+1}, C_t) = 1$, $\varphi(C_{t+1}, C_t) = C_t$, $\varphi(C_{t+1}, C_t) = C_{t+1}$, or powers of these. The asymptotic normality of the estimator and its bootstrap approximation can be used for inference, and this can be justified along the lines of what we did for θ_0.

Likewise, we could test for homogeneity of marginal utility. This is a commonly used assumption in the literature of asset pricing. This assumption is typically made for convenience, without

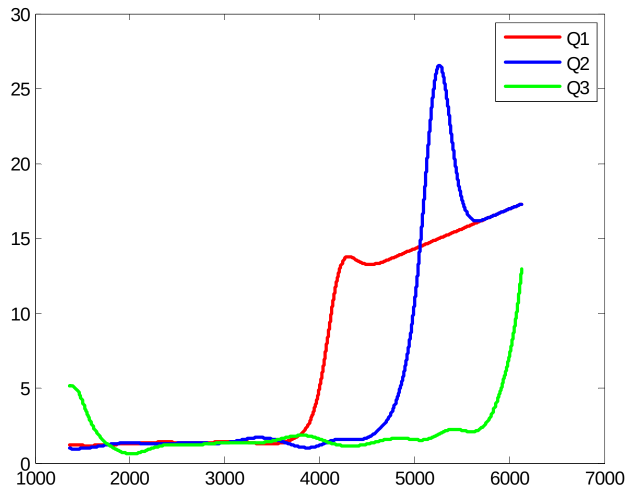

Figure 24: Average relative risk aversion with risky returns, previous level of consumption all three quartiles.

reference to economic theory, and it is often justified because it leads to marginal utilities that depend on consumption growth rather than consumption levels, thereby avoiding the need to address the non-stationary in consumption. Our nonparametric approach with household level data allows for an empirical evaluation of this assumption. A simple evaluation can be based on the following observation. Euler's theorem of homogenous functions implies that

$$a(C_{t+1}, C_t) := C_{t+1}\frac{g(C_{t+1}, C_t)}{\partial C_{t+1}} + C_t\frac{g(C_{t+1}, C_t)}{\partial C_t} = \lambda g(C_{t+1}, C_t),$$

for a real number $\lambda \in \mathbb{R}$. This in turn is equivalent to the random variable

$$r(C_{t+1}, C_t) = \frac{a(C_{t+1}, C_t)}{g(C_{t+1}, C_t)}$$

being a constant. Formal inference, accounting for estimating uncertainty, can be based on the following quantity

$$\delta_{Hom} = E\left[r(C_{t+1}, C_t)\varphi(C_{t+1}, C_t)\right],$$

for a fixed function φ satisfying $E\left[\varphi(C_{t+1}, C_t)\right] = 0$. Under homogeneity $\delta_{Hom} = 0$. An estimator for δ_{Hom} can be constructed similarly to that for θ_0, and bootstrap inference can be used to formally test for $\delta_{Hom} = 0$. Again, a battery of moment functions φ could be used (e.g. centered versions of $\varphi(C_{t+1}, C_t) = C_t$, $\varphi(C_{t+1}, C_t) = C_{t+1}$ and powers of them thereof).

8 Conclusions

In this article, we investigate nonparametric identification and estimation of marginal utilities and discount factors in consumption-based asset pricing Euler equations. The main insights of our nonparametric identification results are: (i) the decomposition of the pricing kernel into the primitives, as in equation (1), and (ii) the use of shape restrictions (positive marginal utilities). Based on our identification arguments, we have proposed a new nonparametric estimator for marginal utilities and 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 very easy to implement. We establish the limiting distribution theory for the discount factor and for the $ARRA$ parameter. Other semiparametric functionals can be studied based on our asymptotic results.

We have applied our nonparametric methods to household-level CEX data and obtained sensible estimates of the discount factor and the $ARRA$. The estimators are insensitive to the asset used (risk-free vs risky), which supports our nonparametric model. We find empirical support for the presence of habits. We have also uncovered substantial heterogeneity in risk attitudes, which might be hard to explain with parametric/semiparametric models.

9 Appendix

9.1 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 W_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 W_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}, $j = 1, ..., J$. 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 $\{W_t\}_{t=0}^{\infty}$. The Lagrangean is

$$E\left[\sum_{t=0}^{T} b^t U(C_t, V_t) - \left(C_t + (D_t - \delta D_{t-1}) P_t + \sum_{j=1}^{J} (A_{jt} - A_{jt-1} R_{jt})\right) \lambda_t\right] \tag{11}$$

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 $C_{t-1}, C_{t-2}, ...C_{t-L}$, 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 (11) as

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

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

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

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

$$g(C_t, V_t) = bE\left[g(C_{t+1}, V_{t+1}) R_{jt+1} \mid C_t, V_t\right] \qquad j = 1, ..., J, \tag{12}$$

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.

Although we will focus our attention on the asset equations, one also obtains an Euler equation associated with durables,

$$g_d(C_t, V_t) = P_t g(C_t, V_t) + \delta b E\left[g(C_{t+1}, V_{t+1}) P_{t+1} \mid C_t, V_t, P_t\right]. \tag{13}$$

Given estimates of the function g, equation (13) would then provide an equation for estimating the function g_d. When habits are external, it would also be possible to estimate g and g_d simultaneously, imposing the additional constraint from Young's theorem that

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

9.2 Lemmas

The following lemma follows from Einmahl and Mason (2005). Let Γ be a class of measurable real-valued functions of $W = (R', C', V', C, V)$. We denote by $\psi := (\varphi, c, v)$ a generic element of the set $\Psi := \Gamma \times S$. Let $f(c, v)$ denote the density of (C, V) evaluated at (c, v). Define the regression function $m(\psi) := E[\varphi(W) | C = c, V = v]$. Then, an estimator for $m(\psi)$ is given by

$$\widehat{m}_h(\psi) = \frac{1}{nh^\ell \widehat{f}(c, v)} \sum_{i=1}^{n} \varphi(W_i) K\left(\frac{c - C_i}{h}\right) \prod_{j=1}^{\ell_1} K\left(\frac{v - V_{ji}}{h}\right).$$

Define the class of functions

$$\mathcal{M} = \{(c, v) \to E[\varphi(W_i) | C = c, V = v] : \varphi \in \Gamma\}.$$

Define $\|g\|_\infty = \sup_{(c,v) \in T} |g(c, v)|$. Henceforth, we abstract from measurability issues that may arise in quantities such as $\sup_{g \in \mathcal{G} : \|g\| \leq 1} \left\| \widehat{A}g - Ag \right\|$ (see Pollard (1990, Chapter 9) for situations where more care is required regarding the issue of measurability). Consider the following assumption:

Assumption AA3:

1. $\{W_i\}_{i=1}^n$ is iid.

2. For each $\varepsilon > 0$, $\log N_{[\cdot]}(\varepsilon, \Gamma, \|\cdot\|) \leq C\varepsilon^{-v}$ for some $v < 2$. The class Γ has an envelope F such that $\sup_{(c,v)\in S} \mathbb{E}[|F(C', V')|^{\delta} |C = c, V = v] < \infty$ for some $\delta > 2$. Functions in $\mathcal{M}$ are $r-$times continuous differentiable with uniformly equicontinuous $r-th$ derivative on S.

3. The probability density function $f(c, v)$ is bounded, bounded away from zero and $r-$times continuous differentiable with continuous $r-th$ derivative on S.

4. The kernel function K is a $r-$th order that is Lipschitz continuous symmetric with support $[-1, 1]$ for some $r \geq 2$.

5. The possibly stochastic bandwidth h_n satisfies $P(l_n \leq h_n \leq u_n) \to 1$ as $n \to \infty$, for deterministic sequences of positive numbers l_n and u_n such that: $u_n \to 0$ and $l_n^{\ell} n/\log n \to \infty$.

Lemma A1. *Let Assumption AA3 hold. Then, we have,*

$$\sup_{a_n\leq h\leq b_n} \sup_{\psi\in\Psi} |\widehat{m}_h(\psi) - m(\psi)| = O_P\left(\sqrt{\frac{\ln n}{nl_n^{\ell}}} + u_n^r\right).$$

Proof. It follows from a simple variation of Theorem 4 in Einmahl and Mason (2005). ■

Lemma A2. *Under Assumption A3,*

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

and

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

as $n \to \infty$.

Proof. Follows from the definition of $\widehat{A}$ and Lemma A1. ■

Lemma A3. *Under Assumptions A3 and A4, for any* $\varphi \in \mathcal{L}^2(r)$, *it holds that*

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

Proof. Define

$$\widehat{T}g(c, v) = \frac{1}{n}\sum_{i=1}^{n} g_i' R_i' K_{hi}(c, v),$$

and note that $\widehat{A}g(c, v) = \widehat{T}g(c, v)/\widehat{f}(c, v)$. Using standard arguments, we write

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

where

$$a_n(c,v) = f^{-1}(c,v)\left(\widehat{T}g(c,v) - Tg(c,v) - Ag(c,v)\left(\widehat{f}(c,v) - f(c,v)\right)\right),$$

$Tg(c,v) := f(c,v)\,Ag(c,v)\,, \widehat{T}g(c,v) := \widehat{f}(c,v)\,\widehat{A}g(c,v)$ and

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

Lemma A1 and our conditions on the bandwidth imply $\|r_n\|_\infty = o_P(n^{-1/2})$. It then follows that $\left\langle\left(\widehat{A} - A\right)g, \varphi\right\rangle$ has the following expansion

$$\int \varphi(c,v)[\widehat{T}g(c,v) - Tg(c,v)]dcdv \tag{14}$$

$$- \int \varphi(c,v)Ag(c,v)\,[\widehat{f}(c,v) - f(c,v)]dcdv \tag{15}$$

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

We now look at terms (14)-(15). Firstly, it follows from standard arguments and A4.4 that the difference between $Tg(c,v)$ and $E[\widehat{T}g(c,v)]$ is $o_P(n^{-1/2})$. Hence,

$$\begin{aligned}
&\int \varphi(c,v)[\widehat{T}g(c,v) - Tg(c,v)]dcdv = \int \varphi(c,v)[\widehat{T}g(c,v) - E(\widehat{T}g(c,v))]dcdv + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n} g_i' R_i' \int \varphi(c,v)K_{hi}(c,v)\,dcdv - \int \varphi(c,v)E(g_i'R_i'K_{hi}(c,v))dcdv + o_P(n^{-1/2}),\\
&= \frac{1}{n}\sum_{i=1}^{n} \varphi(C_i,V_i)g_i'R_i' - E[\varphi(C_i,V_i)\,Ag(C_i,V_i)] + o_P(n^{-1/2}),
\end{aligned}$$

where the last equality follows from the standard change of variables argument and our Assumption A4. Likewise, the term (15) becomes $n^{-1/2}\sum_{i=1}^{n}\varphi(C_i,V_i)Ag(C_i,V_i) - E[\varphi(C_i,V_i)\,Ag(C_i,V_i)] + o_P(n^{-1/2})$. In conclusion, we have

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

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 by Assumption A3.2. ■

Lemma A4. *Let Assumptions A3 and A4 hold. If* $\varphi \in \mathcal{N}^{\perp}(L)$, *so* $\varphi = L^*\varphi^*$ *for some* φ^*, *and* $\varphi_s^* \in \mathcal{L}^2(r)$, *then*

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

Proof. Note that by (16) below and the adjoint property

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

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

$$
b\sqrt{n}\left\langle \left(\widehat{A}-A\right)g, s\left\langle g,\varphi^{*}\right\rangle - \varphi^{*}\right\rangle = -b\sqrt{n}\left\langle \left(\widehat{A}-A\right)g,\varphi_{s}^{*}\right\rangle + o_{P}(1).
$$

Then, the result follows from the last display and Lemma A3. ■

9.3 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 $T\in\mathcal{B}(\mathcal{X},\mathcal{X})$ 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).

Proof of Theorem 1: By Assumption A1 the set of countable eigenvalues of A has zero as a limit point, and thus, the set of eigenvalues which inverse is 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. 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 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 $\langle g,s\rangle\neq 0$, 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, $j = 1, ..., N_\varepsilon \equiv N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|)$, covering $\mathcal{G}$ with respect to $\|\cdot\|$. Assume without loss of generality that the kernel $k \geq 0$. Then, $[Al_j, Au_j]$, $j = 1, ..., N_\varepsilon$, 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) we obtain

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

and

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

where C is a real positive number that depends only on b, $\widetilde{g} = sgn\left(\langle\widehat{g}, g\rangle\right) g / \|g\|_n$ (sgn is the sign function, i.e., $sgn(x) = 1(x > 0) - 1(x < 0)$). By Lemma A2, $\left\|\widehat{A} - A\right\| = o_P(1)$. Also, from Lemma A2, $\|\widehat{g} - \widetilde{g}\|_\infty = o_P(1)$. Then, $\|sgn(\widehat{g}) - sgn(\widetilde{g})\|_\infty = o_P(1)$, and hence $sgn(\widehat{g}) = sgn\left(\langle\widehat{g}, g\rangle\right)$ with one for large n. The normalization $\widehat{g}\left(c_0, v_0\right) > 0$ then implies $\widetilde{g} = g / \|g\|_n$, and by the Law of Large Numbers and the normalization $\|g\| = 1$, it holds $\|\widetilde{g} - g\| = o_P(1)$. ■

Proof of Theorem 4: By definition

$$\widehat{b}\widehat{A}\widehat{g} - bAg = \widehat{g} - g.$$

Write the left hand side of the last display as

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

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

$$\langle bA(\widehat{g} - g), s\rangle = \langle\widehat{g} - g, s\rangle,$$

we obtain

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

Assumption A4.4, Lemma A1, 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, $\langle\widehat{g}, s\rangle \to_p \langle g, s\rangle \equiv 1$, and by Slutzky Theorem

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

Hence, the result follows from Lemma A3. ■

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

$$\begin{aligned} 0 &= \widehat{L}\widehat{g} - Lg \\ &= L(\widehat{g} - g) + (\widehat{L} - L)g + (\widehat{L} - L)(\widehat{g} - g). \end{aligned} \tag{16}$$

First, from previous results it is straightforward to show that

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

and

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

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

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

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

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = \sqrt{n}\left(\mathbb{P}_n\widehat{\psi} - \mathbb{P}\widehat{\psi}\right) + \sqrt{n}\left(\mathbb{P}\widehat{\psi} - \mathbb{P}\psi\right).$$

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

$$\sqrt{n}\left(\mathbb{P}_n\widehat{\psi} - \mathbb{P}\widehat{\psi}\right) = \sqrt{n}\left(\mathbb{P}_n\psi - \mathbb{P}\psi\right) + o_P(1).$$

Since $\widehat{g} - g$ is bounded with probability tending to one, we can apply integration by parts and use Assumption A5 to write

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

where the last equality follows from the Mean Value Theorem and the uniform consistency of $\widehat{g}$. Note that $\varphi \in \mathcal{N}^\perp(L)$, since

$$\begin{aligned} \langle g, \varphi\rangle &= E[d(C, V)] \\ &= 0. \end{aligned}$$

Then, by Lemma A4

$$\sqrt{n}\left(\mathbb{P}\widehat{\psi}-\mathbb{P}\psi\right)=\frac{-b}{\sqrt{n}}\sum_{i=1}^{n}\varphi_s^*(C_i,V_i)\varepsilon_i+o_P(1),$$

and therefore

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi(C_i,V_i)-\mathbb{P}\psi)-b\varphi_s^*(C_i,V_i)\varepsilon_i+o_P(1).$$

The result then follows from the a standard central limit theorem for strong mixing processes 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] Anatolyev, S. (1999), "Nonparametric Estimation of Nonlinear Rational Expectation Models," *Economics Letters*, 62, 1-6.

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

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

[6] 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.

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

[8] 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.

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

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

[11] 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.

[12] 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.

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

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

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

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

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

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

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

[20] 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.

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

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

[23] Escanciano, J. C. and S. Hoderlein (2010, 2012), "Nonparametric Identification of Euler Equations," unpublished manuscript.

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

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

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

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

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

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

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

[31] 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.

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

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

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

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

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

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

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

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

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

[41] 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.

[42] 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.