UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

Nonparametric Euler Equation Identification and Estimation

Juan Carlos Escanciano
Universidad Carlos III de Madrid

Stefan Hoderlein
Boston College

Arthur Lewbel
Boston College

Oliver Linton
University of Cambridge

Sorawoot Srisuma
University of Surrey

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.

Reference Details

2064 Cambridge Working Papers in Economics
2020/60 Cambridge-INET Working Paper Series

Published 4 July 2020

Key Words Euler equations, marginal utility, pricing kernel, Fredholm equations, integral equations, nonparametric identification, asset pricing
JEL Codes C14, D91, E21, G12

Websites www.econ.cam.ac.uk/cwpe
www.inet.econ.cam.ac.uk/working-papers

Nonparametric Euler Equation Identification and Estimation*

Juan Carlos Escanciano
Universidad Carlos III de Madrid

Stefan Hoderlein
Boston College

Arthur Lewbel
Boston College

Oliver Linton
University of Cambridge

Sorawoot Srisuma
University of Surrey

July 04, 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 paper we study the nonparametric identification and estimation of such Euler equations.

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

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

where b is the subjective discount factor, C_t is consumption at time t, V_t is a vector of other economic variables such as durables or lagged consumption (for habits) that might affect utility, R_t is the gross return of an asset, and g is the time homogeneous marginal utility function of consumption.[1] Equation (1) is the first order condition that equates in real terms the marginal cost of an extra unit of the asset, purchased today, to the expected marginal benefit of the extra payoff received tomorrow.[2]

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

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

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

[1] This model assumes time separability, however, the separability is conditional, in that it depends on V_t as well as C_t. So our model permits current utility to depend on many lags of consumption to accommodate habits or durables (e.g. see Campbell and Cochrane (1999)).

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

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

Having established identification, we next propose a novel nonparametric kernel estimator for the marginal utility function and discount factor based on our identification arguments. We provide asymptotic distribution theory for the discount factor, the marginal utility function, and for semiparametric functionals of the marginal utility function such as the 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

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

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

Nonparametric estimators of equation (2) and similar models (taking the form of nonparametric instrumental variables models) have been proposed, by, e.g., Gallant and Tauchen (1989), Chapman (1997), Newey and Powell (2003), Ai and Chen (2003) and Darolles, Fan, Florens, and Renault (2011). However, in these applications identification is assumed rather than proved, by way of high level completeness assumptions. These models have the structure of Fredholm equations of the first kind (also called Type I equations). Solving these types of equations involves ill-posed inverse problems that can be severe, and as a result, fully nonparametric estimators of $M_{t+1} = M(C_{t+1}, V_{t+1}, C_t, V_t)$ based on (2) can have very slow convergence rates and possibly unstable inference.

In contrast, we start by writing the pricing kernel problem in the form of equation (1) instead of

equation (2), thereby estimating g instead of M.[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 impose the additional behavioral restriction that utility is increasing in consumption, which implies that the function g is positive. Applying an infinite-dimensional extension of the Perron-Frobenius theorem (see Kreĭn and Rutman (1950)) yields uniqueness of a positive eigenvalue-eigenfunction pair, which then provides nonparametric point identification.

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

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

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

[3]This simplification does not come for free. It requires that the pricing kernel model be derived from an Euler equation model of the form given by equation (2).

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

The asymptotic theory we present in this paper is based on weakly dependent stationary household-level consumption and asset data. Therefore our existing results would not be suitable to analyze aggregated consumption data that appear to be non-stationary (particularly non-recurrent time series). Otherwise, our estimates allow applied researchers to engage in a nonparametric consumption-based asset pricing theory. For example, Abbott and Gallipoli (2018) illustrate the applicability of our methods by using them to estimate nonparametrically human wealth and permanent income with PSID data.

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

2 Literature Review

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

Nonparametric instrumental variables is a leading example of estimation based on a Type I Fredholm equation, yielding associated ill-posed inverse problems on estimation. Newey and Powell (2003) note that assuming statistical completeness (a high level assumption) is essentially the same as just assuming identification of this type of model. Other related examples of nonparametric and semiparametric ill-posed inverse estimation problems include Carrasco and Florens (2000), Ai and Chen (2003), Hall and Horowitz (2005), Chen and Pouzo (2009), Chen and Reiss (2010), Darolles, Fan, Florens and Renault (2011) and, more recently, Cai, Ren and Sun (2015). A particularly relevant example is Chen and Ludvigson (2009), who studied identification and estimation of a semiparametric specification of the Type-I equation (2). Their model assumes g has the semiparametric form

$g(C_t, V_t) = C_t^{\eta} h(V_t)$, where η is a constant that determines risk aversion and h is an unknown function of current and lagged values of C_t/C_{t-1} representing habits. Virtually all parametric estimators of the asset pricing model, going back to Hansen and Singleton (1982) and including Dunn and Singleton (1986), and Campbell and Cochrane (1999), use the form of equation (2) rather than equation (1).

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

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

Some versions of Perron-Frobenius have been applied before in Euler equation and other similar models. Hansen and Scheinkman (2009, 2012, 2013) used Perron-Frobenius theory for a different problem of identification than ours in a continuous-time setting, using Markov theory. In our notation, they give conditions for identification of the positive eigenfunction and eigenvalue of the operator $\phi \rightarrow E[M_{t+1}\phi(C_{t+1}, V_{t+1}) \mid C_t, V_t]$, assuming that the SDF M_{t+1} is known. In contrast, we show that M_{t+1} itself is identified, by obtaining identification of b and g. Christensen (2015, 2017) applies Kreĭn-Rutman theory to a discrete-time version of Hansen and Scheinkman (2009), and Ross (2015) applies the classical finite-dimensional Perron-Frobenius theorem to identify the pricing kernel and the natural probability distribution from state prices.

Regarding identification, perhaps the closest work to ours is Chen, Chernozhukov, Lee and Newey (2014). Although their paper mainly concerns local nonparametric identification, in their Euler equation application they consider a semiparametric rather than a nonparametric model like ours. Specifically, their model is the same functional form as Chen and Ludvigson (2009) described above, but allowing for a more general conditioning set. They cite working paper versions of our paper as

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 $(g, b) \in \Theta \equiv \mathcal{G} \times (0, 1)$, where $\mathcal{G}$ denotes the parameter space of marginal utility functions, which satisfies some conditions below. From equation (1) it is clear that, for a given b, the Euler equation cannot distinguish between g and 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] Section 2.5 in Carrasco et al. (2007) discusses generic methods for estimation of the singular values and associated eigenfunctions. This problem is different from estimating the principal eigenvalue and eigenfunction of a non-selfadjoint operator, but our kernel estimation strategy follows closely their logic.

DEFINITION. *Let $S, S' \subseteq \mathbb{R}^\ell$ denote the supports of (C_t, V_t) and (C_{t+1}, V_{t+1}) respectively. Let μ be a probability measure, with support $S_\mu \subseteq S \cap S'$, and let $\mathcal{L}^2$ denote the Hilbert space $L_2(S_\mu, \mu)$ of (equivalence classes of) square μ-integrable functions equipped with the inner product $\langle g, f \rangle = \int g f d\mu$ and the corresponding norm $\|g\|^2 = \langle g, g \rangle$ (we drop the domain of integration for simplicity of exposition).*

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

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

We assume that Ag is well-defined and $Ag \in \mathcal{M}$ for $g \in \mathcal{M}$. Examples of μ and $\mathcal{M}$ are given below. With our notation, (1) can be written in a compact form as $bAg = g$. The parameter space for g, $\mathcal{G}$, will be a subset of $\mathcal{M}$ incorporating normalization restrictions. Marginal utilities may not have finite moments around zero (where they may diverge). To overcome this problem, by suitable redefinition of g we can rewrite equation (1) in the form

$$bE[C' g(C', V') \, (C/C') \, R' \mid C, V] = C g(C, V). \tag{4}$$

This reparameterizes the problem in terms of $Cg(C, V)$, which under natural economic assumptions is bounded; see Lucas (1978). This identity also gives an alternative way to estimate the marginal utility function and other objects of interest, which we shall discuss further below. More generally, we could apply parametrizations with powers of consumption if necessary (i.e. with $(C')^\delta \, g(C', V') \, (C/C')^\delta$ and $C^\delta g(C, V)$ in the left and right hand side of (4), respectively). The parameter δ is chosen by the researcher (i.e. it is not estimated). Allowing for $\delta > 1$ is useful to accommodate CRRA utilities with risk aversion larger than one.

We introduce the assumption of correct specification and a formal definition of identification.

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

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

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

Therefore, the Euler equation is point identified if Θ_0 is a singleton. To provide some insights on our identification and estimation strategies we consider first the case where A in (3) has a finite-dimensional range. In this case, we can write

$$Ag(\cdot) = \sum_{i=1}^{I} L_i(g)\phi_i(\cdot), \tag{5}$$

for a set of functions $\{\phi_i\}$ that span the range of A, $\mathcal{R}(A) = \{f \in \mathcal{M} : \exists g \in \mathcal{M},\ Ag = f\}$, and linear operators $L_i(g)$, $i = 1, \ldots, I$. This situation arises, for example, when the support S is discrete and finite. Under (5), any potential solution of (1) has to have necessarily the form $g(\cdot) = \sum_{i=1}^{I} \beta_i \phi_i(\cdot)$ for a vector $\beta = (\beta_1, \ldots, \beta_I)$ satisfying the Euler equation

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

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

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

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

As we shall show, the previous arguments extend to the general infinite-dimensional case replacing the finite-dimensionality of $\mathcal{R}(A)$ by the compactness of A. A linear operator A is compact if it transforms bounded sets into relatively compact sets (relatively compact sets in $\mathcal{M}$ are those whose closure its compact). The compactness assumption is standard in the literature and is useful for both identification and for obtaining asymptotics of continuous functionals of g. Note, however, that if there are overlapping elements in (C_{t+1}, V_{t+1}) and (C_t, V_t) compactness rules out the case $\mathcal{M} = \mathcal{L}^2$; see Carrasco, Florens and Renault (2007, Example 2.5, p. 22). We could deal with the lack of compactness of A on the whole $\mathcal{L}^2$ by conditioning on (i.e. fixing) the overlapping components, as is common in the literature (see e.g. Blundell, Chen and Kristensen (2007, p. 1629)). From the identification point of view there is little loss of generality by following this "conditioning" approach, however, for deriving asymptotics compactness is convenient, as it guarantees that inference will be based on well-posed generalized inverses (see the discussion at the end of this section).

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

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

THEOREM 1. *Suppose that Assumptions S and C hold. Then, B_0 is a finite set and $\mathcal{G}_0$ is a finite union of finite-dimensional subspaces intersected with $\mathcal{G}$.*

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

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

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

$$\mathcal{P} \equiv \{g \in \mathcal{M} : g \geq 0 \ \ \mu - a.s.\} \tag{6}$$

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

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

Assumption I is a mild condition that extends the classical assumption of a positive matrix in the Perron-Frobenius theorem to an infinite-dimensional setting, see Abramovich and Aliprantis (2002, Chapter 9) and Schaefer (1974). With our shape and normalization restrictions the parameter space is $\mathcal{G} \subseteq \{g \in \mathcal{P} : \|g\| = 1\}$. Note that Assumption S and I imply that $g > 0 \ \mu - a.s.$ for all $g \in \mathcal{G}_0$ because $g = bAg > 0$.

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

Identification can be established under weaker conditions than those of Theorem 2, however, we do not pursue these conditions here because the stronger conditions of Theorem 2 will facilitate our later asymptotic inference results (see earlier working paper versions of our paper for these weaker conditions). Our proof of Theorem 2 also shows that $b = 1/\rho(A)$, where $\rho(A)$ is the spectral radius of A (see the Appendix for a definition of the spectral radius of a linear bounded operator). A key sufficient condition for identification of g is that A is irreducible; see Abramovich and Aliprantis (2002, Chapter 9) for a definition of irreducibility in a general setting.

We could consider other sufficient conditions that replace conditions on A by conditions on a power of A, i.e. we could require that Assumptions C and I hold for A^n, for some $n \geq 1$. It is hard to interpret these conditions, however, in a possibly non-Markovian environment, so we do not pursue them here. 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 μ and $\mathcal{M}$, which lead to simple conditions for identification 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 μ has Lebesgue density f_μ on a common support $S_\mu = S = S'$ (e.g. $S_\mu = [0, \infty)$). Then, taking $\mathcal{M}$ equals to $\mathcal{L}^2$, the operator equation $bAg = g$ can be written as

$$b \int k(c', c) g(c') f_\mu(c') dc' = g(c),$$

where $k(c', c) = r(c', c) f(c', c) / [f_\mu(c') f(c)]$ 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

$$\int \int k^2(c', c) f_\mu(c') f_\mu(c) dc' dc < \infty,$$

see Example 2.3 in Carrasco, Florens and Renault (2007, p. 5659); while Assumption I holds if $k(c', c) > 0$ $\mu \times \mu$ a.s. A viable data-driven choice of f_μ is the pooled density $f_\mu = 0.5f + 0.5f'$, provided it satisfies the squared integrability condition above. Note that these assumptions do not require consumption to be stationary (i.e. $f \neq f'$ is allowed), but they impose restrictions on the set of allowable $\mu' s$ and their supports (e.g., $S_\mu = S = S' = [0, \infty)$) so g can be embedded in a common space (as the correct specification of the Euler equation implies).

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

for inference. For example, in the next sections we obtain rates of convergence for estimation of g that are the same as those of ordinary nonparametric regression.

4 Estimation from Individual level-data

Our estimation strategy follows the identification strategy described above. For estimation we assume that we have a sample of household-level data $\{(R_{t_i+1}, C_{t_i+1,i}, V_{t_i+1,i}, C_{t_i,i}, V_{t_i,i})\}_{i=1}^n$ for n households, with possibly overlapping time periods $t_1 \leq t_2 \leq \cdots \leq t_n$. To simplify notation denote $W_i = (R_i', C_i', V_i', C_i, V_i) \equiv (R_{t_i+1}, C_{t_i+1,i}, V_{t_i+1,i}, C_{t_i,i}, V_{t_i,i})$, where $V_i = (V_{1i}, \ldots, V_{\ell 1 i})$ and $V_i' = (V_{1i}', \ldots, V_{\ell_1 i}')$ with $\ell = \ell_1 + 1$. We assume that the data $\{W_i\}_{i=1}^n$ are drawn from a common distribution with underlying parameter $\theta_0 \equiv (g_0, b_0) \in \Theta$. We allow the observations to be dependent across households possibly reflecting common features in their investment opportunities, but this dependence should be weak enough to permit laws of large numbers and central limit theorems to apply. To be concrete we suppose that there is an ordering of the households such that the series can be considered stationary and mixing, although this ordering need not be known by the econometrician. This type of assumption has been made in finance to the cross section of stock returns. See, e.g., Connor and Korajczyk (1993). We shall henceforth assume that Assumptions S, C and I hold, so that θ_0 is point-identified. Particularly, we consider $g_0 \in \mathcal{G} \subseteq \{g \in \mathcal{P} : \|g\| = 1\}$.

We assume that the vector W_i is continuously distributed (the discrete case is simpler). As in the example above, we denote the Lebesgue density of (C_i, V_i) by f and that of (C_i', V_i') by f'. We consider the setting described in the identification section where μ is a probability measure with Lebesgue density f_μ and support $S_\mu \subseteq S \cap S'$. Henceforth, g and b denote generic elements in $\mathcal{G}$ and $(0, 1)$, respectively.

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

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

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

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

with product kernel

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

Here, K is a univariate kernel function and $h \equiv h_n$ is a possibly stochastic bandwidth. Note that contrary to A, the operator $\widehat{A}$ has a finite-dimensional closed range (that is spanned by the functions

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

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

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

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

which in matrix notation can be simply written as

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

where $\widehat{A}_n$ is an $n\times n$ matrix with ij-th element $a_{ij}=\phi_j(C_i',V_i')R_i'/n$, and $\widehat{\beta}=(\widehat{\beta}_1,\ldots,\widehat{\beta}_n)^\intercal$ (henceforth, $v^\intercal$ denotes the transpose of v). Thus, let $\widehat{\lambda}$ denote the largest eigenvalue in modulus of $\widehat{A}_n$ and $\widehat{\beta}=(\widehat{\beta}_1,\ldots,\widehat{\beta}_n)^\intercal$ its corresponding eigenvector. Our estimators for b_0 and g_0 are, respectively,

$$\hat{b}=1/\widehat{\lambda}\qquad\text{and}\qquad\widehat{g}\,(c,v)=n^{-1}\sum_{i=1}^{n}\widehat{\beta}_i\phi_i(c,v). \tag{7}$$

Marginal utilities are identified up to scale and we consider the normalization $\|\widehat{g}\|=1$, which is implemented by setting $\widehat{\beta}^\intercal\widehat{\Omega}\widehat{\beta}=1$, where $\widehat{\Omega}$ is the $n\times n$ matrix with entries

$$\omega_{ij}=\frac{1}{n^2}\int\phi_i(c,v)\phi_j(c,v)f_\mu(c,v)dcdv.$$

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

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

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). The next section addresses this point.

4.1 Estimation with multiple assets

Suppose that we have J assets, and let $\hat{b}_j$ denote our estimator for the discount factor based on asset j-th, $j = 1, ..., J$. We aim to find weights $w_b^* = (w_{1,b}^*, ..., w_{J,b}^*)^\intercal$ satisfying

$$w_b^* = \arg\min_{w_{j,b}} Avar\left(\sum_{j=1}^{J} w_{j,b}\hat{b}_j\right) \text{ such that } \sum_{j=1}^{J} w_{j,b} = 1,$$

where $Avar$ denotes the asymptotic variance. By our asymptotic results below, the $J \times 1$ vector $\hat{b}^{(J)}$ with components $\hat{b}_j$, $j = 1, ..., J$, has an asymptotic variance proportional to

$$\Sigma_J \equiv \lim_{n\to\infty} \text{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\varepsilon_i\right) < \infty, \tag{8}$$

where s_i is a fixed function defined in the next section and the $J \times 1$ vector $\varepsilon_{i,J}$ has the $j-th$ component $\varepsilon_{i,j} = g_0\left(C_i', V_i'\right) R_{i,j}' - b_0^{-1} g_0\left(C_i, V_i\right)$ for the $j-th$ asset $R_{i,j}'$. Thus, in vector notation the problem above is equivalent to

$$w_b^* = \arg\min_{w_b} w_b^\intercal \Sigma_J w_b \text{ such that } w_b^\intercal \mathbf{1} = 1,$$

where $w_b = (w_{1,b}, ..., w_{J,b})^\intercal$ and $\mathbf{1}$ is a $J \times 1$ vector of ones. By Luenberger (1997, Theorem 2, p. 65) the solution to this optimization problem is

$$w_b^* = \frac{\Sigma_J \mathbf{1}}{\mathbf{1}^\intercal \Sigma_J \mathbf{1}}.$$

Given data, we suggest to estimate the optimal weights w_b^* by the sample analogue

$$\hat{w}_b^* = \frac{\hat{\Sigma}_J \mathbf{1}}{\mathbf{1}^\intercal \hat{\Sigma}_J \mathbf{1}},$$

where $\hat{\Sigma}_J$ is any consistent long run variance estimator of Σ_J. Then, form the estimator

$$\hat{b} = (\hat{w}_b^*)^\intercal \hat{b}^{(J)}.$$

A similar approach can be used for functionals of the marginal utility. We will discuss below that under suitable conditions estimation of the weights w_b^* will not have an impact on the asymptotic

first 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 developed 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 is the discount factor corresponding to asset $R'_{i,j}$ under misspecification), as a test of the linearity constraints $w^\intercal b^{(2)} = 0$, for $w = (1, -1)$. The relevant asymptotic theory to carry out this test is the same as for $(\hat{w}_b^*)^\intercal \hat{b}^{(J)}$ (and somewhat simpler, since weights are not estimated).

An alternative approach that combines moments rather than estimators is as follows. Let $\widehat{A}_{n,j}$ denote the $n \times n$ matrix with ik-th element $a_{ik} = \phi_k(C'_i, V'_i)R'_{i,j}/n$ from the previous section, with $j = 1, ..., J$ indexing the asset. Stack all matrices $(\widehat{A}_{n,j} - \lambda I_n)$ for $j = 1, ..., J$, in a large $(n \times J) \times n$ matrix $B_{n,\lambda}$. A GMM type estimator can be constructed as the minimizer

$$\min_{\beta,\lambda} (\mathbf{1} \otimes \beta)^\intercal B_{n,\lambda}^\intercal \hat{W} B_{n,\lambda}^\intercal (\mathbf{1} \otimes \beta), \tag{9}$$

subject to a normalization constraint, where $\otimes$ denotes Kronecker product and $\hat{W}$ denotes a $(n \times J) \times (n \times J)$ positive definite matrix of GMM weights. Our estimator for one asset corresponds to $J = 1$ and $\hat{W} = I_n$ (just identified case). If $(\mathbf{1} \otimes \beta)$ above is replaced by an (unconstrained) generic $(n \times J) \times 1$ vector then the problem again boils down to our setting with one asset, which is amenable to our asymptotic theory. The general constrained case in (9), however, requires high dimensional numerical optimization. Providing asymptotic theory for this high-dimensional GMM setting is beyond the scope of this paper and is left for future research. For moderate and large sample sizes n, our approach combining estimators offers a feasible compromise between efficiency and computational simplicity.

5 Asymptotic Theory

In this section we provide conditions for the consistency and limiting distribution theory of our estimators as defined in the previous section. We give high-level conditions that allow for general estimators of A, and provide in the Appendix low-level conditions for our leading kernel-type estimator under weakly dependent β-mixing strictly stationary sequences. For a generic subset G of $\mathcal{M}$ define the restricted operator norm

$$||A||_G := \sup_{g \in G \subset \mathcal{M}: ||g|| \leq 1} ||Ag||.$$

Denote by G_0 the eigenspace associated to b_0^{-1} and A.

ASSUMPTION E:

1. *The estimator* $\widehat{g}$ *satisfies* $\|\widehat{g}\| = 1$ *and* $\langle \widehat{g}, 1 \rangle > 0$.

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

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 $\mathcal{L}^2$-consistency of our estimators.

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

We remark that Theorem 3 also holds in the partially identified case where Assumption I is dropped and the $\mathcal{L}^2$-distance between $\widehat{g}$ and g_0 is replaced by the gaps between the eigenspaces of $\widehat{A}$ and A associated to the eigenvalues $\hat{b}^{-1} = \rho(\widehat{A})$ and $b_0^{-1} = \rho(A)$, respectively; see Osborn (1975).

To obtain asymptotic distribution theory for our estimators, we impose the following additional assumptions and notation. Let A^* denote the adjoint operator of A, that is, the linear compact operator such that $\langle Ag_1, g_2 \rangle = \langle g_1, A^* g_2 \rangle$ for all $g_1, g_2 \in \mathcal{M}$. Note that b_0^{-1} is also an eigenvalue for A^*; eigenvalues of A^* are complex conjugates of those of A. Similarly as we did for g_0, it can be shown that under our assumptions there exists a unique (up to scale) strictly positive eigenfunction of A^* associated to b_0^{-1} (see Theorem 7.C in Zeidler (1986, vol. 1, p. 290)).

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

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

$$\varepsilon_i = g_0\left(C_i', V_i'\right) R_i' - b_0^{-1} g_0\left(C_i, V_i\right), \qquad i = 1, \ldots, n. \tag{10}$$

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

ASSUMPTION N.

1. $||\widehat{A} - A||_{\mathcal{G}} = o_P(n^{-1/4})$ *and* $P\left(\widehat{g} \in \mathcal{G}\right) \to 1$ *as* $n \to \infty$.

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

3. *Furthermore,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\varepsilon_i \xrightarrow{d} N\left(0,\Sigma_s\right),$$

where $\Sigma_s \equiv \lim_{n\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i\varepsilon_i\right) < \infty$.

THEOREM 4. *Let Assumptions S, C, I, E and N hold. Then, as* $n \to \infty$,

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

The proof of Theorem 4 can be found in the Appendix. We can estimate the asymptotic variance of $\widehat{b}$ by standard long run variance estimators based on $\{\widehat{s}_i\widehat{\varepsilon}_i\}_{i=1}^n$, see e.g. Newey and West (1987), where $\widehat{\varepsilon}_i = \widehat{g}\left(C_i', V_i'\right) R_i' - \widehat{b}^{-1}\widehat{g}\left(C_i, V_i\right)$, and $\widehat{s}$ is computed as our estimator $\widehat{g}$, with the normalization $\langle\widehat{g},\widehat{s}\rangle_n = 1$. 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

$$\begin{aligned}\sqrt{n}\left((\hat{w}_b^*)^\intercal \hat{b}^{(J)} - b_0\right) &= (\hat{w}_b^*)^\intercal \sqrt{n}\left(\hat{b}^{(J)} - b_0\mathbf{1}\right)\\ &\quad + \sqrt{n}\left(\hat{w}_b^* - w_b^*\right)^\intercal b_0\mathbf{1}.\end{aligned}$$

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,

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

where Σ_J is defined in (8).

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

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

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

Recall $L^\dagger$ denotes the Moore-Penrose pseudoinverse of $L = b_0 A - I$, which under our conditions is linear and continuous (cf. Section 3.1).

THEOREM 5. *Let Assumptions S, C, I, E and N hold. Then, we have*

$$(\widehat{g} - g_0)(c, v) = b_0 L^{\dagger} \Delta_n(c, v) + R_n(c, v),$$

where as $n \to \infty$, $\|R_n\| = o_P(n^{-\rho/2})$ *for any* $0 < \rho < 1$.

This result implies that the rates of convergence of $\widehat{g} - g_0$ in $\mathcal{L}^2$ are the same as those of the NW kernel estimator of $E[\varepsilon_i | C_i = c, V_i = v]$. We could use the expansion of Theorem 5 to test parametric hypotheses about g, i.e., $H_0 : g_0(c, v) = g_{\eta_0}(c, v)$, against nonparametric alternatives, where the function $g_{\eta_0}(c, v)$ is known up to a finite-dimensional unknown parameter η_0 (e.g. power utility). A test can be based on the discrepancy

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

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

6 Summary Measures

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

Define the $ARRA$ functional by

$$\gamma(g) \equiv E\left[\frac{-C \partial g(C, V)/\partial c}{g(C, V)}\right]. \tag{11}$$

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

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

Under the assumptions for Theorem 6 below, $\widehat{g}$ is differentiable and bounded away from zero with probability tending to one, so $\gamma_n(\widehat{g})$ 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

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

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

$$\chi_s = \chi - \langle g_0, \chi \rangle \langle g_0, s \rangle^{-1} s. \tag{14}$$

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

We also introduce a class of smooth function $\mathcal{C}^{\eta}(T)$ for a generic closed and convex set T. For any vector a of ℓ integers define the differential operator $\partial_x^a \equiv \partial^{|a|_1}/\partial x_1^{a_1} \ldots \partial x_\ell^{a_\ell}$, where $|a|_1 \equiv \sum_{i=1}^{\ell} a_i$. For any smooth function $h : T \subset \mathbb{R}^\ell \to \mathbb{R}$ and some $\eta > 0$, let $\underline{\eta}$ be the largest integer smaller than or equal to η, and

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

Further, let $\mathcal{C}_M^{\eta}(T)$ be the set of all continuous functions $h : T \subset \mathbb{R}^\ell \to \mathbb{R}$ with $\|h\|_{\infty,\eta} \leq M$ (for an integer η, the η-th derivative is assumed to be continuous). Since the constant M is irrelevant for our results, we drop the dependence on M and denote $\mathcal{C}^{\eta}(T)$.

The *ARRA* estimator behaves asymptotically as a sample average, with an influence function given by

$$\xi_i = (\zeta_i - E[\zeta_i]) - b_0 \chi_s^*(C_i, V_i)\varepsilon_i, \tag{15}$$

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

Assumption CE.

1. *The class $\mathcal{D}$ defined in (12) is P-Donsker*[5].

2. *The measure μ is the probability measure of (C, V) with a support S that can be written as $S = [l_c, u_c] \times S_V$, for some l_c, u_c with $l_c < u_c$. Furthermore, $\lim_{c \to l_c} cf(c, v) = 0 = \lim_{c \to u_c} cf(c, v)$ for all $v \in S_V$ and $P(\min\{g_0, \widehat{g}\} > \varepsilon) \to 1$ for some $\varepsilon > 0$.*

3. *The function d in (13) satisfies $d \in \mathcal{L}^2$, $\{\xi_i\}$ in (15) satisfies*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \xrightarrow{d} N(0, \Xi),$$

where $\Xi \equiv \lim_{n \to \infty} \text{var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i\right) < \infty$ and $\chi_s^* \in \mathcal{C}^r(S)$.

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 $\gamma_n(\widehat{g})$ is well defined. Assumption CE.3 implies that the asymptotic variance of $\gamma_n(\widehat{g})$ is finite.

THEOREM 6. *Let Assumptions S, C, I, E, N and CE hold. Then,*

$$\sqrt{n}(\gamma_n(\widehat{g}) - \gamma(g_0)) \xrightarrow{d} N(0, \Xi),$$

where ξ_i is defined in (15).

Estimating the asymptotic variance of $\gamma_n(\widehat{g})$ by plug-in methods would be possible but complicated. An alternative is to use the block bootstrap, which can be justified along the lines of Radulović (1996) and Chen, Linton and Van Keilegom (2003).

Now consider some other functionals of interest. The asymptotic normality of each can be established using the same methods as Theorem 6. As with $\gamma_n(\widehat{g})$, we can use the bootstrap to estimate their limiting distributions. For the remainder of this section we drop the i subscript for clarity. Closely related to the *ARRA* are local averages defined by

$$\rho(q, j) = E\left[\frac{-C_{t+1} \partial g_0(C_{t+1}, C_t)/\partial C_{t+1}}{g_0(C_{t+1}, C_t)} | C_{t+1} \in Q_q, C_t \in S_j\right], \tag{16}$$

[5]Let P_n be the empirical measure with respect to P. Using a standard empirical process notation, define $\mathbb{G}_n = \sqrt{n}(P_n - P)$. Then $\mathcal{D}$ is P-Donsker if $\mathbb{G}_n$ converges weakly to $\mathbb{G}$ in the space of uniformly bounded functions on $\mathcal{D}$, $l^\infty(\mathcal{D})$, where $\mathbb{G}$ is a mean-zero Gaussian process with uniformly continuous sample paths, see Doukhan, Massart and Rio (1995) for further details.

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 $\rho(q, j)$ can be shown to be asymptotically normal under the same conditions above used for the $ARRA$. That is, with the simplified notation $\rho(q) \equiv \rho(q, q)$ for the parameter and $\rho_n(q) \equiv \rho_n(q, q)$ for the plug-in estimator, it can be shown

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

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

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

The t-statistics are constructed as

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

for suitable consistent estimates $\sigma_n^2(q)$ of the asymptotic variances $\sigma^2(q)$, for $q = 1, 2, 3$ and 4. We then reject H_{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

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

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

7 Monte Carlo Experiment

In this section we illustrate the finite-sample performance of our estimator described in the previous sections based on a $CRRA$ utility function so that $g_0(c, v) = c^{-\eta_0}$, where η_0 in this case equals the

ARRA. The model is then given by the Euler equation

$$b_0 E\left[C_{t+1}^{-\eta_0} R_{t+1} | C_t\right] = C_t^{-\eta_0}.$$

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

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

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

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

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

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

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

where $g^*(C_{t+1}, V_{t+1}) \equiv C_{t+1} g(C_{t+1}, V_{t+1})$, $g^*(C_t, V_t) \equiv C_t g(C_t, V_t)$ and $R_{t+1}^* \equiv (C_t/C_{t+1}) R_{t+1}$. With these definitions the procedure remains as described in Section 4 after redefining the return variable, from R_{t+1} to R_{t+1}^*. The procedure then yields an estimate of g^*, from which the marginal utility

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

For each finite-dimensional parameter and summary measure we consider, we report the mean, standard deviation, 2.5th percentile, 97.5th percentile, 95% coverage probability based on normal distribution, their bootstrap counterparts and the root mean square error.[6] Table 1 reports estimates of the discount factor from our three estimators, $CRRA$, $NP-1$, and $NP-2$. Table 2 reports estimates of the $ARRA$, which for the $CRRA$ model is just the estimated constant η_0, while for the nonparametric estimators the $ARRA$ is $\gamma(g_0)$ defined by equation (11). Table 1 shows that all of the estimators succeed in estimating the discount factor b very accurately. This is in contrast to many macro models, which often calibrate the discount factor due to the difficulty in estimating it accurately. Table 2 shows somewhat more difficulty in estimating the $ARRA$, but the relative accuracy of our nonparametric estimates to the parametric alternative is similar. In both tables the root mean squared errors of our nonparametric estimates are seen to shrink with sample size and increase with dimensionality at rates that are generally consistent with asymptotic theory.

Figures 1 and 2 show plots of the one-dimensional nonparametric (i.e., $NP-1$) estimated marginal utility function g_0 as a function of C_t. Figure 1 is $n = 500$ while Figure 2 is $n = 2000$. For each figure, the solid line denotes the mean, the dotted line denotes the 95% confidence interval, and the dashed line is the true curve. One can see from these figures that $NP-1$ quite accurately tracks the true function. The precision of these fits can also be summarized by their integrated mean square error (weighted with respect to the true density), which is 0.0014 for $n = 500$ and 0.0005 for $n = 2000$.

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

One unexpected finding is that estimates of $\rho(q,j)$ display substantially larger biases and root mean squared errors for larger values of q and j than for smaller values, suggesting that our $NP-2$

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

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

estimates of the marginal utility function tend to be less accurate at higher consumption levels. This can also be seen for $NP-1$ in Figure 1, where the standard error bands widen at higher consumption levels.

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

	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 value of b_0 is 0.95. $CRRA$, $NP-1$ and $NP-2$ refer respectively to the parametric, one-dimensional nonparametric, and two-dimensional nonparametric estimators.

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

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

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

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

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

Table 4: Summary statistics of Monte Carlo estimates of $\delta(g_0)$, used to test for the presence of habit effects. The true value of each $\delta(g_0)$ is 0. The $\tau(C_{t+1}, C_t)$ column lists the functions that are used to define $\delta(g_0)$.

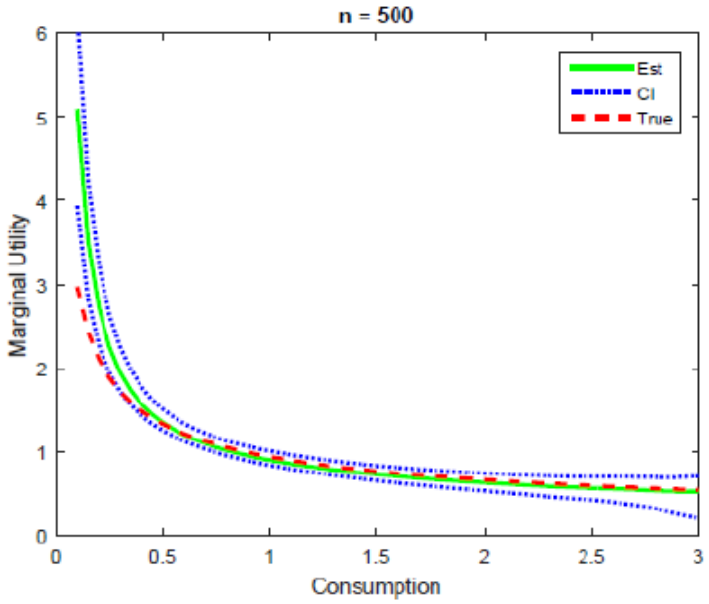

Figure 1: Estimates of the marginal utility function g_0 using simulated data with $n = 500$. Est, CI, and $True$ represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true curve.

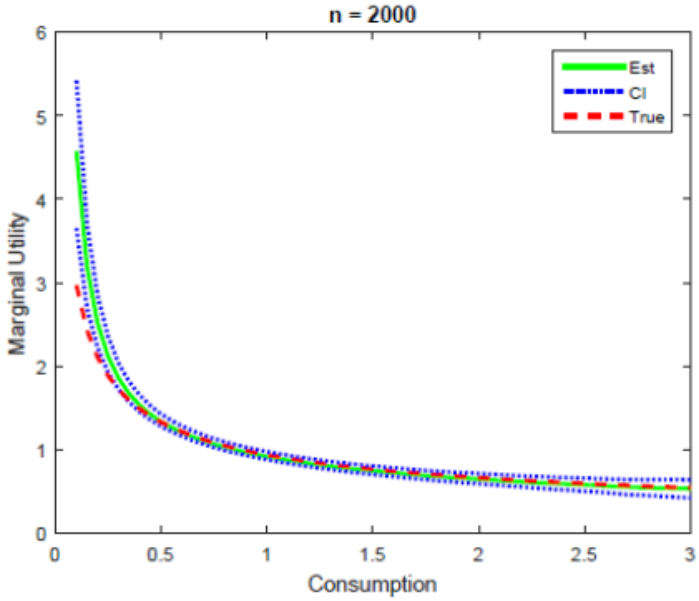

Figure 2: Estimates of the marginal utility function g_0 using simulated data with $n = 2000$. Est, CI, and $True$ represent respectively the one-dimensional nonparametric estimator, its 95% confidence interval, and the true curve.

8 Conclusions

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

In the older version of our paper, we apply our nonparametric methods to household-level CEX data and find evidence against the common assumption of constant relative risk aversion across consumers. Our estimates are fairly insensitive to the choice of asset used (risk-free vs risky), which supports our nonparametric model. We find empirical evidence for the presence of habits, and evidence that risk aversion varies across current and lagged consumption levels in ways that are not fully captured by standard parametric or even semiparametric specifications of habits in asset pricing models. However, there are two caveats that underlie our findings. One is we used common asset returns with repeated cross-sections of household consumption since we do not have data on household-level assets. Our existing theory requires suitable variation in the returns and extending it to clustering data is beyond the scope of this paper. Another concern with the dataset we used is the presence of measurement error in consumption, see e.g. Alan, Attanasio and Browning (2009) and references therein. Escanciano (2019) has recently shown that the identification and estimation of the discount factor in this paper is robust to the presence of measurement error under rather general conditions. Measures of risk aversion, however, are more sensitive to measurement error and may require different estimation strategies to achieve robustness. These strategies are well beyond the scope of this paper and will be investigated in future research.

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). *An Invitation to Operator Theory*. 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," *Journal of Applied Econometrics*, 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," *Economics Letters*, 62, 1-6.

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

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

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

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

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

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

[13] Carrasco, M. and J. P. Florens (2000), "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 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," *Handbook of Econometrics*, 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," *The Journal of Finance*, 52, 1383–1410.

[17] Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2014), "Identification in Semiparametric and Nonparametric Conditional Moment Models," *Econometrica*, 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," *Annals of Statistics*, 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," *Journal of Applied Econometrics*, 24, 1057-1093.

[22] Chen, X. and D. Pouzo (2009), "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 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", *Econometrica*, 85, 1501-1536.

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

[27] Connor, G. and R. A. Korajczyk (1993), "A Test for the Number of Factors in an Approximate Factor Model" *The Journal of Finance*, 48, 1263-1291.

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

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

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

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

[32] Doukhan, P., (1994), Mixing. Properties and examples, Springer, Lecture Notes in Statistics.

[33] Doukhan, P., Massart, P. and Rio, E. (1995), "Invariance principles for absolutely regular empirical processes," *Annales de l'I.H.P. Probabilités et statistiques*, 31, 2, 393-427.

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

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

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

[37] Escanciano, J.C. (2019), "Semiparametric Identification and Fisher Information," https://arxiv.org/abs/1609.06421.

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

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

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

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

[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," *Annals of Statistics*, 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" *Annals of Statistics*, 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," *Bernoulli*, 15, 380-401.

[47] Hansen, B. (2008), "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory,* 24, 726–748.

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

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

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

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

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

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

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

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

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

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

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

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

[60] Lewbel, A. (1994), "Aggregation and Simple Dynamics," *American Economic Review,* 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," *Journal of Monetary Economics*, 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," *Mathematics of Computation*, 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," *Journal of Finance*, 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). *Banach Lattices and Positive Operators*, 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," *Journal of Business and Economic Statistics*, 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). *Weak Convergence and Empirical Processes with Applications to Statistics*, 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). *Nonlinear Functional Analysis and its Applications-Fixed Point Theorems.* Springer-Verlag, New York, Inc.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$g(C_t, V_t) = bE\left[g(C_{t+1}, V_{t+1}) R_{jt+1} \mid C_t, V_t\right] \qquad j = 1, \ldots, J, \tag{18}$$

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

9.2 Results on kernel estimators

9.2.1 Assumptions

This section collects some results on kernel estimates, providing primitive assumptions for the general conditions of the main text. Let $Y_i \in \mathbb{R}^{d_Y}$ denote the elements of (C_i', V_i') that do not overlap with (C_i, V_i), let $X_i \in \mathbb{R}^{d_X}$ denote the overlapping elements of (C_i', V_i') and (C_i, V_i) and let $Z_i \in \mathbb{R}^{d_Z}$ denote the elements of (C_i, V_i) that do not overlap with (C_i', V_i'). Denote $\xi_i = (Y_i, X_i, Z_i)$, for $i \in \mathbb{Z}$. Define the class of functions

$$\mathcal{F} = \{\xi_i \to g(C_i', V_i') R_i' : g \in \mathcal{G}\}.$$

To measure the complexity of the class $\mathcal{F}$, or any other class, we can employ covering or bracketing numbers. Given two functions l, u, a bracket $[l, u]$ is the set of functions $f \in \mathcal{F}$ such that $l \leq f \leq u$. An ε-bracket with respect to $\|\cdot\|$ is a bracket $[l, u]$ with $\|l - u\| \leq \varepsilon$, $\|l\| < \infty$ and $\|u\| < \infty$ (note that u and l not need to be in $\mathcal{F}$). The *covering number with bracketing* $N_{[\cdot]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of ε-brackets with respect to $\|\cdot\|$ needed to cover $\mathcal{F}$. Let $N(\varepsilon, \mathcal{G}, \|\cdot\|)$ be the *covering number with respect to* $\|\cdot\|$, i.e. the minimal number of ε-balls with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. An envelope for $\mathcal{G}$ is a function G, such that $G(c, v) \geq \sup_{g \in \mathcal{G}} |g(c, v)|$ for all (c, v).

Denote by $\mathcal{K}(r)$ the class of bounded functions $k(t) : \mathbb{R} \to \mathbb{R}$ such that for some $r \geq 2$: $\int u^l k(u)\, du = \delta_{l0}$ for $l = 0, \ldots, r-1$, where $\delta_{ll'}$ denotes Kronecker's delta, and $\int |u^r k(u)|\, du < \infty$. Furthermore, for some $\Lambda_1 < \infty$ and $L < \infty$, either $k(u) = 0$ for $|u| > L$ and k is Lipschitz with constant Λ_1 or k is differentiable $|\partial k(t)/\partial t| \leq \Lambda_1$ and for some $v > 1$, $|\partial k(t)/\partial t| \leq \Lambda_1 |t|^{-v}$ for $|t| > L$. These assumptions are extensively discussed in Hansen (2008).

The following regularity conditions are needed for the subsequent asymptotic analysis. Let $\mathcal{F}_s^t \equiv \mathcal{F}_s^t(\xi_i)$ denote the σ-algebra generated by $\{\xi_j,\ j = s,\ldots,t\}$, $s \leq t$, $s,t \in \mathbb{Z}$. Define the β-mixing coefficients as (see, e.g., Doukhan (1994))

$$\beta_t = \sup_{m\in\mathbb{Z}} \sup_{A\in\mathcal{F}_{t+m}^{\infty}} E\left|P(A|\mathcal{F}_{-\infty}^{m}) - P(A)\right|.$$

Assumption A0:

1. $\{\xi_i\}_{i\in\mathbb{Z}}$ *is a strictly stationary and absolutely regular (β-mixing), with mixing coefficients of order $O(t^{-b})$, for some b such that $b > \delta/(\delta-2)$, where $2 < \delta < \infty$, and δ is as in A1.1 below.*

Assumption A1:

1. *For each $\varepsilon > 0$, $\log N_{[\cdot]}(\varepsilon, \mathcal{F}, \|\cdot\|) \leq C\varepsilon^{-v}$ for some $v < 2 - 2\delta/b(\delta-1)$. The class $\mathcal{G}$ is such that $g_0 \in \mathcal{G}$ and has an envelope G such that $\sup_{(c,v)\in S_\mu} E[|G(C_i', V_i')R_i'|^{\delta}\,|C_i = c, V_i = v] < C$ for some $\delta > 2$. Moreover,*

$$\lim_{\delta\to 0} \sup_{|(c_1,v_1)-(c_2,v_2)|<\delta} \sup_{g\in\mathcal{G}} \|\mathbb{E}[g(C',V')R'|C = c_1, V = v_1] - \mathbb{E}[g(C',V')R'|C = c_2, V = v_2]\| = 0.$$

2. *The density function $f(\cdot)$ is bounded away from zero on S_μ and is continuous on S.*

3. *The kernel satisfies $K \in \mathcal{K}(2)$.*

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

Examples of classes $\mathcal{F}$ satisfying A1.1 abound in the literature; see van der Vaart and Wellner (1996). The remaining conditions in Assumption A1 are self-explanatory. For A1.3 we could also use kernels with unbounded support that satisfy some smoothness and integrability conditions. Finally, note that A1.4 allows for data-driven bandwidth choices, which are common in applied work.

For asymptotic normality of our estimators we require the following assumption.

Assumption A2.

1. *The density function f satisfies $f \in \mathcal{C}^r(S_\mu)$, where r as in A2.4 below.*

2. *$Ag \in \mathcal{C}^r(S_\mu)$ for all $g \in \mathcal{G}$.*

3. *The function s given in Definition 3 above satisfies $s \in \mathcal{C}^r(S_\mu)$.*

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

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

9.2.2 Some generic results

We denote by $\psi \equiv (\varphi, c, v)$ a generic element of the set $\Psi \equiv \mathcal{F} \times S_\mu$. Let $f(c,v)$ denote the density of (C_i, V_i) evaluated at (c,v). Define the regression function $m(\psi) \equiv E[\varphi(C_i', V_i')R_i'|C_i = c, V_i = v]$, which does not depend on i. Then, an estimator for $m(\psi)$ is given by

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

Henceforth, we abstract from measurability issues that may arise (see van der Vaart and Wellner (1996) for ways to deal with lack of measurability). The following lemma is used in subsequent results.

Lemma B1. *Suppose that Assumptions A0-A1 hold. Then,*

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} |\widehat{m}_h(\psi) - m(\psi)| = o_P(1). \tag{19}$$

If, in addition, A2 holds, then

$$\sup_{l_n \le h \le u_n} \sup_{\psi \in \Psi} |\widehat{m}_h(\psi) - m(\psi)| = O_P\left(\sqrt{\frac{1}{nl_n^\ell}} + u_n^r\right). \tag{20}$$

Proof. By the triangle inequality

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

where $T(\psi) \equiv m(\psi) f(c,v)$. We obtain uniform rates for $\widehat{T}_h(\psi) - E[\widehat{T}_h(\psi)]$; the rates for $\widehat{f}(c,v) - E[\widehat{f}(c,v)]$ follow analogously and are simpler to obtain.

Define the class of functions

$$\mathcal{K}_0 := \left\{(C_i, V_i) \to K\left(\frac{c - C_i}{h}\right) \prod_{j=1}^{\ell_1} K\left(\frac{v_j - V_{ji}}{h}\right) : (c,v) \in S_\mu, h \in (0,1]\right\}.$$

By the proof of Lemma B.3 in Escanciano, Jacho-Chávez and Lewbel (2014) $\mathcal{K}_0$ is a VC class and hence $N_{[\cdot]}(\epsilon, \mathcal{K}_0, \|\cdot\|_2) \leq C\varepsilon^{-\alpha_K}$ for some $\alpha_K \geq 1$. On the other hand, Lemma A.1 of the same reference yields

$$\log N_{[\cdot]}(\epsilon, \mathcal{F}\cdot\mathcal{K}_0, \|\cdot\|_2) \leq \log N_{[\cdot]}(C\epsilon, \mathcal{F}, \|\cdot\|_2) + \log N_{[\cdot]}(C\epsilon, \mathcal{K}_0, \|\cdot\|_2).$$

By Assumption A1.1 this is bounded by $C\varepsilon^{-v}$.

Theorem 3 and (2.15) in Doukhan, Massart and Rio (1995) applied to the class $\mathcal{F}\cdot\mathcal{K}_0$ then imply

$$\sup_{l_n \leq h \leq u_n} \sup_{\psi \in \Psi} \left| \widehat{T}_h(\psi) - E\left[\widehat{T}_h(\psi)\right] \right| = O_P\left(\sqrt{\frac{1}{nl_n^\ell}}\right),$$

provided $\|f\|_{2,\beta}^2 \leq h^\ell$ for all $f \in \mathcal{F}\cdot\mathcal{K}_0$, where for any function f,

$$\|f\|_{2,\beta}^2 = \int_0^1 \beta^{-1}(u) Q_f^2(u) du,$$

and where β^{-1} is the inverse cadlag of the decreasing function $u \to \beta_{\lfloor u \rfloor}$ ($\lfloor u \rfloor$ being the integer part of u, and β_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)

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

Hence,

$$\|f\|_{2,\beta}^2 \leq \int_0^1 \beta^{-1}(u) \frac{Ch^\ell}{u} du \leq Ch^\ell \int_0^1 u^{b-1} du = \frac{Ch^\ell}{b},$$

where the latter inequality follows from Assumption A0.

On the other hand, Lemma 2 in Einmahl and Mason (2005) and the uniform equicontinuity of Assumption A1.1 yield

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

and likewise for the density bias term. This together with the above expansion for $\widehat{m}_h - m$ completes the proof of (19).

To obtain rates for the bias terms we need the smoothness conditions of Assumption A2. A standard Taylor expansion argument, the higher-order property of the kernel and the Lipschitz property of the $r-th$ derivative imply that

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

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

Lemma B2. *Suppose that Assumptions A0-A1 hold. Then, as* $n \to \infty$,

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

If, in addition, A2 holds, then

$$\sup_{l_n \leq h \leq u_n} \left\|\widehat{A} - A\right\|_{\mathcal{F}} = O_P\left(\sqrt{\frac{1}{nl_n^{\ell}}} + u_n^r\right).$$

Proof. The result follows directly from Lemma B1. ■

We introduce a useful class of functions:

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

Lemma B3. *Suppose that Assumptions A0, A1 and A2 hold. Then, for any* $\varphi \in \mathcal{L}^2(r)$, *it holds that*

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

and then

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

Proof. Define

$$\widehat{T} g_0(c, v) = \frac{1}{n} \sum_{i=1}^{n} g_{0i}' R_i' K_{hi}(c, v),$$

with $g_{0i}' \equiv g_0(C_i', V_i')$ and note that $\widehat{A} g_0(c, v) = \widehat{T} g_0(c, v) / \widehat{f}(c, v)$. Using standard arguments, we write

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

where

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

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

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

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

$$\int \varphi(c,v)[\widehat{T}g_0(c,v) - Tg_0(c,v)]dcdv \tag{21}$$

$$- \int \varphi(c,v) A g_0(c,v) [\widehat{f}(c,v) - f(c,v)]dcdv + o_P(n^{-1/2}). \tag{22}$$

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 $E[\widehat{T}g_0(c,v)]$ is $O_P(u_n^r) = o_P(n^{-1/2})$ by the condition $nu_n^{2r} \to 0$. Hence,

$$\int \varphi(c,v)[\widehat{T}g_0(c,v) - Tg_0(c,v)]dcdv = \int \varphi(c,v)[\widehat{T}g_0(c,v) - E(\widehat{T}g_0(c,v))]dcdv + o_P(n^{-1/2})$$

$$= \frac{1}{n}\sum_{i=1}^{n} g_{0i}' R_i' \int \varphi(c,v) K_{hi}(c,v)\, dcdv - \int \varphi(c,v) E(g_0' R_i' K_{hi}(c,v)) dcdv + o_P(n^{-1/2}),$$

$$= \frac{1}{n}\sum_{i=1}^{n} \varphi(C_i,V_i) g_{0i}' R_i' - E[\varphi(C_i,V_i) A g_0(C_i,V_i)] + o_P(n^{-1/2}),$$

where the last equality follows from the standard change of variables argument and our Assumption A2. Likewise, the term (22) becomes $n^{-1/2}\sum_{i=1}^n \varphi(C_i,V_i) A g_0(C_i,V_i) - E[\varphi(C_i,V_i) A g_0(C_i,V_i)] + o_P(n^{-1/2})$. In conclusion, we have

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

Then, the result follows from a standard central limit theorem for β-mixing sequences. ■

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

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

Also for $r \in \mathcal{N}^\perp(L) = \mathcal{R}(L^*)$ denote by r^* the unique minimum norm solution of $r = L^* r^*$. Note that for $r \in \mathcal{R}(L^*)$, r_s^* does not depend on the solution r^* considered of $r = L^* r^*$ (whether or not this is minimum norm). This follows because under our conditions $\mathcal{N}(L^*)$ is the linear span generated by s. That is, all solutions φ of $r = L^*\varphi$ are of the form $\varphi = r^* + \lambda s$, with $\lambda \in \mathbb{R}$, and for such solution

$$\begin{aligned} \varphi - \langle g_0, \varphi\rangle \langle g_0, s\rangle^{-1} s &= r^* + \lambda s - \langle g_0, r^*\rangle \langle g_0, s\rangle^{-1} s - \lambda \langle g_0, s\rangle \langle g_0, s\rangle^{-1} s \\ &= r^* - \langle g_0, r^*\rangle \langle g_0, s\rangle^{-1} s. \end{aligned}$$

Lemma B4. *Let Assumptions S, C, I, E, N and A0-A2 hold. If $\varphi \in \mathcal{N}^\perp(L)$, so $\varphi = L^*\varphi^*$ for some φ^*, and if $\varphi_s^* \in \mathcal{L}^2(r)$, then*

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

Proof. Note that by (25) below and the adjoint property

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

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

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

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

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

$$\left|\widehat{b}^{-1}-b_0^{-1}\right| \leq M||\widehat{A}-A||_{G_0} \tag{23}$$

and

$$\|\widehat{g}-\widetilde{g}\| \leq M||\widehat{A}-A||_{G_0}, \tag{24}$$

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

$$\left| \widehat{b} - b_0 \right| \leq M \left| \widehat{b} \times b_0 \right| ||\widehat{A} - A||_{G_0} \leq M ||\widehat{A} - A||_{G_0},$$

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

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

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

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

Next,

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

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

Proof of Theorem 4. By definition

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

Write the left hand side of the last display as

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

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

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

we obtain

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

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

$$\left\| \widehat{R} \right\| \leq C \left\{ ||\widehat{A} - A||_{G_0}^2 + ||\widehat{A} - A||_{\mathcal{G} - \{g_0\}} \|\widehat{g} - g_0\| \right\}$$

and

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

which implies by Assumption N.1

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

Then, Cauchy-Schwarz inequality yields

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

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

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

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

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

$$0 = \widehat{L}\widehat{g} - Lg_0 = L(\widehat{g} - g_0) + (\widehat{L} - L)g_0 + (\widehat{L} - L)(\widehat{g} - g_0). \tag{25}$$

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

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

and

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

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

$$L(\widehat{g} - g_0) = -b_0(\widehat{A} - A)g_0 + R_n,$$

where R_n satisfies the conditions of the Theorem. ■

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

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

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

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

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

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

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

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

and therefore

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

The result then follows from Assumption CE.3. ■