Nonparametric Euler Equation Identi…cation and Estimation

Juan Carlos Escanciano Stefan Hoderlein Arthur Lewbel

Oliver Linton Sorawoot Srisuma

revised: March, 2020

Abstract

We consider nonparametric identi…cation and estimation of pricing kernels, or equivalently of marginal utility functions up to scale, in consumption based asset pricing Euler equations. Ours is the …rst paper to prove nonparametric identi…cation 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 identi…cation 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 …nite sample performance for our estimators.

JEL Codes: C14, D91, E21, G12. Keywords: Euler equations, marginal utility, pricing kernel, Fredholm equations, integral equations, nonparametric identi…cation, 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 Identi…cation and Estimation," by Lewbel and Linton (2010), and by Lewbel, Linton, and Srisuma (2012), and replaces "Nonparametric Identi…cation 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 ...rst-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

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

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

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

equation (2), thereby estimating g instead of M.[3] 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 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.[4] 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 $(g;b) \in \mathcal{G} \times (0;1)$, where $\mathcal{G}$ denotes the parameter space of marginal utility functions, which satisfies some conditions below. From equation (1) it is clear that, for a given b, the Euler equation cannot distinguish between g and 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

$S; S$ $\mathbb{R}^{\ell}$ $(C_t; V_t)$ $(C_{t+1}; V_{t+1})$

S $S \setminus S$ L^2 $L_2(S ; \ )$

$\langle g; f \rangle = \ \ gfd$

$\|g\|^2 = \langle g; g \rangle$

Let $M \ \ L^2$ be a linear subspace; and de…ne the linear operator $A : (M; \| \ \|) \ ! \ (M; \| \ \|)$ by

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

We assume that Ag is well-de…ned and $Ag \in M$ 3)

Ag

Theorem 1 shows that without further assumptions the Euler equation is partially identi…ed, with b identi…ed up to a …nite set corresponding to eigenvalues larger than one, and g is identi…ed up to a corresponding set of eigenfunctions. The discount factor b is also , meaning that for any b 2 B_0 there is an open neighborhood of b that does not contain any other element in B_0. Essentially, compactness of A ensures that B_0 is at most countable, and the economic restriction that discount factors lie in $(0; 1)$ ensures that B_0 is …nite.

The identi…ed 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 identi…ed set for each, and the true $(g; b)$ must lie in the intersection of these identi…ed sets. One might further shrink the identi…ed set by imposing the restriction that $bg(C_{t+1}; V_{t+1})R_{t+1} \quad g(C_t; V_t)$ is uncorrelated with all variables in the information set at time t, not just measurable functions of $(C_t; V_t)$.

Assumptions S and C do not su¢ ce for point identi…cation in general. We consider now a shape restriction on marginal utilities, which is a common behavioral assumption

We could consider other su¢ cient conditions that replace conditions on A by conditions on a power of $A;$ i.e. we could require that Assumptions C and I hold for $A^n;$ for some n 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 overidenti…ed 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 and M$;$ which lead to simple conditions for identi…cation by Theorem 2. Assume for simplicity that V_{t+1} and V_t are empty, and denote by $f(c;c);$ $f(c)$ and $f(c)$ the joint and marginal densities of $(C_{t+1};C_t);$ respectively. Assume has Lebesgue density f on a common support $S = S = S$ (e.g. $S = [0;1))$: Then, taking M equals to $L^2;$ the operator equation $bAg = g$ can be written as

$$b \quad k(c;c)g(c)f(c)dc = g(c);$$

where $k(c;c) = r(c;c)f(c;c)/[f(c)f(c)]$ and $r(c;c) = E[R_{t+1}|C_{t+1} = c;C_t = c]$ a.s. Then, it is well known that Assumption C holds if

$$k^2(c;c)f(c)f(c)dc\,dc < 1;$$

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

$_i(c; v); \, i = 1; :::; n):$ Therefore, similar to our discussion of identi…cation in Section 3, the number of eigenvalues and eigenfunctions of A is …nite 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 $n^{\ 1} \ {}^{n}_{i=1} \ {}_{i} \ {}_{i}(c; v);$ for some coe¢ cients $\ _i; \, i = 1; :::; n;$ satisfying for its corresponding eigenvalue the equation

$$\frac{1}{n^2} \ {}^{n}_{i=1} \ {}^{n}_{j=1} \ {}_{j} \ {}_{j}(C_i; V_i) R_i \ {}_{i}(c; v) = \frac{1}{n} \ {}^{n}_{i=1} \ {}_{i} \ {}_{i}(c; v):$$

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

$$\frac{1}{n} \ {}^{n}_{j=1} \ {}_{j} \ {}_{j}(C_i; V_i) R_i = \ {}_{i};$$

which in matrix notation can be simply written as

$$A_n \ = \ ;$$

where A_n is an $n \ \ n$ matrix with ij-th element $a_{ij} = \ _j(C_i; V_i) R_i = n;$ and $\ = (\ _1; :::; \ _n) \$ (henceforth, v denotes the transpose of v)$:$ Thus, let denote the largest eigenvalue in modulus of A_n and $\ = (\ _1; :::; \ _n) \$ its corresponding eigenvector. Our estimators for b_0 and g_0 are, respectively,

$$\hat{b} = 1= \qquad \text{and} \qquad g(c; v) = n^{\ 1} \ {}^{n}_{i=1} \ {}_{i} \ {}_{i}(c; v): \tag{7}$$

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

$$! _{ij} = \frac{1}{n^2} \ {}_{i}(c; v) \ {}_{j}(c; v) f \ (c; v) dcdv:$$

As a practical recommendation, we could also normalize $g(C_i; V_i)$ to have unit standard deviation. Also, we impose the sign normalization $\langle g; 1 \rangle > 0:$ The estimator $(g; \hat{b})$ can be easily obtained with any statistical package that computes eigenvalues and eigenvectors of matrices. There are also e¢ cient algorithms for the computation of the so-called Perron-Frobenius root $;$ see e.g. Chanchana (2007(r)(c)9(e)

The easiest way to consider simultaneously di¤erent 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 $\hat{b}_j$

…rst order behavior of $\hat{b}$; 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 identi…cation hold. We can then test the restriction $b_1 = b_2$ (where b_j^2

:

1. g $\|g\| = 1$ $\langle g, 1\rangle > 0.$

2. $\|A \quad A\|_{G_0} \to_p 0$

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 G_0 consists of the linear span of g_0. More generally, under Assumption C, 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.

. $\hat{b} \to_p b_0$ $\|g \quad g_0\| \to_p 0$

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 A and A associated to the eigenvalues $\hat{b}^{-1} = \quad (A)$ and $b_0^{-1} = \quad (A)$

3.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\varepsilon_i \xrightarrow{d} N(0; {}_s);$$

${}_s \quad \lim_n \quad \mathrm{var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i\varepsilon_i\right) < \infty:$

. $\qquad n \to \infty;$

$$\sqrt{n}\left(b - b_0\right) \xrightarrow{d} N\left(0; b_0^4 \, {}_s\right):$$

The proof of Theorem 4 can be found in the Appendix. We can estimate the asymptotic variance of b by standard long run variance estimators based on $\{s_i\varepsilon_i\}_{i=1}^{n};$ see e.g. Newey and West (1987), where $\varepsilon_i = g(C_i; V_i)R_i - b^{-1}g(C_i; V_i);$ and s is computed as our estimator $g;$ with the normalization $\langle g; s\rangle_n = 1:$ An alternative to plug-in asymptotic methods is to use block bootstrap, see e.g. Radulović (1996).

For the estimator based on J assets proposed in Section 4.1, note that

$$\begin{aligned}\sqrt{n}\,(\hat{w}_b)\left(\hat{b}^{(J)} - b_0\right) &= (\hat{w}_b)\sqrt{n}\left(\hat{b}^{(J)} - b_0\right)\\ &+ \sqrt{n}\,(\hat{w}_b - w_b)\, b_0 :\end{aligned}$$

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,

$$\begin{aligned}\sqrt{n}\,(\hat{w}_b)\left(\hat{b}^{(J)} - b_0\right) &= \sqrt{n}\,(w_b)\left(\hat{b}^{(J)} - b_0\right) + o_P(1)\\ &\xrightarrow{d} N\left(0; b_0^4 (w_b) \, {}_J w_b\right);\end{aligned}$$

where ${}_J$ is defined in (8).

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

$$\sqrt{nh_n^{\ell}}\; {}_n(c; v) \xrightarrow{d} N(0; \ (c; v));$$

with $(c; v) = f^{-1}(c; v)\ {}^2(c; v)\ {}_2;\ {}_2 = \int K^2(u)du$ and ${}^2(c; v) = E[\varepsilon_i^2 | C_i = c; V_i = v]$ 6.03553250.7071 Tf 9.9552

p

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

$$\mathcal{D} = \left\{ (c;v) \to c\frac{\partial \log(g(c;v))}{\partial c} : g \in \mathcal{G} \right\}; \tag{12}$$

and the functions

$$d(c;v) \quad \frac{\partial\,(c \quad f(c;v))}{\partial c}\frac{1}{f(c;v)} \qquad \text{and} \qquad (c;v) \quad \frac{d(c;v)}{g_0(c;v)}. \tag{13}$$

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

$$\ _s = \quad \langle g_0; \ \rangle \langle g_0; s\rangle^{\ 1} s. \tag{14}$$

The function $\ _s$ has a geometrical interpretation as the value of 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 $\ _s = L\ r$ in r; i.e. $\ _s = \arg\min\{\|r\| : \ _s = L\ r\}$; which is well defined because $\ _s \in \mathcal{N}\ (L) = \mathcal{R}(L\ )$; see Luenberger (1997, Theorem 3, p. 157) for the latter equality. Here $\mathcal{N}\ (L)$ 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 $\mathcal{C}\ (T)$ for a generic closed and convex set T. For any vector a of ℓ integers define the differential operator $\partial_x^a \quad \partial^{|a|_1}/\partial x_1^{a_1} \ldots \partial x_\ell^{a_\ell}$; where $|a|_1 \quad \sum_{i=1}^{\ell} a_i$. For any smooth function $h : T \quad \mathbb{R}^\ell \to \mathbb{R}$ and some $\ > 0$, let $\underline{\ }$ be the largest integer smaller or equal than , and

$$\|h\|_{\ ;} \quad \max_{|a|_1 \ \underline{\ }} \sup_{x \ T} |\partial_x^a h(x)| + \max_{|a|_1 = \underline{\ }} \sup_{x \neq x'} \frac{|\partial_x^a h(x) \quad \partial_x^a h(x')|}{|x \quad x'|^{\ \underline{\ }}}.$$

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

1. D P- 5

2. $(C; V)$ S $S =$ $[l_c; u_c]$ $S_V;$ $l_c; u_c$ $l_c < u_c:$ $\lim_{c \to l_c} cf(c; v) = 0 = \lim_{c \to u_c} cf(c; v)$ $v \in S_V$ $P(\min\{g_0; g\} > \varepsilon) \to 1$ $\varepsilon > 0$

3. d $d \in L^2; \{ \ _i\}$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ _i \xrightarrow{d} N(0; \ );$$

$\lim_n$ var $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ _i < \infty$ $_s \in C^r(S)$

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 $_n(g)$ is well de.gned. Assumption CE(3) implies that

g 210(r)10T 0 3 8-5077lTJ/-15.5TJ/F23 7 Tf 23.13m6J/F(1)1nC

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 $QRRA$ (quartile relative risk aversion).

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

$$\sqrt{n}(\ _n(q) - \ (q)) \to^{d} N\left(0, \ ^2(q)\right),$$

for a suitable asymptotic variance $\ ^2(q)$, $q = 1, 2, 3$ and 4. Moreover, by definition, $\sqrt{n}(\ _n(q) - \ (q))$ and $\sqrt{n}(\ _n(j) - \ (j))$ 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 $q \neq j$,

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

The t-statistics are constructed as

$$t_{qj} = \frac{\sqrt{n}(\ _n(q) - \ _n(j))}{\sqrt{\ ^2_n(q) + \ ^2_n(j)}},$$

for suitable consistent estimates $\ ^2_n(q)$ of the asymptotic variances $\ ^2(q)$, 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

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

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

ARRA. The model is then given by the Euler equation

$$b_0 E\left[C_{t+1}^{-\gamma_0} R_{t+1} \mid C_t\right] = C_t^{-\gamma_0}.$$

We set $b_0 = 0.95$ and $\gamma_0 = 0.5$. We draw a random sample of (C_t, C_{t+1}) from the distribution

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

and construct $R_{t+1} = b_0^{-1}(1+\varepsilon_t)(C_{t+1}/C_t)^{\gamma_0}$, where ε_t is distributed uniformly on $[-0.5, 0.5]$ and drawn independently of (C_t, C_{t+1}). This design was chosen to generate data that satisfies the Euler equation model, has realistic parameter values and consumption distribution, and avoids the apfunction g is then recovered [redacted]ation $g(c; v) = g(c; v)=c$. Throughout we set the ban[redacted] to be $1{:}06\ n^{1=3:5}$, where [redacted]le standard deviation of C_t. This is essentially Silve[redacted] rule applied to the rate $n^{1=}$[redacted]ur estimators for g_0 are normalized to have a unit st[redacted] deviation.

For each …nite-dimensio[redacted]er and summary measure we consider, we report the[redacted] standard deviation, $2{:}5$th p[redacted]$:5$th percentile, 95% coverage probability based on [redacted] distribution, their bootstrap[redacted]s and the root mean square error.[6] Table 1 reports es[redacted] of the discount factor from [redacted]stimators, ***CRRA***, ***NP*** 1, and ***NP*** 2. Table 2 [redacted] estimates of the $_0$

estimates of the marginal utility function tend to be less accurate at higher consumption levels. This can also be seen for $\boldsymbol{NP}$ 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
$n = 500$	***CRRA***		0.000	0.012	0.926	0.975	0.946	0.012	0.926	0.974	0.940	0.012
	NP	1	0.006	0.027	0.917	0.971	0.984	0.018	0.915	0.980	0.929	0.028
	NP	2	0.009	0.041	0.808	0.983	0.963	0.031	0.895	1.012	0.932	0.042
$n = 2000$	***CRRA***		0.000	0.006	0.938	0.961	0.960	0.006	0.938	0.962	0.950	0.006
	NP	1	0.004	0.020	0.936	0.960	0.992	0.009	0.932	0.965	0.924	0.020
	NP	2	0.005	0.028	0.862	0.965	0.974	0.021	0.922	0.994	0.946	0.028

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

	QRRA	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
n = 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
n = 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

	$(C_{t+1}; C_t)$	Bias	Std	Lpc	Upc	Cov	B-Std	B-Lpc	B-Upc	B-Cov	Rmse
$n = 500$	C_{t+1}	-0.002	0.111	-0.111	0.132	0.975	0.118	-0.255	0.200	0.975	0.111
	C_t										

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% con…dence interval, and the true.

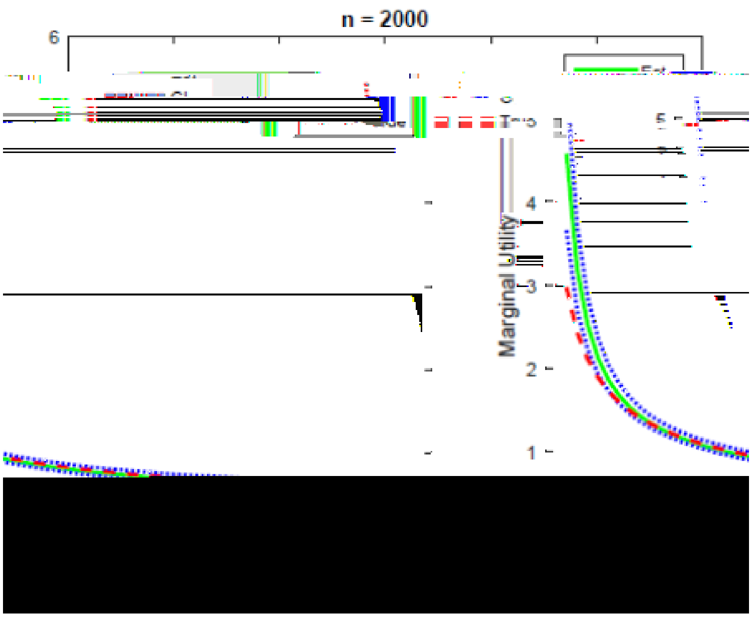

Figure 2: Estimates of the marginal utility function g_0 using simulated data with n = 2000. ***Est***, ***CI***, and ***True*** represent respectively the one-dimensional nonparametric estimator, its 95% con…dence interval, and the true.

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 e¤ects. 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 a¤ect utility and are known at time t. Let V_t denote the vector of all variables other than C_t that a¤ect utility in time t. In particular, V_t contains Z_t, V_t contains D_t if durables matter, and V_t contains lagged consumption C_{t-1}, C_{t-2} and so on if habits matter.

The consumer's time separable utility function is

$$\max_{C_t; D_t \; t=1} E \sum_{t=0} b^t U(C_t; V_t) \; :$$

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 . Then without frictions the consumer's budget constraint can be written as, for each period t,

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

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

$$E \sum_{t=0}^{T} b^t U(C_t; V_t) \quad C_t + (D_t \quad D_{t-1}) P_t + \sum_{j=1}^{J} (A_{jt} - A_{jt-1} R_{jt}) \quad {}_t \qquad (17)$$

with Lagrange multiplier[(t)]TJ/F151rasse1 wanoj8Td[(C)]TJ/F237.J-7l8 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

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

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

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

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

$$\begin{aligned} {}_t &= b^t g(I_t) \\ {}_t &= E[\ {}_{t+1} R_{jt+1} \mid I_t] \qquad j = 1, \ldots, J \\ {}_t P_t &= b^t g_d(C_t, V_t) \quad E[\ {}_{t+1} P_{t+1} \mid I_t] \end{aligned}$$

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

$$\begin{aligned} b^t g(I_t) &= E\ \ b^{t+1} g(I_{t+1}) R_{jt+1} \mid I_t \qquad j = 1, \ldots, J \\ b^t g(I_t) P_t &= b^t g_d(C_t, V_t) \quad E\ \ b^{t+1} g(I_t \end{aligned}$$

3.

$$\sup_{l_n \quad h \quad u_n} \sup |m_h(\ ) \quad m(\ )| = O_P \quad \overline{\frac{1}{n l_n^{`}}} + u_n^r \quad . \tag{20}$$

Proof. By the Triangle inequality

$$|m_h(\ ) \quad m(\ )|$$

$$m_h(\ ) \quad \frac{E[T_h(\ )]}{E[f(c;v)]} + \frac{E[T_h(\ )]}{E[f(c;v)]} \quad m(\ )$$

$$\frac{1}{f(c;v)} \ T_h(\ ) \quad E[T_h(\ )] + \frac{E[T_h(\ )]}{f(c;v)\,E[f(c;v)]} \ f(c;v) \quad E[f(c;v)]$$

$$+ \frac{1}{E[f(c;v)]} \ E[T_h(\ )] \quad T(\ ) + \frac{|T(\ )|}{E[f(c;v)]f(c;v)} \ E[f(c;v)] \quad f(c;v) \ ;$$

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

Define the class of functions

$$c;v$$

$$T$$

$$K_0 := \quad (C_i; V_i \quad c j$$

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

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

$\mathcal{L}^2(\mathbf{r})$ $\varphi \in \mathcal{L}^2$ φ $j=$ $E \; \varphi_i \varepsilon_i \varphi_j \varepsilon_j <$
1 φ $\mathbf{r}$

Lemma B3. $\varphi \in \mathcal{L}^2(\mathbf{r});$

$$\sqrt{n} \quad A \quad A \;\; g_0; \varphi \quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_i \varepsilon_i + o_P(1);$$

$$\sqrt{n} \quad A \quad A \;\; g_0; \varphi \quad \xrightarrow{d} N(0; \;\; \cdot):$$

Proof. Define

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

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

$$A \quad A \;\; g_0(c;v) = a_n(c;v) + r_n(c;v),$$

where

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

$Tg_0(c;v) \quad f(c;v)\, Ag_0(c;v); Tg_0(c;v) \quad f(c;v)\, Ag_0(c;v)$ and

$$r_n(c;v) \qquad \frac{f(c;v) \quad f(c;v)}{f(c;v)} a_n(c;v):$$

Lemma B1 and our conditions on the bandwidth imply $\|r_n\| = o_P(n^{\;\;1=2})$. It then follows that $A \quad A \;\; g_0; \varphi$ has the following expansion

$$\varphi(c;v)[Tg_0(c;v) \quad Tg_0(c;v)]dcdv \tag{21}$$

$$\varphi(c;v) Ag_0(c;v)\,[f(c;v) \quad f(c;v)]dcdv \tag{22}$$

$$+\, o_P(n^{\;\;1=2}).$$

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

Hence,

$$\int\int \varphi(c;v)[Tg_0(c;v) - Tg_0(c;v)]dcdv = \int\int \varphi(c;v)[Tg_0(c;v) - E(Tg_0(c;v))]dcdv + o_P(n^{-1/2})$$

$$= \frac{1}{n}\sum_{i=1}^{n} g_{0i}R_i \left[\int\int \varphi(c;v)K_{hi}(c;v)\,dcdv - \int\int \varphi(c;v)E(g_0R_iK_{hi}(c;v))dcdv\right] + o_P(n^{-1/2}),$$

9.3 Main Proofs

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

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

Proof of Theorem 2. Let A^* denote the adjoint of $A;$ which is also compact and positive by well known results in functional analysis. Assumption S implies that $\rho(A) > 0.$ Also notice that the eigenvalues of A^* are complex conjugates of those of A (in particular, $\rho(A) = \rho(A^*)$)$.$ Then, by the Kreĭn-Rutman's theorem (see Theorem 7.C in Zeidler (1986, vol. 1, p. 290)) there is exactly one solution to $bAg = g$ with $g > 0$ and $\|g\| = 1$ and a solution to $bA^*s = s$ with $s > 0$. Note $\langle g; s\rangle = b\langle Ag; s\rangle = b\langle g; A^*s\rangle = b\rho(A)\langle g; s\rangle$. Hence, since g and s are strictly positive, $\langle g; s\rangle \neq 0;$ and then $b = \rho^{-1}(A)$. ■

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

$$\left|\hat{b}^{-1} - b_0^{-1}\right| \leq M\|\hat{A} - A\|_{\mathcal{G}_0} \tag{23}$$

and

$$\|\hat{g} - \tilde{g}\| \leq M\|\hat{A} - A\|_{\mathcal{G}_0}; \tag{24}$$

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

$$\begin{aligned}\left|\hat{b} - b_0\right| &\leq M \hat{b} b_0 \|\hat{A} - A\|_{\mathcal{G}_0} \\ &\leq M\|\hat{A} - A\|_{\mathcal{G}_0};\end{aligned}$$

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

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

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

so $\langle \tilde{g}, g_0 \rangle \geq 0$ for large enough $n.$

Next,

$$\begin{aligned} 1 &= \|\hat{g}\| \text{ (by normalization)} \\ &= \|\tilde{g}\| + o_P(1) \text{ (by } \|\hat{g} - \tilde{g}\| \leq M\|\hat{A} - A\|_{G_0}) \\ &= |\langle \tilde{g}, g_0 \rangle| + o_P(1); \text{ (by definition of } \tilde{g}) \end{aligned}$$

which then implies $\|\tilde{g} - g_0\| = |\langle \tilde{g}, g_0 \rangle - 1| = o_P(1).$ Hence, by the triangle inequality, $\|\hat{g} - g_0\| = o_P(1).$ ■

Proof of Theorem 4. By definition

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

Write the left hand side of the last display as

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

where $R = \left(\hat{b} - b_0\right)\left(\hat{A} - A_0\right)\hat{g} + b_0\left(\hat{A} - A\right)(\hat{g} - g_0).$ Then, after noticing that (by definition of s),

$$\langle b_0 A(\hat{g} - g_0); s \rangle = \langle \hat{g} - g_0; s \rangle;$$

we obtain

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

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

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

and

$$\begin{aligned} \|\hat{g} - g_0\| &\leq \|\hat{g} - \tilde{g}\| + \|\tilde{g} - g_0\| \\ &\leq M\|\hat{A} - A\|_{G_0} + |\|\tilde{g}\| - 1| \text{ (by } \langle \tilde{g}; g_0 \rangle \geq 0) \\ &\leq 2M\|\hat{A} - A\|_{G_0}; \text{ (by } |\|\tilde{g}\| - 1| \leq \|\hat{g} - \tilde{g}\|) \end{aligned}$$

which implies by Assumption N.1

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

Then, Cauchy-Schwarz inequality yields

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

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

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

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

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

$$\begin{aligned} 0 &= \hat{L}\hat{g} - L g_0 \\ &= L(\hat{g} - g_0) + (\hat{L} - L)g_0 + (\hat{L} - L)(\hat{g} - g_0). \end{aligned} \tag{25}$$

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

$$\left\|(\hat{L} - L)(\hat{g} - g_0)\right\| = o_P(n^{-1/2})$$

and

$$\left\|(\hat{L} - L)g_0 - b_0(\hat{A} - A)g_0\right\| = O_P\left(n^{-1/2}\right).$$

Hence, in $L^2;$

$$L(\hat{g} - g_0) = -b_0(\hat{A} - A)g_0 + R_n;$$

where R_n satisfies the conditions of the Theorem. ■

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

$$\sqrt{n}\left(\hat{\theta}_n(\hat{g}) - \theta(g_0)\right) = \sqrt{n}\left(P_n \hat{\psi} - P\hat{\psi}\right) + \sqrt{n}\left(P\hat{\psi} - P\psi\right).$$

By the P-Donsker property of $\mathcal{D}; P(\hat{g} \in \mathcal{G}) \to 1$ and the consistency of $\hat{g};$

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

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

$$\begin{aligned} \sqrt{n}\left(P\hat{\psi} - P\psi\right) &= \sqrt{n}\langle \log(\hat{g}) - \log(g_0); d\rangle + o_P(1) \\ &= \sqrt{n}\langle \hat{g} - g_0; \delta\rangle + o_P(1); \end{aligned}$$

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

$$\sqrt{n}\left(P\hat{\psi} - P\psi\right) = \frac{b_0}{\sqrt{n}} \sum_{i=1}^{n} \phi_s(C_i; V_i)\varepsilon_i + o_P(1);$$

and therefore

$$\sqrt{n}\,(\quad _n(g) \quad (g_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\; (C_i; V_i) \quad P \;) \quad b_0' \; _s(C_i; V_i)''_i + o_P(1):$$

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). . Graduate Studies in Mathematics 50. American Mathematical Society.

[3] Ai, C. and X. Chen (2003), “E¢ cient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions,” , 71, 1795-1844.

[4] Alan, S., Attanasio, O. and M. Browning (2009), “Estimating Euler Equations With Noisy Data: Two Exact GMM Estimators,” , 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,” , 62, 1-6.

[7] Andrews, D. W. K. (1995), “Nonparametric Kernel Estimation for Semiparametric Models,” 11, 560–596.

[8] Banks, J., R. Blundell, and S. Tanner (1998), “Is There a Retirement-Savings Puzzle?” , 88, 769-788.

[9] Battistin, E., R. Blundell, and A. Lewbel, (2009), “Why is consumption more log normal than income? Gibrat’s law revisited,” , 117, 1140-1154.

[10] Bosq, D. (2000), . Springer, New York.

[11] Cai, Z., Ren, Y. and L. Sun, (2015), “Pricing Kernel Estimation: A Local Estimating Equation Approach,” , 31, 560-580.

[12] Campbell, J. Y., and J. Cochrane, (1999), “Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” , 107, 205-251.

[13] Carrasco, M. and J. P. Florens (2000), “Generalization of GMM to a Continuum of Moment Conditions,” , 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,” , 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,” , 52, 1383–1410.

[17] Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014), “Identi…cation in Semiparametric and Nonparametric Conditional Moment Models,” , 82, 785-809.

[18] Chen, X., Hansen, L. P. and J. Scheinkman (2000), “Nonlinear Principal Components and Long-Run Implications of Multivariate Di¤usions,” unpublished manuscript.

[19] Chen, X., Hansen, L. P. and J. Scheinkman (2009), “Nonlinear Principal Components and Long-Run Implications of Multivariate Di¤usions,” , 37, 4279–4312.

[20] Chen, X., D.T. Jacho-Chavez and O.B. Linton, (2016), “Averaging of an Increasing Number of Moment Condition Estimators,” , 32, 30-70.

[21] Chen, X. and S. C. Ludvigson (2009), “Land of addicts? An Empirical Investigation of Habit-Based Asset Pricing Models,” , 24, 1057-1093.

[22] Chen, X. and D. Pouzo (2009), “E¢ cient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals,” , 152, 46-60.

[23] Chen, X. and M. Reiss (2010), “On Rate Optimality For Ill-Posed Inverse Problems In Econometrics,” , 27, 497-521.

[24] Christensen, T.M. (2015), “Nonparametric Identi…cation of Positive Eigenfunctions”, , 31, 1310-1330.

[25] Christensen, T.M. (2017), “Nonparametric Stochastic Discount Factor Decomposition”, , 85, 1501-1536.

[26] Cochrane, J. (2001). 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., Ho¤mann, M. and Reiss, M. (2004), "Nonparametric Estimation of Scalar Di¤usions Based on Low Frequency Data," , 26, 2223-2253.

[44] Hall, R. E. (1978), 'Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,'

[56] Kubler, F. and K. Schmedders (2010): “Non-Parametric Counterfactual Analysis in Dynamic General Equilibrium,” , 45, 181-200.

[57] Kress, R. (1999). . Springer.

[58] Lawrance, E. C., (1991), “Poverty and the Rate of Time Preference: Evidence from Panel Data,” , 99, 54-77.

[59] Lewbel, A. (1987), “Bliss Levels That Aren't,” 95, 211-215.

[60] Lewbel, A. (1994), “Aggregation and Simple Dynamics,” 84, 905-918.

[61] Lucas, R. E. (1978): “Asset Prices in an Exchange Economy,” , 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," , 10, 417-425.

[64] Newey, W. and J. Powell (2003), ”Instrumental Variables Estimation of Nonparametric Models,” , 71, 1557-1569.

[65] Newey, W. K., and K. D. West (1987), “A Simple, Positive Semi-De…nite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” , 55, 703-708.

[66] Osborn, J. E. (1975), “Spectral Approximation for Compact Operators,” , 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,” , 70, 615-648.

[70] Rothenberg, T. J. (1971). “Identi…cation in parametric models,” Econometrica, 39, 577-591.

[71] Sargan, J. D. (1983). “Identi…cation and lack of identi…cation.” Econometrica, 51, 1605-1633.

[72] Schaefer, H.H. (1974). , 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," , 20, 518-529.

[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).