

# Nonparametric Euler Equation Identification and Estimation

Juan Carlos Escanciano

Stefan Hoderlein

Arthur Lewbel

Oliver Linton

Sorawoot Srisuma

revised: March, 2020

## 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 nonparametric instrumental variables estimators. We derive limiting distributions for our estimator and for relevant associated functionals. A Monte Carlo shows a satisfactory finite sample performance for our estimators.

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, anonymous referees, and seminar participants at University of Miami, UC San Diego, joint MIT-Harvard, Semiparametric Methods in Economics and Finance Workshop (London, 2010), Cowles workshop (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 and Linton (2010), and by Lewbel, Linton, and Srisuma (2012), and replaces “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 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

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 semi-parametric functionals of the marginal utility function such as the Average Relative Risk Aversion (**ARRA**) parameter defined below.

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

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

where  $M_{t+1} = b \frac{g(C_{t+1}; V_{t+1})}{g(C_t; V_t)}$  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

$$E[M_{t+1}(R_{t+1} - R_{f,t+1}) | C_t; V_t] = E\left[b \frac{g(C_{t+1}; V_{t+1})}{g(C_t; V_t)} (R_{t+1} - R_{f,t+1}) | C_t; V_t\right] = 0$$

equation (2), thereby estimating  $g$  instead of  $M$ .<sup>3</sup> 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, 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 establish asymptotic normality of a nonparametric estimator of the

$g(C_t; V_t) = C_t h(V_t)$  ; where  $\gamma$  is a constant that determines risk aversion and

prior knowledge. They first use completeness conditions to identify the parametric  $RRA$  and then use Perron-Frobenius to identify the role of habits. In contrast, we do not require a constant  $RRA$  or require completeness conditions for identification. Thus, the setting and identification approaches of this paper and those of Chen et al. (2014) are quite different.

An alternative to our kernel based estimation would be the use of sieves. Although we focus on kernel estimates, our asymptotic theory is developed in a way that can be easily adapted to other nonparametric estimation methods, including sieves (e.g. splines) and local polynomial methods. 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.<sup>4</sup> However, their results cannot be applied to our model, since in our case the associated operator is not self-adjoint. Christensen (2017) proposes a nonparametric sieve estimator for the discrete-time 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 several advantages as summarized in the previous section, mainly attributable to our method of exploiting the well-posedness of equation (1). In particular, with our methods we obtain novel asymptotic distribution theory for functionals of the nonparametric utility, such as the  $ARRA$  functional. This asymptotic theory is of independent interest and has wide applicability in other situations where type-II equations arise.

### 3 Identification

Since our goal is the study of Euler equations, we shall take as primitives the pair  $(g; b) \in G \times (0; 1)$ , where  $G$  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  $h$  if there exists some constant  $k_0 \in \mathbb{R}$  such that  $g = k_0 h$  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).

---

4

Definition.  $S; S \in \mathbb{R}^d$

$$S = S \setminus S \quad (C_t; V_t) \quad (C_{t+1}; V_{t+1})$$

$$L^2 \quad L_2(S; \cdot)$$

$$\|g\|_2 = \|f\|_2$$

Let  $M \subset L^2$  be a linear subspace; and define the linear operator  $A : (M; k, k) \rightarrow (M; k, k)$  by

$$Ag(c; v) = E[g(C_{t+1}; V_{t+1})R_{t+1} | C_t = c; V_t = v]; \quad (3)$$

We assume that  $Ag$  is well-defined and  $Ag \in M$



*Ag*

Theorem 1 shows that without further assumptions the Euler equation is partially identified, with  $\mathbf{b}$  identified up to a finite set corresponding to eigenvalues larger than one, and  $\mathbf{g}$  is identified up to a corresponding set of eigenfunctions. The discount factor  $\beta$  is also  $\beta > 1$ , meaning that for any  $\beta \in \mathbf{B}_0$  there is an open neighborhood of  $\beta$  that does not contain any other element in  $\mathbf{B}_0$ . Essentially, compactness of  $\mathbf{A}$  ensures that  $\mathbf{B}_0$  is at most countable, and the economic restriction that discount factors lie in  $(0;1)$  ensures that  $\mathbf{B}_0$  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  $(\mathbf{g}; \mathbf{b})$  must lie in the intersection of these identified sets. One might further shrink the identified set by imposing the restriction that  $\mathbf{b}g(\mathbf{C}_{t+1}; \mathbf{V}_{t+1})R_{t+1} - g(\mathbf{C}_t; \mathbf{V}_t)$  is uncorrelated with all variables in the information set at time  $t$ , not just measurable functions of  $(\mathbf{C}_t; \mathbf{V}_t)$ .

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 ent

We could consider other sufficient conditions that replace conditions on  $\mathbf{A}$  by conditions on a power of  $\mathbf{A}$ ; i.e. we could require that Assumptions C and I hold for  $\mathbf{A}^n$ ; 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. 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  $t$ .

For illustration, we consider the following examples of  $\mathbf{A}$  and  $\mathbf{M}$ ; which lead to simple conditions for identification by Theorem 2. Assume for simplicity that  $\mathbf{V}_{t+1}$  and  $\mathbf{V}_t$  are empty, and denote by  $\mathbf{f}(c; c)$ ;  $\mathbf{f}(c)$  and  $\mathbf{f}(c)$  the joint and marginal densities of  $(\mathbf{C}_{t+1}; \mathbf{C}_t)$ ; respectively. Assume  $\mathbf{f}$  has Lebesgue density  $\mathbf{f}$  on a common support  $\mathbf{S} = \mathbf{S} = \mathbf{S}$  (e.g.  $\mathbf{S} = [0; 1]$ ): Then, taking  $\mathbf{M}$  equals to  $L^2$ ; the operator equation  $\mathbf{bAg} = \mathbf{g}$  can be written as

$$\mathbf{b} \int k(c; c) \mathbf{g}(c) \mathbf{f}(c) dc = \mathbf{g}(c);$$

where  $k(c; c) = r(c; c) \mathbf{f}(c; c) = [\mathbf{f}(c) \mathbf{f}(c)]$  and  $r(c; c) = E[\mathbf{R}_{t+1} | \mathbf{C}_{t+1} = c; \mathbf{C}_t = c]$  a.s. Then, it is well known that Assumption C holds if

$$\int k^2(c; c) \mathbf{f}(c) \mathbf{f}(c) dc < 1 ;$$

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. For estimation we assume that we have a sample of household-level data  $f(\mathbf{R}_{t_j+1}; \mathbf{C}_{t_j+1}; \mathbf{V}_{t_j+1}; \mathbf{C}_{t_j}; \mathbf{V}_{t_j}) g_{i=1}^n$  for  $n$  households, with possibly overlapping time periods  $t_1, t_2, \dots, t_n$ . To simplify notation denote  $\mathbf{W}_i = (\mathbf{R}_i; \mathbf{C}_i; \mathbf{V}_i; \mathbf{C}_i; \mathbf{V}_i) = (\mathbf{R}_{t_j+1}; \mathbf{C}_{t_j+1}; \mathbf{V}_{t_j+1}; \mathbf{C}_{t_j}; \mathbf{V}_{t_j})$ ; where  $\mathbf{V}_i = (\mathbf{V}_{1i}; \dots; \mathbf{V}_{1i})$  and  $\mathbf{V}_i = (\mathbf{V}_{1i}; \dots; \mathbf{V}_{1i})$

$f_i(c; v); i = 1; \dots; n$ ): Therefore, similar to our discussion of identification in Section 3, the number of eigenvalues and eigenfunctions of  $A$  is finite and bounded by  $n$ , and they can be computed by solving a linear system. Indeed, any eigenfunction  $g(c; v)$  of  $A$  necessarily has the form  $n^{-1} \sum_{i=1}^n \alpha_i f_i(c; v)$ ; for some coefficients  $\alpha_i; i = 1; \dots; n$ ; satisfying for its corresponding eigenvalue  $\lambda$  the equation

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \alpha_j f_j(C_i; V_i) R_{ij}(c; v) = \lambda \frac{1}{n} \sum_{i=1}^n \alpha_i f_i(c; v):$$

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

$$\frac{1}{n} \sum_{j=1}^n \alpha_j f_j(C_i; V_i) R_{ij} = \lambda \alpha_i$$

which in matrix notation can be simply written as

$$A_n \alpha = \lambda \alpha;$$

where  $A_n$  is an  $n \times n$  matrix with  $ij$ -th element  $a_{ij} = \frac{1}{n} f_j(C_i; V_i) R_{ij}$ ; and  $\alpha = (\alpha_1; \dots; \alpha_n)$  (henceforth,  $v$  denotes the transpose of  $v$ ): Thus, let  $\lambda$  denote the largest eigenvalue in modulus of  $A_n$  and  $\alpha = (\alpha_1; \dots; \alpha_n)$  its corresponding eigenvector. Our estimators for  $b_0$  and  $g_0$  are, respectively,

$$\hat{b} = 1 = \frac{1}{n} \sum_{i=1}^n \alpha_i \quad \text{and} \quad g(c; v) = \frac{1}{n} \sum_{i=1}^n \alpha_i f_i(c; v): \tag{7}$$

Marginal utilities are identified up to scale and we consider the normalization  $\|g\| = 1$ ; which is implemented by setting  $\alpha = 1$ ; where  $\alpha$  is the  $n \times n$  matrix with entries

$$\alpha_{ij} = \frac{1}{n^2} \int f_i(c; v) f_j(c; v) f(c; v) dc dv:$$

As a practical recommendation, we could also normalize  $g(C_i; V_i)$  to have unit standard deviation. Also, we impose the sign normalization  $\langle g; \hat{b} \rangle > 0$ : The estimator  $(g; \hat{b})$  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  $\lambda$ ; see e.g. Chanchana (2007)(r)(c)9(e)

The easiest way to consider simultaneously different assets in our estimation strategy is to obtain individual estimates of the marginal utility for each asset by the method above and then combine the resulting estimators to reduce the variance; see e.g. Chen, Jacho-Chavez and Linton (2016). Next section addresses this point.

## 4.1 Estimation with multiple assets

Suppose that we have  $J$  assets, and let  $\hat{b}_j$

first order behavior of  $\hat{b}_j$ ; and thus its asymptotic distribution will follow from the results obtained in the next section.

Similar asymptotic results to those develop above can be used to test for overidentifying restrictions. Take for simplicity the case  $J = 2$ ; and assume our conditions for identification hold. We can then test the restriction  $b_1 = b_2$  (where  $b_j^2$

Assumption E:

1.  $\|g\| = 1$  and  $\langle g, \mathbf{1} \rangle > 0$ :
2.  $\inf_{g \in \mathcal{G}_0} \|\mathbf{A}g\| > 0$

Condition E.1 is just a convenient normalization for our setting: Assumption E.2 is a mild consistency condition. Note that by our identification results  $\mathcal{G}_0$  consists of the linear span of  $g_0$ . More generally, under Assumption C,  $\mathcal{G}_0$  is finite dimensional, which makes E.2 easy to check; see the Appendix for primitive conditions for kernel estimators. Our next result shows the strong  $L^2$ -consistency of our estimators:

Theorem 3.  $\|\hat{b} - b_0\| \leq C \|g - g_0\|$

We remark that Theorem 3 also holds in the partially identified case where Assumption I is dropped and the  $L^2$ -distance between  $g$  and  $g_0$  is replaced by the gaps between the eigenspaces of  $\mathbf{A}$  and  $\hat{\mathbf{A}}$  associated to the eigenvalues  $\hat{b}^{-1} = \lambda(\hat{\mathbf{A}})$  and  $b_0^{-1} = \lambda(\mathbf{A})$



3.

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \mathbf{s}_i \mathbf{s}_i' \xrightarrow{p} \mathbf{N}(0; \mathbf{s}); \\ \lim_n \text{var} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{s}_i \mathbf{s}_i' \right) & < 1; \end{aligned}$$

Theorem 4.

$$\frac{1}{n} \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i' \xrightarrow{p} \mathbf{N}(0; \mathbf{b}_0^A \mathbf{s});$$

The proof of Theorem 4 can be found in the Appendix. We can estimate the asymptotic variance of  $\mathbf{b}$  by standard long run variance estimators based on  $\frac{1}{n} \sum_{i=1}^n \mathbf{s}_i \mathbf{s}_i'$ ; see e.g. Newey and West (1987), where  $\mathbf{s}_i = \mathbf{g}(\mathbf{C}_i; \mathbf{V}_i) - \mathbf{g}(\mathbf{C}_i; \mathbf{V}_i)$ ; and  $\mathbf{s}$  is computed as our estimator  $\mathbf{g}$ ; with the normalization  $\mathbf{h}(\mathbf{g}; \mathbf{s})_n = 1$ : An alternative to plug-in asymptotic methods is to use block bootstrap, see e.g. Radulović (1996).

For the estimator based on  $\mathbf{J}$  assets proposed in Section 4.1, note that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{w}}_b)_i \hat{\mathbf{b}}^{(j)} \mathbf{b}_0 & = (\hat{\mathbf{w}}_b) \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{b}}^{(j)} \mathbf{b}_0 \\ & + \frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{w}}_b - \mathbf{w}_b)_i \mathbf{b}_0; \end{aligned}$$

Since the second term is exactly zero, by construction of the weights, we expect, by consistency of the long run variance estimator and the proof of Theorem 4 above,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{w}}_b)_i \hat{\mathbf{b}}^{(j)} \mathbf{b}_0 & = \frac{1}{n} \sum_{i=1}^n (\mathbf{w}_b)_i \hat{\mathbf{b}}^{(j)} \mathbf{b}_0 + o_p(1) \\ & \xrightarrow{p} \mathbf{N}(0; \mathbf{b}_0^A(\mathbf{w}_b) \mathbf{J} \mathbf{w}_b); \end{aligned}$$

where  $\mathbf{J}$  is defined in (8).

Our next result establishes an asymptotic expansion for  $\mathbf{g} - \mathbf{g}_0$ : This expansion can be used to obtain rates for  $\mathbf{g} - \mathbf{g}_0$  and to establish asymptotic normality of (semiparametric) functionals of  $\mathbf{g}$ . Define the process  $\frac{1}{n} \sum_{i=1}^n \mathbf{s}_i(\mathbf{c}; \mathbf{v})$ ; where recall that  $\mathbf{s}_i(\mathbf{c}; \mathbf{v}) = \mathbf{K}_{hi}(\mathbf{c}; \mathbf{v}) - \mathbf{f}(\mathbf{c}; \mathbf{v})$ : Note that a standard result in kernel estimation is that for all  $(\mathbf{c}; \mathbf{v})$  in the interior of  $\mathbf{S}$ ; under suitable conditions,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{s}_i(\mathbf{c}; \mathbf{v}) \xrightarrow{p} \mathbf{N}(0; \mathbf{h}(\mathbf{c}; \mathbf{v}));$$

with  $\mathbf{h}(\mathbf{c}; \mathbf{v}) = \mathbf{f}'(\mathbf{c}; \mathbf{v}) \mathbf{h}(\mathbf{c}; \mathbf{v}) \mathbf{f}'(\mathbf{c}; \mathbf{v})$  and  $\mathbf{h}(\mathbf{c}; \mathbf{v}) = \mathbf{E}[\mathbf{s}_i^2 | \mathbf{C}_i = \mathbf{c}; \mathbf{V}_i = \mathbf{v}]$ .



Under the assumptions for Theorem 6 below,  $g$  is differentiable and bounded away from zero with probability tending to one, so  $\hat{g}_n(g)$  is well-defined for large  $n$ . Define the class of functions

$$D = \{ (c; v) \mid c \frac{\partial \log(g(c; v))}{\partial c} : g \in G \}; \quad (12)$$

and the functions

$$d(c; v) = \frac{\partial (c \frac{f(c; v)}{\partial c})}{\partial c} \frac{1}{f(c; v)} \quad \text{and} \quad (c; v) \frac{d(c; v)}{g_0(c; v)}; \quad (13)$$

Also, we need to introduce some notation to be used in the asymptotic normality of  $\hat{g}_n(g)$ : Assuming  $\int L^2$ ; define

$$s = \frac{hg_0; \int hg_0; \int s}{\int hg_0; \int hg_0; \int s}; \quad (14)$$

The function  $s$  has a geometrical interpretation as the value of  $g_0$  projected parallel to  $s$  on a subspace of functions orthogonal to  $g_0$ . Let  $L$  denote the adjoint operator of  $L$ ; and let  $s$  denote the minimum norm solution of  $s = Lr$  in  $r$ ; i.e.  $s = \arg \min \|r\| : s = Lr$ ; which is well defined because  $s \in N(L) = R(L)$ ; see Luenberger (1997, Theorem 3, p. 157) for the latter equality. Here  $N(L)$  denotes the orthogonal complement of the null space of  $L$ , see Luenberger (1997, p. 52) for a definition.

We also introduce a class of smooth function  $C(T)$  for a generic closed and convex set  $T$ . For any vector  $a$  of  $\ell$  integers define the differential operator  $\partial_x^a = \partial_{x_1}^{a_1} \cdots \partial_{x_\ell}^{a_\ell}$ ; where  $\sum_{i=1}^{\ell} a_i = |a|$ . For any smooth function  $h : T \rightarrow \mathbb{R}$  and some  $\epsilon > 0$ , let  $\ell_\epsilon$  be the largest integer smaller or equal than  $\ell$ , and

$$\|h\|_{\ell, \epsilon} = \max_{|a| \leq \ell_\epsilon} \sup_{x \in T} |\partial_x^a h(x)| + \max_{|a| = \ell_\epsilon} \sup_{x = x^0} \frac{|\partial_x^a h(x) - \partial_x^a h(x^0)|}{|x - x^0|^\epsilon}.$$

Further, let  $C_M(T)$  be the set of all continuous functions  $h : T \rightarrow \mathbb{R}$  with  $\|h\|_{\ell, \epsilon} \leq M$  (for an integer  $\ell$ ; the  $\ell$ -th derivative is assumed to be continuous). Since the constant  $M$  is irrelevant for our results, we drop the dependence on  $M$  and denote  $C(T)$  (tsc]TJ/F2

1.  $D$   $P$  5
2.  $(C; V)$   $S$   $S =$   
 $[l_c; u_c] S_V; l_c; u_c l_c < u_c: \lim_c l_c cf(c; v) = 0 = \lim_c u_c cf(c; v)$   
 $v \in S_V P(\inf_{g_0; g} g > \epsilon) = 1 \quad \epsilon > 0$
3.  $d$   $d \in L^2; f_i g$   
 $\frac{1}{n} \sum_{i=1}^n \epsilon_i \sim N(0; \sigma^2);$   
 $\lim_n \text{var} \frac{1}{n} \sum_{i=1}^n \epsilon_i < 1 \quad s \in C^r(S)$

Assumption CE.1 is standard in the semiparametric literature, see, e.g. Chen, Linton and Van Keilegom (2003). Assumption CE.2 is similar to other assumptions required in estimation of average derivatives, see Powell, Stock and Stoker (1989). This assumption guarantees that  $\hat{g}_n(g)$  is well defined. [https://doi.org/10.815910\(t\)4450](https://doi.org/10.815910(t)4450)

where  $Q_q$  denotes the interval between the  $q - 1$  and  $q$  quartile of  $C_{t+1}$ , and  $S_j$  denotes the interval between the  $j - 1$  and  $j$  quartile of  $C_t$  for  $q, j = 1; 2; 3; 4$ . We refer to each of these local averages of the **RRA** between different quartiles as a **QRRRA** (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 **QRRRA** parameters  $\rho(q; j)$  can be shown to be asymptotically normal under the same conditions above used for the **ARRA**: That is, with the simplified notation  $\rho(q) = \rho(q; q)$  for the parameter and  $\hat{\rho}_n(q) = \hat{\rho}_n(q; q)$  for the plug-in estimator, it can be shown

$$\sqrt{n}(\hat{\rho}_n(q) - \rho(q)) \xrightarrow{d} N(0; \sigma^2(q));$$

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

$$H_{0qj}: \rho(q) = \rho(j) \quad \text{vs} \quad H_{1qj}: \rho(q) \neq \rho(j);$$

The t-statistics are constructed as

$$t_{qj} = \frac{\sqrt{n}(\hat{\rho}_n(q) - \hat{\rho}_n(j))}{\sqrt{\frac{\sigma^2(q)}{n} + \frac{\sigma^2(j)}{n}}};$$

for suitable consistent estimates  $\hat{\sigma}_n^2(q)$  of the asymptotic variances  $\sigma^2(q)$ ; for  $q = 1; 2; 3$  and  $4$ : We then reject  $H_{0qj}$  when  $t_{qj}$  is large in absolute value, using that  $t_{qj}$  converges to a standard normal under  $H_{0qj}$ :

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

$$g(g) = E \frac{\partial g(C_{t+1}; C_t)}{\partial C_t} (C_{t+1}; C_t);$$

for various positive functions  $g(\cdot)$ . When there is no habit effect  $g_0 = 0$  for any choice of  $g$ . As with  $g_0$ , for each choice of function  $g$  we estimate  $g_0$  by plugging in  $g$  for  $g_0$

**ARRA.** The model is then given by the Euler equation

$$b_0 E C_{t+1}^{-\theta} R_{t+1} C_t = C_t^{-\theta}$$

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

$$(\log C_t, \log C_{t+1}) \sim N(0; \begin{matrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{matrix});$$

and construct  $R_{t+1} = b_0^{-1} (1 + \epsilon_t) (C_{t+1}/C_t)^{\theta}$ , where  $\epsilon_t$  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 ap-

function  $g$  is then recovered using the function  $g(c; v) = g(c; v) = c$ . Throughout we set the bandwidth  $h$  to be  $1.06 n^{-1/3.5}$ , where  $n$  is the sample size. The standard deviation of  $C_t$ . This is essentially Silverman's rule applied to the rate  $n^{-1/3.5}$ . Our estimators for  $g_0$  are normalized to have a unit standard deviation.

For each finite-dimensional parameter and summary measure we consider, we report the standard deviation, 2.5th percentile, 5th percentile, 95th percentile, 95% coverage probability based on the bootstrap distribution, their bootstrap standard errors, and the root mean square error.<sup>6</sup> Table 1 reports estimates of the discount factor from three estimators, *CRR*, *NP*<sub>1</sub>, and *NP*<sub>2</sub>. Table 2 reports estimates of the o

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  $(g_0)$  that can be used to test for the presence of habits in  $g_0$ . In our experiments estimates of  $(g_0)$  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.



	$b_0$	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n = 500$	<b>CRRRA</b>	0.000	0.012	0.926	0.975	0.946	0.012	0.926	0.974	0.940	0.012
	<b>NP 1</b>	0.006	0.027	0.917	0.971	0.984	0.018	0.915	0.980	0.929	0.028
	<b>NP 2</b>	0.009	0.041	0.808	0.983	0.963	0.031	0.895	1.012	0.932	0.042
$n = 2000$	<b>CRRRA</b>	0.000	0.006	0.938	0.961	0.960	0.006	0.938	0.962	0.950	0.006
	<b>NP 1</b>	0.004	0.020	0.936	0.960	0.992	0.009	0.932	0.965	0.924	0.020
	<b>NP 2</b>	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  $b_0 = 0.95$ . **CRRRA**, **NP 1** and **NP 2** refer respectively to the parametric, one-dimensional

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



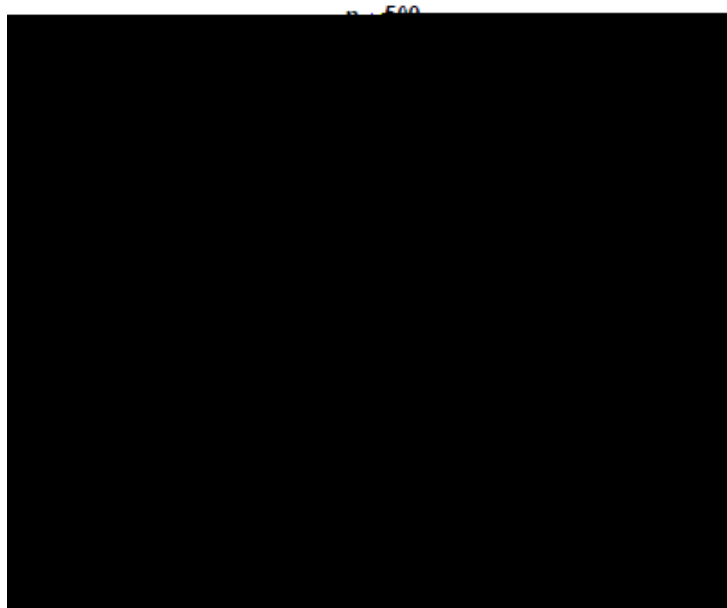


Figure 1: Estimates of the marginal utility function  $g_0$  using simulated data with  $n = 500$ . *Est*, *CI*, and *True* represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true.

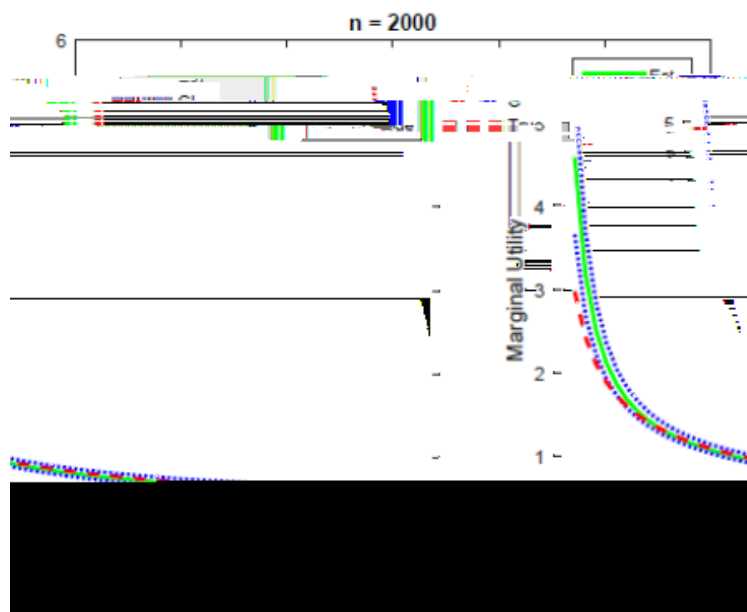


Figure 2: Estimates of the marginal utility function  $g_0$  using simulated data with  $n = 2000$ . *Est*, *CI*, and *True* represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true.



## 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  $Z_t$  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

$$\max_{C_t; D_t} E \sum_{t=0}^{\infty} b^t U(C_t; V_t) :$$

The consumer saves by owning durables and by owning quantities of risky assets  $A_{jt}$ ,  $j = 1; \dots; J$ . 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  $\delta$ . Then without frictions the consumer's budget constraint can be written as, for each period  $t$ ,

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

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  $f_{Z_t} g_{t=0}$ . The Lagrangean is

$$E \sum_{t=0}^{\infty} b^t U(C_t; V_t) - \sum_{t=0}^{\infty} \lambda_t [C_t + (D_t - \delta D_{t-1}) P_t + \sum_{j=1}^J A_{jt} - \sum_{j=1}^J A_{j,t-1} R_{jt}] \quad (17)$$

with Lagrange multiplier  $\lambda_t$  [TJ/F151rasse1 wanoj8Td[(C)]TJ/F237.J-718

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  $g$  by

$$g(I_t) = \sum_{s=0}^L b^s E \frac{\partial U(C_{t+s}; V_{t+s})}{\partial C_t} \mid I_t .$$

where  $L$  is such that  $V_t$  contains  $C_{t-1}; C_{t-2}; \dots; C_{t-L}$ , 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  $g(I_t) = g(C_t; V_t)$ , while for internal habits define

$$g(C_t; V_t) = E[g(I_t) \mid C_t; V_t].$$

With this notation, regardless of whether habits are internal or external, we may write the first order conditions associated with the Lagrangean (17) as

$$\begin{aligned} \lambda_t &= b^t g(I_t) \\ \lambda_t &= E[\lambda_{t+1} R_{j,t+1} \mid I_t] \quad j = 1; \dots; J \\ \lambda_t P_t &= b^t g_d(C_t; V_t) - E[\lambda_{t+1} P_{t+1} \mid I_t] \end{aligned}$$

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

$$\begin{aligned} b^t g(I_t) &= E[b^{t+1} g(I_{t+1}) R_{j,t+1} \mid I_t] \quad j = 1; \dots; J \\ b^t g(I_t) P_t &= b^t g_d(C_t; V_t) - E[b^{t+1} g(I_{t+1}) P_{t+1} \mid I_t] \end{aligned}$$





3.

$$\sup_{I_n} \sup_{h, u_n} |m_h(\cdot) - m(\cdot)| = O_p \left( \frac{1}{n^{1/2}} + u_n \right). \quad (20)$$

Proof. By the Triangle inequality

$$\begin{aligned} |m_h(\cdot) - m(\cdot)| &= \left| \frac{E[T_h(\cdot)]}{E[f(\mathbf{c}; \mathbf{v})]} + \frac{E[T_h(\cdot)]}{E[f(\mathbf{c}; \mathbf{v})]} - m(\cdot) \right| \\ &= \left| \frac{1}{f(\mathbf{c}; \mathbf{v})} T_h(\cdot) - \frac{E[T_h(\cdot)]}{E[f(\mathbf{c}; \mathbf{v})]} + \frac{E[T_h(\cdot)]}{f(\mathbf{c}; \mathbf{v})} - \frac{E[T_h(\cdot)]}{E[f(\mathbf{c}; \mathbf{v})]} \right| \\ &+ \left| \frac{1}{E[f(\mathbf{c}; \mathbf{v})]} E[T_h(\cdot)] - T(\cdot) \right| + \left| \frac{|T(\cdot)|}{E[f(\mathbf{c}; \mathbf{v})]} - \frac{|T(\cdot)|}{f(\mathbf{c}; \mathbf{v})} \right| \end{aligned}$$

where  $T(\cdot) = m(\cdot) f(\mathbf{c}; \mathbf{v})$ . We obtain uniform rates for  $T_h(\cdot) - E[T_h(\cdot)]$ ; the rates for  $f(\mathbf{c}; \mathbf{v}) - E[f(\mathbf{c}; \mathbf{v})]$  follow analogously and are simpler to obtain.

Define the class of functions

$$K_0 := \{C_i; V_i \mid C_j\}$$

$\mathbf{c}; \mathbf{v}$

$T$

and where  $\psi^{-1}$  is the inverse cadlag of the decreasing function  $u \mapsto \psi(u)$  ( $\psi(u)$  being the integer part of  $u$ , and  $\psi_t$  being the mixing coefficient) and  $Q_\psi$  is the inverse cadlag of the tail function  $u \mapsto P(|f| > u)$  (see Doukhan, Massart and Rio (1995)). Note that by Assumption A1 and Pollard (1984, p. 36)

$$P(|f| > z) \leq \frac{E[f^2]}{z^2} + \frac{Ch}{z^2}.$$

Definition 4.  $L^2(\mathbf{r})$   $\int_0^1 \int_0^1 \mathbf{r}(\mathbf{c}; \mathbf{v}) \mathbf{r}(\mathbf{c}; \mathbf{v})^T d\mathbf{c} d\mathbf{v} < \infty$

Lemma B3.  $\int_0^1 \int_0^1 \mathbf{r}(\mathbf{c}; \mathbf{v}) \mathbf{r}(\mathbf{c}; \mathbf{v})^T d\mathbf{c} d\mathbf{v} < \infty$

$$\frac{1}{n} \sum_{i=1}^n \mathbf{r}_i \mathbf{r}_i^T = \int_0^1 \int_0^1 \mathbf{r}(\mathbf{c}; \mathbf{v}) \mathbf{r}(\mathbf{c}; \mathbf{v})^T d\mathbf{c} d\mathbf{v} + o_p(1);$$

$$\frac{1}{n} \sum_{i=1}^n \mathbf{r}_i \mathbf{r}_i^T \xrightarrow{d} N(0, \int_0^1 \int_0^1 \mathbf{r}(\mathbf{c}; \mathbf{v}) \mathbf{r}(\mathbf{c}; \mathbf{v})^T d\mathbf{c} d\mathbf{v});$$

Proof. Define

$$Tg_0(\mathbf{c}; \mathbf{v}) = \frac{1}{n} \sum_{i=1}^n g_{0i} R_i K_{hi}(\mathbf{c}; \mathbf{v});$$

with  $g_{0i} = g_0(\mathbf{C}_i; \mathbf{V}_i)$  and note that  $Ag_0(\mathbf{c}; \mathbf{v}) = Tg_0(\mathbf{c}; \mathbf{v}) = \mathbf{f}(\mathbf{c}; \mathbf{v})$ . Using standard arguments, we write

$$Ag_0(\mathbf{c}; \mathbf{v}) = a_n(\mathbf{c}; \mathbf{v}) + r_n(\mathbf{c}; \mathbf{v}),$$

where

$$a_n(\mathbf{c}; \mathbf{v}) = \mathbf{f}^{-1}(\mathbf{c}; \mathbf{v}) Tg_0(\mathbf{c}; \mathbf{v}) - Tg_0(\mathbf{c}; \mathbf{v}) Ag_0(\mathbf{c}; \mathbf{v}) + \mathbf{f}(\mathbf{c}; \mathbf{v}) \mathbf{f}(\mathbf{c}; \mathbf{v})^{-1};$$

$Tg_0(\mathbf{c}; \mathbf{v}) = \mathbf{f}(\mathbf{c}; \mathbf{v}) Ag_0(\mathbf{c}; \mathbf{v})$ ;  $Tg_0(\mathbf{c}; \mathbf{v}) = \mathbf{f}(\mathbf{c}; \mathbf{v}) Ag_0(\mathbf{c}; \mathbf{v})$  and

$$r_n(\mathbf{c}; \mathbf{v}) = \frac{\mathbf{f}(\mathbf{c}; \mathbf{v}) - \mathbf{f}(\mathbf{c}; \mathbf{v})}{\mathbf{f}(\mathbf{c}; \mathbf{v})} a_n(\mathbf{c}; \mathbf{v});$$

Lemma B1 and our conditions on the bandwidth imply  $\|r_n\| = o_p(n^{-1/2})$ . It then follows that  $Ag_0(\mathbf{c}; \mathbf{v})$  has the following expansion

$$\int_0^1 \int_0^1 [Tg_0(\mathbf{c}; \mathbf{v}) - Tg_0(\mathbf{c}; \mathbf{v})] d\mathbf{c} d\mathbf{v} \quad (21)$$

$$\int_0^1 \int_0^1 Ag_0(\mathbf{c}; \mathbf{v}) [\mathbf{f}(\mathbf{c}; \mathbf{v}) - \mathbf{f}(\mathbf{c}; \mathbf{v})] d\mathbf{c} d\mathbf{v} \quad (22)$$

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

We now look at terms (21)-(22). Firstly, it follows from standard arguments and A2.5 that the difference between  $Tg_0(\mathbf{c}; \mathbf{v})$  and  $E[Tg_0(\mathbf{c}; \mathbf{v})]$  is  $O_p(u_n^r) = o_p(n^{-1/2})$  by the condition  $nu_n^r \rightarrow 0$ :

Hence,

$$\begin{aligned} & \int (c; v) [Tg_0(c; v) - E(Tg_0(c; v))] dc dv = \int (c; v) [Tg_0(c; v) - E(Tg_0(c; v))] dc dv + o_p(n^{-1/2}) \\ & = \frac{1}{n} \sum_{i=1}^n g_{0i} R_i \int (c; v) K_{hi}(c; v) dc dv - \int (c; v) E(g_0 R_i K_{hi}(c; v)) dc dv + o_p(n^{-1/2}), \end{aligned}$$

### 9.3 Main Proofs

The spectral radius  $\rho(\mathbf{A})$  of a linear continuous operator  $\mathbf{A}$  on a Banach space  $X$  is defined as  $\sup_{\lambda \in \sigma(\mathbf{A})} |\lambda|$ , where  $\sigma(\mathbf{A}) \subset \mathbb{C}$  denotes the spectrum of  $\mathbf{A}$ . Any compact operator  $\mathbf{A}$  has a discrete spectrum, so that  $\sigma(\mathbf{A})$  is simply the set of eigenvalues of  $\mathbf{A}$ . For more definitions and further details see Kress (1999, Chapter 3.2). The operator  $\mathbf{B}$  is called positive if  $\langle \mathbf{B}g, g \rangle \geq 0$  when  $g \in P$ .

**Proof of Theorem 1.** By Assumption C the set of countable eigenvalues of  $\mathbf{A}$  has zero as a limit point, and thus, the set of eigenvalues with  $|\lambda| \geq \epsilon > 0$  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  $\mathbf{A}^*$  denote the adjoint of  $\mathbf{A}$ ; which is also compact and positive by well known results in functional analysis. Assumption S implies that  $\rho(\mathbf{A}^*) > 0$ : Also notice that the eigenvalues of  $\mathbf{A}^*$  are complex conjugates of those of  $\mathbf{A}$  (in particular,  $\rho(\mathbf{A}^*) = \rho(\mathbf{A})$ ): Then, by the Kreĭn-Rutman's theorem (see Theorem 7.C in Zeidler (1986, vol. 1, p. 290)) there is exactly one solution to  $\mathbf{A}^*g = \lambda g$  with  $\lambda > 0$  and  $\|g\| = 1$  and a solution to  $\mathbf{A}s = \lambda s$  with  $\lambda > 0$ . Note  $\langle g, s \rangle = \langle \mathbf{A}^*g, s \rangle = \langle g, \mathbf{A}s \rangle = \lambda \langle g, s \rangle$ . Hence, since  $g$  and  $s$  are strictly positive,  $\langle g, s \rangle \neq 0$ ; and then  $\lambda = \rho(\mathbf{A})$ . ■

**Proof of Theorem 3.** By Theorems 1 and 2 in Osborn (1975), there is a constant  $M$  such that

$$\|b^{-1} - b_0^{-1}\| \leq M \|j\mathbf{A} - \mathbf{A}j\|_{G_0} \quad (23)$$

and

$$\|kg - gk\| \leq M \|j\mathbf{A} - \mathbf{A}j\|_{G_0}; \quad (24)$$

where  $g = \langle g, g_0 \rangle g_0$  is the projection of  $g$  on  $g_0$ . Thus, by  $0 < b_0 < b < 1$ ; a.s,

$$\begin{aligned} \|b - b_0\| &\leq M \|b - b_0\| \|j\mathbf{A} - \mathbf{A}j\|_{G_0} \\ &\leq M \|j\mathbf{A} - \mathbf{A}j\|_{G_0}; \end{aligned}$$

and by Assumption E.2  $\|b - b_0\| = o_P(1)$ .

To conclude that  $\|kg - gk\| = o_P(1)$  we need to show that  $\|kg - gk\| = o_P(1)$ . First, we show that  $\langle g, g_0 \rangle$  is non-negative for sufficiently large  $n$ : To see this, note

$$\begin{aligned} \langle g, 1 \rangle &= \langle g, 1 \rangle + o_P(1) \\ &= \langle g, g_0 \rangle \langle g_0, 1 \rangle + o_P(1) \\ &\geq 0; \end{aligned}$$

so  $\|hg; g_0\| \rightarrow 0$  for large enough  $n$ :

Next,

$$\begin{aligned} 1 &= \|g\| \quad (\text{by normalization}) \\ &= \|g\| + o_P(1) \quad (\text{by } \|g - g_0\| = o_P(1)) \\ &= \|hg; g_0\| + o_P(1); \quad (\text{by definition of } g) \end{aligned}$$

which then implies  $\|g - g_0\| = \|hg; g_0\| + o_P(1) = o_P(1)$ : Hence, by the triangle inequality,  $\|g - g_0\| = o_P(1)$ : ■

Proof of Theorem 4. By definition

$$bAg - b_0Ag_0 = g - g_0$$

Write the left hand side of the last display as

$$b - b_0 \quad A g + b_0 \quad A \quad A \quad g_0 + b_0 A(g - g_0) + R;$$

where  $R = b - b_0 \quad A \quad A_0 \quad g + b_0 \quad A \quad A \quad (g - g_0)$ : Then, after noticing that (by definition of  $s$ ),

$$\|b_0 A(g - g_0)\|_s = \|hg - g_0\|_s;$$

we obtain

$$\|b - b_0\|_s \|g\|_s + \|b_0\|_s \|hg - g_0\|_s + \|R\|_s = 0;$$

By the proof of Theorem 3, it is straightforward to show that, for a  $C > 0$ :

$$\|R\|_s \leq C \left( \|A - A_{G_0}\|_s^2 + \|A - A_{G_0}\|_s \|g - g_0\|_s \right)$$

and

$$\begin{aligned} \|g - g_0\|_s &\leq \|g\|_s + \|g_0\|_s \\ &\leq \|M_{jj}A - A_{jj_{G_0}}\|_s + \|g\|_s \quad (\text{by } \|hg; g_0\| \rightarrow 0) \\ &\leq 2\|M_{jj}A - A_{jj_{G_0}}\|_s; \quad (\text{by } \|g\|_s \rightarrow 1) \end{aligned}$$

which implies by Assumption N.1

$$\|R\|_s = o_P(n^{-1/2});$$

Then, Cauchy-Schwarz inequality yields

$$\begin{aligned} \|R\|_s &\leq \|R\|_s \|s\|_s \\ &= o_P(n^{-1/2}); \end{aligned}$$

Then, by continuity of the inner product,  $\langle \mathbf{g}; \mathbf{s} \rangle = \langle \mathbf{g}_0; \mathbf{s} \rangle + o_p(1)$ ; and by Slutsky Theorem

$$\sqrt{n}(\mathbf{b} - \mathbf{b}_0) = \sqrt{n}(\mathbf{b}_0 - \mathbf{b}_0) + o_p(1) = o_p(1)$$

Hence, the result follows from Assumptions N.2 and N3. ■

**Proof of Theorem 5.** Define the operators  $L = \mathbf{b}_0 \mathbf{A}^{-1}$ ; and its estimator  $L_n = \mathbf{b}_n \mathbf{A}^{-1}$ : Then, by definition

$$\begin{aligned} 0 &= L\mathbf{g} - L\mathbf{g}_0 \\ &= L(\mathbf{g} - \mathbf{g}_0) + (L - L_n)\mathbf{g}_0 + (L - L_n)(\mathbf{g} - \mathbf{g}_0) \end{aligned} \quad (25)$$

First, from previous results it is straightforward to show as in Theorem 4

$$(L - L_n)(\mathbf{g} - \mathbf{g}_0) = o_p(n^{-1/2})$$

and

$$(L - L_n)\mathbf{g}_0 = \mathbf{b}_0(\mathbf{A} - \mathbf{A}_n)\mathbf{g}_0 = O_p(n^{-1/2})$$

Hence, in  $L^2$ ;

$$L(\mathbf{g} - \mathbf{g}_0) = \mathbf{b}_0(\mathbf{A} - \mathbf{A}_n)\mathbf{g}_0 + \mathbf{R}_n$$

where  $\mathbf{R}_n$  satisfies the conditions of the Theorem. ■

**Proof of Theorem 6.** Set  $(\mathbf{C}_i; \mathbf{V}_i) = \mathbf{C}_i \otimes \mathbf{g}(\mathbf{C}_i; \mathbf{V}_i) = \mathbf{c} \otimes \mathbf{g}(\mathbf{C}_i; \mathbf{V}_i)$ ; which estimates consistently  $(\mathbf{C}_i; \mathbf{V}_i) = \mathbf{C}_i \otimes \mathbf{g}_0(\mathbf{C}_i; \mathbf{V}_i) = \mathbf{c} \otimes \mathbf{g}_0(\mathbf{C}_i; \mathbf{V}_i)$ : Then, using standard empirical processes notation, write

$$\sqrt{n}(\mathbf{g}_n - \mathbf{g}_0) = \sqrt{n}(\mathbf{P}_n - \mathbf{P})(\mathbf{g}) + \sqrt{n}(\mathbf{P} - \mathbf{P})(\mathbf{g})$$

By the  $\mathbf{P}$ -Donsker property of  $D; \mathbf{P}(\mathbf{g} \in G)$  and the consistency of  $\mathbf{g}$ ;

$$\sqrt{n}(\mathbf{P}_n - \mathbf{P})(\mathbf{g}) = \sqrt{n}(\mathbf{P}_n - \mathbf{P})(\mathbf{g}_0) + o_p(1)$$

Since  $\mathbf{g} - \mathbf{g}_0$  is bounded with probability tending to one, we can apply integration by parts and use Assumption CE to write

$$\begin{aligned} \sqrt{n}(\mathbf{P} - \mathbf{P})(\mathbf{g}) &= \sqrt{n} \int \log(\mathbf{g}) - \log(\mathbf{g}_0); d\mathbf{i} + o_p(1) \\ &= \sqrt{n} \int \mathbf{g} - \mathbf{g}_0; d\mathbf{i} + o_p(1); \end{aligned}$$

where the last equality follows from the Mean Value Theorem and the lower bounds on  $\mathbf{g}$  and  $\mathbf{g}_0$ . Note that  $\int \mathbf{g}_0; d\mathbf{i} = \mathbf{E}[\mathbf{d}(\mathbf{C}; \mathbf{V})] = 0$ : Then, by Lemma B4

$$\sqrt{n}(\mathbf{P} - \mathbf{P})(\mathbf{g}) = \frac{\mathbf{b}_0}{\sqrt{n}} \sum_{i=1}^n \mathbf{d}_i(\mathbf{C}_i; \mathbf{V}_i) + o_p(1);$$



and therefore

$$\bar{g}_n(g_0) = \frac{1}{n} \sum_{i=1}^n (C_i; V_i) P b_0(C_i; V_i) + o_p(1):$$

The result then follows from Assumption CE.3. ■

## References

- [1] Abbott, B. and Gallipoli, G. (2018), "Permanent-Income Inequality," Technical report University of British Columbia.
- [2] Abramovich, Y. A. and Aliprantis, C. D. (2002). *Real Analysis with Economic Applications*. Graduate Studies in Mathematics 50. American Mathematical Society.
- [3] Ai, C. and X. Chen (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.
- [4] Alan, S., Attanasio, O. and M. Browning (2009), "Estimating Euler Equations With Noisy Data: Two Exact GMM Estimators," *Econometrica*, 24, 309-324.
- [5] An, Y. and Y. Hu (2012), "Well-posedness of measurement error models for self-reported data," *Journal of Econometrics*, 168, 259–269.
- [6] Anatolyev, S. (1999), "Nonparametric Estimation of Nonlinear Rational Expectation Models," *Econometrica*, 62, 1-6.
- [7] Andrews, D. W. K. (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometrica*, 11, 560–596.
- [8] Banks, J., R. Blundell, and S. Tanner (1998), "Is There a Retirement-Savings Puzzle?" *Econometrica*, 88, 769-788.
- [9] Battistin, E., R. Blundell, and A. Lewbel, (2009), "Why is consumption more log normal than income? Gibrat's law revisited," *Econometrica*, 117, 1140-1154.
- [10] Bosq, D. (2000), *Nonparametric Kernel Estimation*. Springer, New York.
- [11] Cai, Z., Ren, Y. and L. Sun, (2015), "Pricing Kernel Estimation: A Local Estimating Equation Approach," *Econometrica*, 31, 560-580.
- [12] Campbell, J. Y., and J. Cochrane, (1999), "Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Econometrica*, 107, 205-251.
- [13] Carrasco, M. and J. P. Florens (2000), "Generalization of GMM to a Continuum of Moment Conditions," *Econometrica*, 16, 797-834.

- [14] Carrasco, M., J.P. Florens and E. Renault (2007): "Linear Inverse Problems and Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Econometric Theory*, vol. 6, eds. J. Heckman and E. Leamer. North-Holland.
- [15] Chanchana, P. (2007), "An Algorithm for Computing the Perron Root of a Nonnegative Irreducible Matrix" Ph.D. Dissertation, North Carolina State University, Raleigh.
- [16] Chapman, D. A. (1997), "Approximating the Asset Pricing Kernel," *Econometric Theory*, 52, 1383–1410.
- [17] Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014), "Identification in Semiparametric and Nonparametric Conditional Moment Models," *Econometric Theory*, 82, 785-809.
- [18] Chen, X., Hansen, L. P. and J. Scheinkman (2000), "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions," unpublished manuscript.
- [19] Chen, X., Hansen, L. P. and J. Scheinkman (2009), "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions," *Econometric Theory*, 37, 4279–4312.
- [20] Chen, X., D.T. Jacho-Chavez and O.B. Linton, (2016), "Averaging of an Increasing Number of Moment Condition Estimators," *Econometric Theory*, 32, 30-70.
- [21] Chen, X. and S. C. Ludvigson (2009), "Land of addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Econometric Theory*, 24, 1057-1093.
- [22] Chen, X. and D. Pouzo (2009), "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Econometric Theory*, 152, 46-60.
- [23] Chen, X. and M. Reiss (2010), "On Rate Optimality For Ill-Posed Inverse Problems In Econometrics," *Econometric Theory*, 27, 497-521.
- [24] Christensen, T.M. (2015), "Nonparametric Identification of Positive Eigenfunctions", *Econometric Theory*, 31, 1310-1330.
- [25] Christensen, T.M. (2017), "Nonparametric Stochastic Discount Factor Decomposition", *Econometric Theory*, 85, 1501-1536.
- [26] Cochrane, J. (2001). *Asset Pricing*. Princeton University Press.

[28] Darolles, S., J. P. Florens and C. Gouriéroux (2004): "Kernel-based Nonlinear Canonical Analy-

- [42] Gayle, W.-R. and N. Khorunzhina (2014), "Micro-Level Estimation of Optimal Consumption Choice with Intertemporal Nonseparability in Preferences and Measurement Errors," Unpublished manuscript.
- [43] Gobet, E., Hoffmann, M. and Reiss, M. (2004), "Nonparametric Estimation of Scalar Diffusions Based on Low Frequency Data," *Journal of Applied Probability*, 26, 2223-2253.
- [44] Hall, R. E. (1978), 'Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,' *Journal of Political Economy*, 86, 971-987.
- [45] Hall, P., and J. L. Horowitz, (2005), "Nonparametric Methods for Inference in the Presence of Instrumental Variables" *Journal of the American Statistical Association*, 33, 2904-2929.
- [46] 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," *Journal of the American Statistical Association*, 15, 380-401.
- [47] Hansen, B. (2008), "Uniform Convergence Rates for Kernel Estimation with Dependent Data,"

- [56] Kubler, F. and K. Schmedders (2010): "Non-Parametric Counterfactual Analysis in Dynamic General Equilibrium," *Econometrica*, 78, 45, 181-200.
- [57] Kress, R. (1999). *Stochastic Processes and their Applications*. Springer.
- [58] Lawrance, E. C., (1991), "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Econometrica*, 59, 99, 54-77.
- [59] Lewbel, A. (1987), "Bliss Levels That Aren't," *Econometrica*, 55, 95, 211-215.
- [60] Lewbel, A. (1994), "Aggregation and Simple Dynamics," *Econometrica*, 62, 84, 905-918.
- [61] Lucas, R. E. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429-1445.
- [62] Luenberger, D. G. (1997). *Optimization by Vector Space Methods*. New York: John Wiley & Sons.
- [63] Mankiw, N. G., (1982), "Hall's Consumption Hypothesis and Durable Goods," *Econometrica*, 50, 10, 417-425.
- [64] Newey, W. and J. Powell (2003), "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1557-1569.
- [65] Newey, W. K., and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
- [66] Osborn, J. E. (1975), "Spectral Approximation for Compact Operators," *Journal of Functional Analysis*, 18, 29, 712-725.
- [67] Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer, Berlin.
- [68] Radulović, D., (1996), The bootstrap for empirical processes based on stationary observations. *Stochastic Processes and their Applications*, 65, 259-279.
- [69] Ross, S. A. (2015): "The Recovery Theorem," *Econometrica*, 83, 70, 615-648.
- [70] Rothenberg, T. J. (1971). "Identification in parametric models," *Econometrica*, 39, 577-591.
- [71] Sargan, J. D. (1983). "Identification and lack of identification." *Econometrica*, 51, 1605-1633.
- [72] Schaefer, H.H. (1974). *Stochastic Processes and their Applications*, Springer-Verlag, New York, Heidelberg, Berlin.

- [73] Stock, J., M. Yogo and J. Wright (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Econometrica*, 70, 1365-1401.
- [74] Tamer, E. (2010). "Partial identification in econometrics." *Annual Review of Economics*, 2(1), 167-195.
- [75] van der Vaart, A. W., and J. A. Wellner (1996). *Functional Data Analysis*, Springer Series in Statistics. Springer-Verlag, New York, 1 edn.
- [76] White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press Orlando, Florida.
- [77] Zeidler, E. (1986). *Asymptotic Theory of Econometric Surveys*. Springer-Verlag, New York, Inc.