Nonparametric Euler Equation Identification and Estimation

Arthur Lewbel* Oliver Linton† Sorawoot Srisuma‡
Boston College London School of Economics London School of Economics

First Draft, June, 2005
This Draft February 2011

Abstract

We consider nonparametric identification and estimation of consumption based asset pricing Euler equations. This entails estimation of pricing kernels or equivalently marginal utility functions up to scale. The standard way of writing these Euler pricing equations yields Fredholm integral equations of the first kind, resulting in the ill posed inverse problem. We show that these equations can be written in a form that resembles Fredholm integral equations of the second kind, having well posed rather than ill posed inverses. We allow durables, habits, or both to affect utility. We extend the usual method of solving Fredholm equations to allow for the presence of habits. Using these results, we show with few low level assumptions that marginal utility functions and pricing kernels are locally nonparametrically identified, and we give conditions for finite set and point identification of these functions. We provide consistent nonparametric estimators for these functions and associated limiting distributions, and an empirical application to US consumption and asset pricing data.

We would like to thank Xiaohong Chen and Don Andrews for helpful comments. All errors are our own. JEL Codes: C14, D91,E21,G12. Keywords: Euler equations, marginal utility, pricing kernel, Fredholm equations, integral equations, nonparametric identification, asset pricing.

*Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Phone: (617) 552-3678. https://www2.bc.edu/arthur~lewbel/ E-mail: lewbel@bc.edu.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. http://econ.lse.ac.uk/staff/olinton/~index_own.html. E-mail address: o.linton@lse.ac.uk.

‡Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. http://personal.lse.ac.uk/srisuma/ E-mail address: s.t.srisuma@lse.ac.uk

1 Introduction

One of original motivations of the generalized method of moments GMM estimator was estimation of the Euler equations associated with rational expectations in consumption and associated consumption based asset pricing models. See, e.g., Hansen and Singleton (1982). More recently, these types of Euler equation models have been used as leading examples of nonparametric instrumental variables estimators. See, e.g., Newey and Powell (2003). Nonparametric instrumental variables models have the structure of Fredholm equations of the first kind, and hence suffer from the ill-posed inverse problem, resulting in nonstandard limiting distribution theory and requiring difficult to interpret, high level assumptions for identification.

In this paper we show that, even with the inclusion of habits, the standard time additive specification of utility in these Euler equation models permits writing these Euler equations in a form that resembles (and in some cases equals) Fredholm equations of the second kind, which (unlike those of the first kind), have well-posed inverses.

Our models permit the marginal utility of consumption in each time period to depend on durables or habits, which have been shown to be empirically important. In a utility function that allows for habits, marginal utility in each time period depends on both current and past consumption levels. This dependence on past consumption generates an Euler integral equation that is more complicated than the standard Fredholm equation of the second kind. However, we show that this more complicated integral equation can be solved in a way that is analogous to the usual solution method for Fredholm equations of the second kind.

Our results make nonparametric Euler equation estimation, and associated asset pricing models, amenable to standard nonparametric estimation methods and associated limiting distribution theory, rather than requiring the nonstandard limiting distribution theory and associated peculiar regularity requirements required for dealing with estimation of ill-posed inverses.

Existing nonparametric Euler equation estimators in the literature assume identification. The conditions for nonparametric identification of Fredholm integral equations of the first kind are high level and difficult to interpret. By recasting the model as an integral equation that equals (or in the model with habits, resembles) a Fredholm equation of the second kind, we can provide more explicit, low level conditions for identification.

In particular, we show that the subjective rate of time preference b equal the inverse of an eigenvalue of an identified matrix, and that the marginal utility function and hence the asset pricing kernel can be recovered from the associated eigenvector. This implies that, with minimal low level assumptions both b and the marginal utility function are identified up to a finite set (and hence is locally identified) and may be globally point identified. Given point or finite set identification of

b, the marginal utility function and pricing kernel are also point or finite set identified as long as there are not multiple roots and hence multiple eigenvectors associated with the eigenvectors in the identified set for b.

These identification results apply to estimation of a single Euler equation. If we observe returns data on multiple assets, and so have multiple Euler equations, then the pricing kernel, the marginal utility function, and b will all in general be point identified, unless by some great coincidence the Fredholm kernels associated with each asset just happen to yield multiple identical eigenvalues and associated marginal utility functions across all assets.

One use of nonparametric estimates is to test for different functional restrictions on utility and hence on the pricing kernel. Let C_t be expenditures on consumption and let V_t be a vector of other variables that may affect utility. These other variables could include durables or lagged consumption values through habit formation. One possible application is to nonparametrically estimate the model including both durables and lagged consumption in V_t, and then test whether one or both belong in the model.

Let $g(C_t, V_t)$ denote a time t marginal utility function (details regarding the construction of g are provided in the next section). Let R_{jt} be the gross return in time period t of owning one unit of asset j in period $t-1$. For a consumer with time separable utility and a rate of time preference b that saves by owning assets j, the Euler equation for maximizing utility is usually represented as

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

where M_t is the time t pricing kernel. An Euler equation in the form of equation (1) holds for each financial asset j, and so pricing equations are often cast in the form of relative returns or net returns

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

where R_{0t+1} denotes either a market rate or a risk free rate.

The goal is estimation of the pricing kernel M_t, or equivalently the marginal utility function g up to scale and the subjective discount factor b. The scale of g is not identified, so an arbitrary scale normalization (which does not affect the resulting estimate of M_t) is imposed.

Estimation typically proceeds by applying Hansen's (1982) Generalized Method of Moments (GMM) to the moment conditions (2) after parameterizing the unknown marginal utility function. Prominent examples of such parameterized models include Hall (1978), Hansen and Singleton (1982), Dunn and Singleton (1986), and Campbell and Cochrane (1999), among many, many others.

This paper considers semiparametric and nonparametric identification and estimation of this model. A difficulty with nonparametric estimation of equation (1) or (2) is that solving for the pricing kernel M_t corresponds to solving a Fredholm integral equation of the first kind (writing out the conditional expectation as an integral). Estimation of equation (1) or (2) with unknown functions is a special case of nonparametric instrumental variables estimation, or conditional GMM estimation with unknown functions, as in Carrasco and Florens (2000), Newey and Powell (2003), Ai and Chen (2003), Hall and Horowitz (2005), Chen and Ludvigson (2009), Chen and Pouzo (2009), Chen and Reiss (2010), and Darolles, Fan, Florens and Renault (2010). Identification conditions for nonparametric instrumental variable models are given by Chen, Chernozhukov, Lee, and Newey (2010).

These estimators require verification of elaborate sets of nonstandard regularity conditions, due to complications arising from the associated Fredholm integral equations of the first kind. Empirical examples of semiparametric estimates of (special cases of) equation (2) include Gallant and Tauchen (1989), Fleissig, Gallant, and Seater (2000), and Chen and Ludvigson (2009).

Let $f(C_{t+1}, V_{t+1} \mid C_t, V_t)$ denote the conditional density function of C_{t+1}, V_{t+1} given C_t, V_t and define $f_j(C_{t+1}, V_{t+1}, C_t, V_t) = E\left(R_j \mid C_{t+1}, V_{t+1}, C_t, V_t\right) f(C_{t+1}, V_{t+1} \mid C_t, V_t)$. Then the Euler equation (1) can be rewritten as

$$g(C_t, V_t) = bE\left[R_{jt+1} g(C_{t+1}, V_{t+1}) \mid C_t, V_t\right] \tag{3}$$

$$g(C_t, V_t) = b \int g(C_{t+1}, V_{t+1}) f_j(C_{t+1}, V_{t+1}, C_t, V_t) dC_{t+1} dV_{t+1}. \tag{4}$$

In a model without habits in utility, V_t will not include lags of $C_{t,}$, and in this case equation (4) is a Fredholm integral equation of the second kind, which can be solved using standard methods for b and the unknown function g. We base our estimators on equation (4), which has a well posed inverse, resulting in a simpler limiting distribution theory with less restrictive assumptions. As noted above, previous nonparametric and semiparametric Euler equation estimators were based on equations (1) or (2), which are Fredholm integral equation of the first kind in M_t and so suffer from the ill-posed inverse problem.

An exception is Anatolyev (1999), who shows consistency of a specific Euler equation estimator with a well posed inverse like ours. Another exception is Escanciano and Hoderlein (2010), who show identification of some versions of this model by imposing shape restrictions on utility functions, and providing sufficient conditions to ensure that only one element of the set we identify satisfies those shape restrictions. Unlike our results, they do not provide an explicit solution to the resulting integral equations which we use to construct estimators. A final exception is Chen, Chernozhukov, Lee, and Newey (2010), who in one example application of their local identification theorems consider semiparametric (not nonparametric) identification of an Euler equation model in an instrumental variables framework.

Interpreting equation (4) as a filtering problem, one would typically estimate b based on the largest eigenvalue associated with the Fredholm equation solution. However, while $1/b$ must be an eigenvalue, there is no economic theory that would require that individual's rate of time preference b to equal the inverse of the largest eigenvalue.

Allowing for habits in utility means that V_t includes lags of C_t, which complicates the resulting integral equation. In particular, the Fredholm kernel function $f_j(C_{t+1}, V_{t+1}, C_t, V_t)$ is then no longer an ordinary density. We show how standard methods for solving Fredholm integral equations of the second kind can be extended to handle this case. With habits the integral equation (4) still has a well posed inverse, though the corresponding integral equation solution and associated estimators become more complicated.

Previous Euler equation estimators either assume nonparametric identification of the model (which requires high level assumptions that are difficult to interpret or verify), or impose parametric or semiparamertric assumptions to attain identification.

Our recasting of the model in the form of a Fredholm equation of the second kind allows to derive more primitive conditions for identification. In particular, with minimal regularity we obtain conditions for local identification, identification up to finite sets, and for point identification.

Given our estimates of g and hence of the pricing kernel M, we may test whether g is independent of durables consumption, lagged consumption, or both, thereby testing whether durable consumption or habit formation plays a role in determining the pricing kernel. We will also want to test various popular functional restrictions on utility.

In addition to the pricing kernel M, other functions of the marginal utility function g that are of interest to estimate are the Arrow Pratt coefficients of relative and of absolute risk aversion (crra and cara), and marginal rates of substitution (mrs),

$$\begin{aligned} crra(x) &= \frac{-c\partial g(c,v)/\partial c}{g(c,v)} \\ cara(x) &= \frac{-\partial g(c,v)/\partial c}{g(c,v)} \\ mrs(x) &= \frac{\partial g(c,v)/\partial v}{\partial g(c,v)/\partial c} \end{aligned}$$

These measures are all independent of the scale of g. One might also be interested in the unconditional means of these functions, corresponding to the average risk aversion or average marginal rates of substitution in the population.

2 Euler Equation Derivation

To encompass a very wide 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 single period utility function, b is the one period subjective discount factor, C_t is expenditures on consumption, D_t is a stock of durables, and W_t is a vector of other variables that affect utility and are known at time t. Let V_t denote the vector of all variables other than C_t that affect utility in time t. In particular, V_t contains W_t, V_t contains D_t if durables matter, and V_t contains lagged consumption C_{t-1}, C_{t-2}, etc.,. if habits matter.

The consumer's time separable utility function is

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

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

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

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

$$E\left[\sum_{t=0}^{T} b^t U(C_t, V_t) - \left(C_t + (D_t - \delta D_{t-1}) P_t + \sum_{j=1}^{J} (A_{jt} - A_{jt-1} R_{jt})\right) \lambda_t\right] \tag{5}$$

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

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

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

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

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

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

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

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

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

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

With this notation, regardless of whether habits are internal or external, we may write the first order conditions associated with the Lagrangean (5) as

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

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

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

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

$$g(C_t, V_t) = bE\left[g(C_{t+1}, V_{t+1}) R_{jt+1} \mid C_t, V_t\right] \qquad j = 1, ..., J \tag{6}$$

which is the source of our estimated Fredholm equations.

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

$$g_d(C_t, V_t) = P_t g(C_t, V_t) + \delta bE\left[g(C_{t+1}, V_{t+1}) P_{t+1} \mid C_t, V_t, P_t\right]. \tag{7}$$

Given estimates of the function g, equation (7) would then provide an equation for estimating the function g_d. When habits are external, it would also be possible to estimate g and g_d simultaneously, imposing the additional constraint from Young's theorem that

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

3 Identification

In the previous section we derived the Euler equations (3) and (4), allowing V_t to depend on durables D_t, past consumption C_{t-1}, etc.,. Here we take these equations as our starting point. Because of the presence of lags, there can be elements of (C_{t+1}, V_{t+1}) that overlap with elements of (C_t, V_t), and we will need to treat those elements differently from others. To handle this overlap, and to simplify notation by eliminating subscripts, define Y to be the vector of elements of the intersection of the sets of elements of (C_t, V_t) and (C_{t+1}, V_{t+1}). Then define X to the elements of (C_{t+1}, V_{t+1}) that are not in Y, and define Z to be the elements of (C_t, V_t) that are not in Y. So, e.g., if $V_t = (C_{t-1}, C_{t-2}, D_t)$, corresponding to the model where utility depends on current consumption, two lags of consumption comprising habit effects, and current durables consumption, then we would have $Y = (C_t, C_{t-1})$, $X = (C_{t+1}, D_{t+1})$, and $Z = (C_{t-2}, D_t)$. If there are no lagged effects, so e.g. if $V_t = D_t$, then we would have $X = (C_{t+1}, D_{t+1})$, $Z = (C_t, D_t)$, and Y would be empty. Note that by construction X and Z will always have the same number of elements, and each element of Z will be a lag of the corresponding element of X. Also, the elements of Y, X, and Z need to be ordered properly so that $(C_t, V_t) = (Y, Z)$ and $(C_{t+1}, V_{t+1}) = (X, Y)$.

In this notation, equation (3) for marginal utility g can be written as

$$g(y, z) = bE\left[g(X, Y)R_j \mid Y = y, Z = z\right] = bE\left[g(X, y)R_j \mid Y = y, Z = z\right] \tag{8}$$

for random vectors X, Y, and Z, and random scalar $R_j = R_{jt}$. In the special case where Y is empty this simplifies to

$$g(z) = bE\left[g(X)R_j \mid Z = z\right] \tag{9}$$

We now show how to construct the set of function g that solve equation (8). The method is an extension of the standard technique for solving Fredholm integral equations of the second kind. When Y is empty, equation (8) reduces to equation (9), which is an ordinary Fredholm equation of the second kind, and in that case our method reduces to a standard solution method for such equations. Make the following Assumption I (for identification):

ASSUMPTION I. X and Z each have the same support Ω, and Y has support Ω_y Let $f(X \mid y, z)$ denote the conditional probability density function of the continuously distributed X, conditional on $Y = y, Z = z$. For each asset $j = 1, ..., J$, define $f_j(x, y, z) = E\left(R_j \mid X = x, Y = y, Z = z\right) f(x \mid y, z)$. Assume $g(y, z)$ is not zero for all z. Without loss of generality, assume some scaling is imposed for $g(y, z)$, such as that the square of $g(y, z)$ integrated over some known measure equals one. Assume $f_j(x, y, z)$ is square integrable, that is, $\int_\Omega \int_{\Omega_y} \int_\Omega f_j(x, y, z)^2 dxdydz$ is finite. Assume there exists functions a_{jk}, b_{jk} and c_{jk} such that

$$f_j(x,y,z) = \sum_{k=1}^{K} a_{jk}(x)b_{jk}(y)c_{jk}(z) \tag{10}$$

where K is either an integer or infinity.

By analogy with standard Fredholm theory, when equation (10) holds with K finite, we may define the Fredholm kernel function $f_j(x,y,z)$ to be degenerate, otherwise it is nondegenerate. A sufficient condition to have equation (10) hold is $f_j(x,y,z)$ analytic, because in that case functions a_{jk}, b_{jk}, and c_{jk} can be constructed by a Taylor series expansion.

Given Assumption I, equation (8) can be written as

$$g(y,z) = b\int_{\Omega} g(x,y)f_j(x,y,z)dx. \tag{11}$$

For each asset $j = 1, ..., J$ and each index k, ℓ, m on $1, ..., K$, define the scalars $A_{jk\ell m}$ and $\Gamma_{jk\ell}$ by

$$A_{jk\ell m} = \int_{\Omega} c_{jm}(z)b_{jk}(z)a_{j\ell}(z)dz$$

$$\Gamma_{jk\ell} = \int_{\Omega}\int_{\Omega_y} g(x,y)a_{jk}(x)dxb_{jk}(y)a_{j\ell}(y)dy.$$

Theorem 1. Let Assumption I hold. Then, For each asset $j = 1, ..., J$ and each index k, ℓ, m on $1, ..., K$,

$$\Gamma_{jk\ell} = b\sum_{m=1}^{K} \Gamma_{jmk}A_{jk\ell m} \tag{12}$$

and

$$g(y,z) = b\sum_{m=1}^{K}\sum_{k=1}^{K} \Gamma_{jkm}c_{jk}(z)b_{jm}(y)c_{jm}(z) \tag{13}$$

PROOF OF THEOREM 1: Define $B_{jk}(y) = \int_{\Omega} g(x,y)a_{jk}(x)dx$ so $\Gamma_{jk\ell} = \int_{\Omega_y} B_{jk}(y)b_{jk}(y)a_{j\ell}(y)dy$. Substitute equation (10) into equation (11) to get

$$g(y,z) = b\sum_{m=1}^{K}\int_{\Omega} g(x,y)a_{jm}(x)dxb_{jm}(y)c_{jm}(z) = b\sum_{m=1}^{K} B_{jm}(y)b_{jm}(y)c_{jm}(z). \tag{14}$$

Multiply this equation by $a_{jk}(y)$ and integrate over y

$$\int_{\Omega_y} g(y,z)a_{jk}(y)dy = b\sum_{m=1}^{K}\int_{\Omega_y} B_{jm}(y)b_{jm}(y)a_{jk}(y)dyc_{jm}(z)$$

which by the definitions of $B_{jk}(z)$ and Γ_{jkm} simplifies to

$$B_{jk}(z) = \sum_{m=1}^{K} \Gamma_{jmk} c_{jm}(z). \tag{15}$$

Now multiply this equation by $b_{jk}(z)a_{j\ell}(z)$ and integrate over z to get

$$\int_{\Omega} B_{jk}(z) b_{jk}(z) a_{j\ell}(z) dz = b \sum_{m=1}^{K} \Gamma_{jmk} \int_{\Omega} c_{jm}(z) b_{jk}(z) a_{j\ell}(z) dz$$

which, by applying the definitions of $\Gamma_{jk\ell}$ and $A_{jk\ell m}$ gives equation (12). Also, equations (14) and (15) together give equation (13).

Theorem 1 forms the basis for identification. Since $f_j(x, y, z)$ is the product of a conditional expectation of observables and a conditional density function of observables, we may assume $f_j(x, y, z)$ is identified. Then functions a_{jk}, b_{jk} and c_{jk} that satisfy equation (10) can be constructed, and from those the constants $A_{jk\ell m}$ can be constructed. We therefore have identification of $A_{jk\ell m}$. Equation (12) relates b and $\Gamma_{jk\ell}$ to $A_{jk\ell m}$, and by equation (13), $g(y, z)$ is identified if b and $\Gamma_{jk\ell}$ are identified. So identification depends on the extent to which b and $\Gamma_{jk\ell}$ can be recovered from equation (12), given $A_{jk\ell m}$.

Let $\Gamma_{jk\cdot}$ be the K vector of elements $\Gamma_{jk\ell}$ for $\ell = 1, ..., K$, let $\Gamma_{j\cdot k}$ be the K vector of elements $\Gamma_{j\ell k}$ for $\ell = 1, ..., K$, and let A_{jk} be the K by K matrix of having $A_{m\ell k}$ in the m'th row and ℓ'th column. Equation (12) can then be written as

$$\Gamma_{jk\cdot} = bA_{jk}\Gamma_{j\cdot k} \qquad \text{for } k = 1, ..., K \tag{16}$$

so

$$\begin{pmatrix} \Gamma_{j1\cdot} \\ \Gamma_{j2\cdot} \\ \vdots \\ \Gamma_{jK\cdot} \end{pmatrix} = b \begin{pmatrix} A_{j1} & 0 & 0 & 0 \\ 0 & A_{j2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & A_{jK} \end{pmatrix} \begin{pmatrix} \Gamma_{j\cdot 1} \\ \Gamma_{j\cdot 2} \\ \vdots \\ \Gamma_{j\cdot K} \end{pmatrix} \tag{17}$$

Let $\widetilde{\Gamma}_j$ be the K^2 element vector on the left in equation (17), let $\widetilde{A}_j$ denote the K^2 by K^2 block diagonal matrix in equation (17), and let Γ_j be the vector having K^2 elements where $\Gamma_{jk\ell}$ is the $(k \times K) + \ell$ element of Γ_j, so Γ_j is final vector in equation (17). We then have $\widetilde{\Gamma}_j = b\widetilde{A}_j\Gamma_j$. Now $\widetilde{\Gamma}_j$ contains all the same elements as Γ_j, but in a different order. For example, the first three elements of $\widetilde{\Gamma}_j$ are Γ_{j11}, Γ_{j12}, and Γ_{j13}, which are the elements in positions 1, $K+1$, and $2K+1$ in Γ_j. Let P denote the permutation matrix that makes $P\widetilde{\Gamma}_j = \Gamma_j$, and define the matrix $A_j = P\widetilde{A}_j$. Then $\Gamma_j = P\widetilde{\Gamma}_j = bA_j\Gamma_j$ so

$$(I - bA_j)\,\Gamma_j = 0 \tag{18}$$

Equation (18) is just a way to rewrite equation (12), so identification now depends on the extent to which b and Γ_j can be recovered from equation (18).

Corollary 1. Let Assumption I hold, removing the variable Y and the function b_{jk} everywhere they appear. In particular, $f_j(x,z) = \sum_{k=1}^{K} a_{jk}(x)c_{jk}(z)$ and equation (9) can then be written as

$$g(z) = b\int_{\Omega} g(x)f_j(x,z)dx. \tag{19}$$

Define $A_{j\ell m} = \int_{\Omega} a_{j\ell}(z)c_{jm}(z)dz$ and $\Gamma_{j\ell} = \int_{\Omega} g(z)a_{j\ell}(z)dz$. Then $\Gamma_{j\ell} = b\sum_{m=1}^{K}\Gamma_{jm}A_{j\ell m}$ and

$$g(z) = b\sum_{m=1}^{K}\Gamma_{jm}c_{jm}(z). \tag{20}$$

If we redefine A to be the K by K matrix of elements $A_{j\ell m}$ and redefine Γ to be the k matrix of elements $\Gamma_{j\ell}$, then $(I - bA_j)\,\Gamma_j = 0$.

PROOF OF COROLLARY 1: $g(z) = b\int_{\Omega} g(x)f_j(x,z)dx = b\int_{\Omega} g(x)\sum_{m=1}^{K} a_{jm}(x)c_{jm}(z)dx = b\sum_{m=1}^{K}\Gamma_{jm}c_{jm}(z)$. Multiply both sides of this expression by $a_{j\ell}(z)$ and integrate over z to get $\Gamma_{j\ell} = \int_{\Omega} g(z)a_{j\ell}(z)dz = b\sum_{m=1}^{K}\Gamma_{jm}\int_{\Omega} c_{jm}(z)a_{j\ell}(z)dz = b\sum_{m=1}^{K}\Gamma_{jm}A_{j\ell m}$, and this equation is $\Gamma_j = bA_j\Gamma_j$.

Corollary 1 is the analog to Theorem 1 when there is no Y variable and so applies to equation (9) instead of Equation (8) (that is, Euler equation models without habit or any other lagged variables in the marginal utility function). Equation (8) is a standard Fredholm equation of the second kind, and Corollary 1 corresponds to a standard method of solving such equations. Corollary 1 yields the same form $(I - bA_j)\,\Gamma_j = 0$ as Theorem 1, though with smaller matrices. Essentially, Corollary 1 directly yields an analog to equation (16) without the k subscripts, which can then be written directly as $(I - bA_j)\,\Gamma_j = 0$, without having to go through the steps of getting from equation (16) to equation (18).

We have now shown that starting from either equation (9) or (8), identification of b and the function g corresponds to recovering b and Γ_j with A_j known in equation (18), or equivalently $\left(\frac{1}{b}I - A_j\right)\Gamma_j = 0$. The equation $(\lambda_j I - A_j)\Gamma_j = 0$ is satisfied by up to K^2 different eigenvalues λ_j. Multiplicity of roots can result in there being more than one eigenvector Γ_j associated with any given eigenvalue λ_j. Let Λ_j be the set of all eigenvalues of A_j.

Point identification of b and g is likely (i.e., we have 'generic' identification in the sense of McManus 1992) when the data consist of multiple assets. Let S_j denote the set of all pairs λ_j, g_j such that λ_j, Γ_j are an eigenvalue and a corresponding eigenvector of A_j, and g_j is given either by equation (13) or (20) using Γ_j and replacing b with $1/\lambda_j$. Given J assets, the true pair b, g will be in the intersection of the sets S_j for $j = 1, ..., J$. Failure of point identification of b and g would therefore require that the exact same eigenvalue and associated g type function appear in every set S_j, which would require a great coincidence among the functions $E(R_j \mid X = x, Y = y, Z = z)$, and hence among the Fredholm kernels f_j for all the assets $j = 1, ..., J$.

Even when we have only one asset j (such as only having the risk free interest rate in models when we are only modeling marginal utility and not asset pricing), we still have useful set or point identification results (see, e.g., Manski 2003 for a general overview of set versus point identification). Let Λ_j be the set of all eigenvalues associated with the matrix A_j. Since K can be infinite, the identified set Λ_j (for the given asset j) that $1/b$ must lie in can be infinite, but is still countable, so we have identification up to a countable set.

It is unreasonable to think that an individual's rate of time preference corresponds to a subjective discount rate that is either above one or arbitrarily close to zero, so assume $1 \geq b \geq b_*$ for some positive constant b_*. The eigenvalues associated with A_j can be ordered from largest to smallest, and will in general be a series that approaches zero, so there can only be a finite number of eigenvalues that satisfy $1 \geq b \geq b_*$. Therefore, assuming $1 \geq b \geq b_*$, we have identification up to a finite set, with $1/b \in \{\lambda_j \in \Lambda_j \mid 1 \geq 1/\lambda_j \geq b_*\}$. We will alternatively have a finite number of eigenvalues, and hence identification up to a finite set, if the Fredholm kernel f_j for any asset j is degenerate, corresponding to a finite K. Note that finite set identification also implies local identification (that is, identified in an open neighborhood of the true; see Rothenberg 1971).

Given point or finite set identification of b, the marginal utility function g (and hence also the pricing kernel) will be point identified (or identified up to the same size set) as long as there are not multiple roots and hence multiple eigenvectors associated with the eigenvectors in the identified set for b.

In a special case of our model, Escanciano and Hoderlein (2010) show point identification of marginal utility and the discount factor after imposing some shape restrictions on marginal utility that we do not impose. This point identification requires additional restrictions, notably bounded support of consumption, that may not be plausible in our data. however, their results suggest that we may further shrink the identified set by imposing shape restrictions beyond smoothness on the marginal utility function g and bounds on the discount factor b. In particular, it may be sensible to impose nonincreasing marginal utility, meaning that $\partial g(x, y) / \partial x \leq 0$.

Finally, if (as is commonly done), the subjective discount rate is assumed to equal a constant the risk free rate, then taking R_j to be this risk free rate will cause b and R_j to drop out of the Euler equation. Equivalently, we could in that case set b and R_j equal to one for the purpose of estimating g.

Based on market efficiency, it may be reasonable to assume

$$E(R_j \mid X = x, Y = y, Z = z) = E(R_j \mid X = x) \qquad j = 1, ..., J \tag{21}$$

meaning that, after conditioning on the time $t+1$ information X, the mean returns of each asset j in time $t+1$ do not depend on information dated time t or earlier. If equation (21) holds then to solve $(I - bA_j)\Gamma_j = 0$ for every asset j we only need to decompose the density function f as $f(x \mid y, z) = \sum_{k=1}^{K} a_k(x)b_k(y)c_k(z)$ for some functions a_k, b_k, and c_k. Then equation (10) will hold for all assets j, with $a_{jk}(x) = a_k(x)E(R_j \mid X = x)$, and $b_{jk}(y) = b_k(y)$, and $c_{jk}(z) = c_k(z)$. This could be an empirically valuable simplification.

4 Estimation Overview

Suppose will be based on aggregate time series data on consumption and asset returns. Then the derivations provided here would be those of a representative consumer, though in many cases it is possible to obtain the same Euler equations by aggregating the demand functions of individual consumers. Then one will need to include sufficient time varying covariates in V_t to have the Fredholm kernel functions f_j (which depend on the conditional density function of consumption and on conditional mean of asset returns) be constant over time. A complication is that aggregate consumption may be nonstationary. In this case, we might specify the model in terms of relative consumption as in, e.g., Chen and Ludvigson (2009).

On the other hand, suppose we have data on individual consumers, so X, Y, and Z are then vectors of consumption in a few time periods and other variables pertaining to individual consumers. If we have cross section data from a single pair of time periods only then each R_j will be constant across observations, so we would not be able to exploit variation in asset prices to aid identification. Panel data would provide some variation, but in that case $E(R_j \mid X = x, Y = y, Z = z)$ will vary only by the time period of the observations X, Y, and Z, so the amount of variation coming from asset prices will be limited unless the panel is long.

For this preliminary overview, assume we have a sample $\{(X_i, Y_i, Z_i, R_{ji}), i = 1, \ldots, n\}$ where X_i, Y_i, and Z_i are each continuously distributed vectors with supports $X_i \in \Omega \subset \mathbb{R}^{\ell_1}$, $Z_i \in \Omega \subset \mathbb{R}^{\ell_1}$, and $Y_i \in \Omega_y \subset \mathbb{R}^{\ell_2}$. Assume the data are stationary. This data could either consist of a few periods of consumption and possibly durables for each of many individuals i, or be time series data with growth removed from consumption in some way. Note R_{ji} generally only varies over time.

For example, in a model with habits based on aggregate consumption data, we could have $i = t = 1, ..., T \to \infty$, $X_i = (C_{t+1}/C_t)$, $Y_i = (C_t/C_{t-1})$, and $Z_i = (C_{t-1}/C_{t-2})$, where we have written the model in the form of relative consumption as in Chen and Ludvigson (2009) to remove possible nonstationarity in the levels of consumption. This would then be imposing the restriction that marginal utility at time t is homogeneous of degree zero in C_t, C_{t-1}, and C_{t-2}, but is otherwise nonparametric. Assets R_{jt} could include a risk free rate like the short term treasury bill rate, stock or bond market index rates, rates on various beta portfolios, etc.,.

For an example using micro data, in a model with habits, durables D, and households h, we could (after conditioning on sufficient observable household characteristics to make marginal utility plausibly homogeneous) take $i = i(h,t)$, $h = 1,...,H \to \infty$, $t = 1,...,T$ fixed; $X_i = (C_{ht}, D_{ht})$, $Y_i = C_{ht-1}$, $Z_i = (C_{ht-2}, D_{ht-1})$, and $R_{ji} = R_{jt}$. If the model with this data did not include habits, then we would instead drop Y and take $X_i = (C_{ht}, D_{ht})$ and $Z_i = (C_{ht-1}, D_{ht-1})$. Dropping D everywhere gives the model without durables, with or without habits.

The goal is estimation of the function g and discount value b in $g(y,z) = b\int_\Omega g(x,y)f_j(x,y,z)dx$, where the Fredholm kernel f_j is given by $f_j(x,y,z) = E(R_j \mid X = x, Y = y, Z = z) f(x \mid y,z)$ for each asset j. Let $\widehat{f}_j(x,y,z)$ be a uniformly consistent estimator of $f_j(x,y,z)$. Below, use a kernel estimator.

For deriving our asymptotic theory, we define our estimators $\widehat{b}$ and $\widehat{g}$ as solutions to the empirical analog of our identifying equations, that is, we define $\widehat{b}$ and $\widehat{g}$ by $\widehat{g}(y,z) = \widehat{b}\int_\Omega \widehat{g}(x,y)\widehat{f}_j(x,y,z)\,dx$, subject to normalizations and the identifying assumptions.

In practice, estimates $\widehat{b}$ and $\widehat{g}$ could be obtained by applying the solution method given in Theorem 1 using $\widehat{f}_j$ in place of f_j, with $1/b$ being the associated eigenvalue and $\widehat{g}$ given by the estimated analog to equation (13).

The integrals over Ω and Ω_y that are used to implement a solution will be numerical, and hence equivalent to summations over a fine grid of points which we may denote as Ω^* and Ω_y^*. Therefore, if desired one could more directly construct estimates $\widehat{g}$ and $\widehat{b}$ as solutions to

$$\widehat{g}(y,z) = \widehat{b}\sum_{x\in\Omega^*}\widehat{g}(x,y)\,\widehat{f}_j(x,y,z), \qquad \text{all } y \in \Omega_y^*,\ z \in \Omega^*,\ j = 1,...,J \tag{22}$$

subject to a scale normalization (imposed without loss of generality) such as $\sum_{y\in\Omega_y^*}\sum_{z\in\Omega^*}\widehat{g}(y,z)^2\varpi(y,z) = 1$ for some known measure $\varpi(y,z)$. It then becomes a numerical search to find the finite number of values of $\widehat{b}$ and $\widehat{g}(x,y)$ for all $x \in \Omega^*$ and $y \in \Omega_y^*$ that satisfy equation (22), the scale normalization for $\widehat{g}$, and any known equalities or inequalities on $\widehat{b}$, specifically $b_* \leq \widehat{b} \leq 1$.

Our estimator yields countable sets of values of $\widehat{b}$ and $\widehat{g}(x,y)$ given $\widehat{f}_j(x,y,z)$. The fact that plausible discount factors lie in a relative narrow range of values means we are likely to have either point identification (if only one candidate $\widehat{b}$ lies in a plausible range) or identification up to a small number of values. If we have more than one asset, so $J > 1$, then we could apply some minimum distance estimator to impose the constraint that the true estimator be the value of $\widehat{g}(y,z)$ and $\widehat{b}$ that is the same across all assets J. As noted earlier, if these restrictions do not yield point identification, then we might further shrink the identified set by imposing shape restrictions beyond smoothness on the marginal utility function g. In particular, it may be sensible to impose nonincreasing marginal utility, meaning that $\partial g(x,y)/\partial x \leq 0$.

For models without habits, all of the above steps can be applied after dropping all of the terms involving y.

5 Asymptotic Theory

Here we provide conditions for consistency and limiting distribution theory for our estimator. We begin by recapping our notation and adding additional notation as needed. Let D_t be a vector of observable variables other than total consumption C_t that affects utility (e.g., it could consumption of just durables). Let $V_t = (C_{t-1}, ...C_{t-L_1}, D_t, ...D_{t-L_2})$ for some non-negative integers L_1 and L_2. If $L_1 = 0$, $L_2 > 0$ then $V_t = (D_t, ...D_{t-L_2})$; if $L_1 > 0$, $L_2 = 0$ then $V_t = (C_{t-1}, ...C_{t-L_2})$; and if $L_1 = L_2 = 0$ then V_t is empty. Define Y_t to be the intersection of the sets of elements of (C_t, V_t) and (C_{t+1}, V_{t+1}), define X_t to be the elements of (C_{t+1}, V_{t+1}) that are not in Y, and define Z_t to be the elements of (C_t, V_t).

Therefore, if L_1 and L_2 are positive then $Y = (C_{t-1}, ...C_{t-L_1+1}, D_t, ...D_{t-L_2+1})$, $X_t = (C_{t+1}, D_{t+1})$, and $Z_t = (C_{t-L_1}, D_{t-L_2})$. If $L_1 = 0$, $L_2 > 0$ then $Y_t = (D_t, ...D_{t-L_2+1})$, $X_t = (C_{t+1}, D_{t+1})$, and $Z_t = (C_t, D_{t-L_2})$. If $L_1 > 0$, $L_2 = 0$ then $Y_t = (C_{t-1}, ...C_{t-L_1+1})$, $X_t = C_{t+1}$, and $Z_t = C_{t-L_1}$. If $L_1 = L_2 = 0$ then Y_t is empty, $X_t = C_{t+1}$, and $Z_t = C_t$.

Assume the elements of Y, X, and Z are ordered properly so that $(C_t, V_t) = (Y_t, Z_t)$ and $(C_{t+1}, V_{t+1}) = (X_t, Y_t)$. Let the support of X_t be $\Omega \subset \mathbb{R}^{\ell_1}$, Z_t has the same support as X_t, and let the support of Y_t be $\Omega_y \subset \mathbb{R}^{\ell_2}$. Let $S \subset \mathbb{R}^{\ell_2} \times \mathbb{R}^{\ell_1}$ denote the support of (Y_t, Z_t)

A_j is a linear operator so that

$$A_j g(y,z) = \int g(x,y) f_j(x,y,z) dx,$$

where

$$m_j(x,y,z) = E[R_{jt+1} | X_t = x, Y_t = y, Z_t = z]$$

$$f_j(x,y,z) = m_j(x,y,z) \times f_{x|y,z}(x|y,z).$$

Let $r_j(x,y,z) = f_j(x,y,z) \times f_{y,z}(y,z)$. Let K_j be an ℓ_j-dimensional product of kernel functions K, let h be a bandwidth, and let $K_{jh}(.) = K_j(./h)/h^{\ell_j}$. We define the following kernel based estimators, where all the summations are from 1 to T.

$$\widehat{r}_j(x,y,z) = \frac{1}{T} \sum R_{jt+1} K_{1h}(x - X_t) K_{2h}(y - Y_t) K_{1h}(z - Z_t)$$

$$\widehat{f}_{x,y,z}(x,y,z) = \frac{1}{T}\sum K_{1h}(x - X_t) K_{2h}(y - Y_t) K_{1h}(z - Z_t)$$

$$\widehat{m}_j(x,y,z) = \frac{\widehat{r}_j(x,y,z)}{\widehat{f}_{x,y,z}(x,y,z)} = \widehat{E}(R_{jt+1} \mid X_t = x, Y_t = y, Z_t = z)$$

$$\widehat{f}_{y,z}(y,z) = \frac{1}{T}\sum K_{2h}(y - Y_t) K_{1h}(z - Z_t)$$

$$\widehat{f}_{x|y,z}(x|y,z) = \frac{\widehat{f}_{x,y,z}(x,y,z)}{\widehat{f}_{y,z}(y,z)}$$

$$\widehat{f}_j(x,y,z) = \frac{\widehat{r}_j(x,y,z)}{\widehat{f}_{y,z}(y,z)} = \widehat{m}(x,y,z)\,\widehat{f}_{x|y,z}(x|y,z)$$

Our (set of) estimators $\left(\widehat{b}, \widehat{g}\right)$ are defined (in Assumption A below) to satisfy the empirical analog linear equations of (11), subject to the same normalizations that we imposed on b and g. Let Λ_j denote the set of real eigenvalues that lie in $(1, \infty)$ associated with the linear operator A_j. We consider the following set of assumptions for our asymptotic results.

Assumption A.

1. A_j is a compact operator for each asset $1 \leq j \leq J$.

2. The eigenspace corresponding to the eigenvalue that lies in $\Lambda = \bigcap_{j=1}^{J} \Lambda_j$ has dimension 1.

3. There exists a $(y_0, z_0) \in S$ such that any eigenfunction that corresponds to an eigenvalue in Λ is non-zero at (y_0, z_0).

4. For $j = 1, \ldots, J$, the estimator $\left(\widehat{b}, \widehat{g}\right)$ is defined to solve $\widehat{g}(y,z) = \widehat{b}\int \widehat{g}(x,y)\,\widehat{f}_j(x,y,z)\,dx$ such that $\int \widehat{g}^2(y,z)\,\widehat{f}_{y,z}(y,z)\,dx = 1$, and $\widehat{g}(y_0,z_0) > 0$ for some $(y_0,z_0) \in S$.

Assumption B.

1. The sequence $(C_t, D_t, R_{1t}, \ldots, R_{Jt})_{t=1}^{T}$ is a strictly stationary and geometrically strong mixing sequence, satisfying the Euler equation (4), whose marginal distribution coincides with the distribution of $(C, D, R_1, \ldots, R_J)$.

2. The support of (Y_t, Z_t), S, is a compact subset of $\mathbb{R}^{\ell_2} \times \mathbb{R}^{\ell_1}$.

3. The probability density function $f_{x,y,z}(x,y,z)$ is continuous and bounded away from zero.

4. The regression function $m_j(x,y,z)$ is continuous.

5. The kernel function K is a bounded, symmetric, 2nd-order kernel with support $[-1,1]$ that integrates to 1.

6. As $T \to \infty, h_T \to 0, Th_T^2 \to 0$ and $Th_T^{2\ell_1+\ell_2}/\log(T)^2 \to \infty$.

7. As $T \to \infty, Th_T^2 \to 0$.

Assumption C.

1. The probability density function $f_{x,y,z}(x,y,z)$ is twice continuously differentiable.

2. The regression function $m_j(x,y,z)$ is twice continuously differentiable.

3. As $T \to \infty, Th_T^{\ell_1+\ell_2+4} \to 0$.

Comments:

The first two conditions of Assumption A are high level conditions. Compactness ensures that Λ_j is countable and accumulates only at 0. No repeated eigenvalues is the other key assumption we have to make. This can be relaxed to the case where we allow for repeated roots as long as the geometric multiplicity is 1. The third condition in Assumption A is a harmless sign normalization, so when we estimate an eigenfunction we incorporate the constraint that $g(y_0,z_0) > 0$. Assumptions B.1 - B.6 and C are standard conditions in the nonparametric statistics literature. Note that, as is common with semiparametric problems, B.7 imposes an undersmoothing condition required to ensure a parametric rate of convergence for the estimator of the discount factor.

We have the following theorems (the proofs are provided in the Appendix.):

Theorem 2. Under Assumptions A and B, if $g_0 = b_0 A_j g_0$ for all $j = 1, \ldots, J$ such that $\int g_0^2(y,z) f_{y,z}(y,z)\,dx = 1$, and, $g(y_0,z_0) > 0$ for some $(y_0,z_0) \in S$ then as $T \to \infty$**:**

1. $\widehat{b} \xrightarrow{p} b_0$.

2. For all $(y,z) \in int(S)$, $|\widehat{g}(y,z) - g_0(y,z)| \xrightarrow{p} 0$.

Theorem 3. Under Assumptions A, B and C, if $g_0 = b_0 A_j g_0$ for all $j = 1, \ldots, J$ such that $\int g_0^2(y,z) f_{y,z}(y,z)\,dx = 1$, and, $g(y_0,z_0) > 0$ for some $(y_0,z_0) \in S$ then as $T \to \infty$**:**

1. Let $\sigma^2(x,y,z) = E[e_t^2 | X_t = x, Y_t = y, Z_t = z]$, where $e_t = R_{jt+1} - m(X_t,Y_t,Z_t)$, then
$$\sqrt{T}\left(\widehat{b} - b_0\right) \xrightarrow{d} N\left(0, E\left[\sigma^2(X_t,Y_t,Z_t)\left[\frac{g_0(X_t,Y_t)\,h_0(Y_t,Z_t)}{f_{y,z}(Y_t,Z_t)}\right]^2\right]\right).$$

2. For all $(x, y) \in int\,(S)$

$$\sqrt{Th_T^{\ell_1+\ell_2}}\left(\widehat{g}\,(y,z) - g_0\,(y,z)\right) = (I-B)^{\dagger}\sqrt{Th_T^{\ell_1+\ell_2}}C_T\,(x,y) + o_p\,(1)\,,$$

where $(I-B)^{\dagger}$ is the Moore-Penrose pseudoinverse of $(I-B)$, and $\sqrt{Th_T^{\ell_1+\ell_2}}C_T\,(x,y) \xrightarrow{d} N\,(0, \Sigma_0\,(x,y))$. The process C_T is defined in (33) and the explicit formula for $\Sigma_0\,(x,y)$ is provided in equation (34) of the Appendix.

In the Appendix we provide explicit formulae for the asymptotic variance of the estimators of both the discount factor $\widehat{b}$ and the corresponding eigenfunction. There are clear sample analogues available to perform pointwise inference. In particular, for the eigenfunction, the parameters of the stochastic process C_T and the operator $(I-B)^{\dagger}$ can be estimated, so it is feasible approximate the asymptotic distribution given above by discretization.

6 Empirical Implementation - *Preliminary*

As in Chen and Ludvigson (2009), we define C_t to be quarterly, seasonally adjusted real per capita expenditures on services and nondurables, constructed from US National Income and Products Account tables 2.3.4 and 2.3.5 (with clothing and shoes excluded from the NIPA definition of nondurables). We use 220 quarters of data starting from the first quarter of 1954. We let $X_t = (C_{t+1}/C_t)$, $Y_t = (C_t/C_{t-1})$, and $Z_+ = (C_{t-1}/C_{t-2})$, where we have written the model in the form of relative consumption to remove nonstationarity in the levels of consumption. We consider two different assets. The riskless asset $j = 1$ is the yield on one-month US treasury bills, and the risky asset $j = 2$ is the excess market return on the US stock market.

Table 1 shows the ten largest eigenvalues between zero of one of the estimated operator, corresponding to possible estimates of the discount factor b, using the riskfree and risky return separately, for various bandwidths where $h = std\,(C_t)\,T^{-1/6}$.

The bolded numbers in these tables show the discount factors used to construct corresponding nonparametric estimates of the marginal utility function. Since the discount rate should be the same regardless of asset, we estimated the bolded choice of discount factor for each bandwidth by minimizing the difference between the estimated factors in the two tables, among all factors in the economically plausible range of .7 to 1. The estimates across the tables were generally within a few percent of each other, which is reassuring for estimates that should be asymptotically the same if the model is correct. Note that while we often found repeated roots, the geometric multiplicity was found to one, so the multiple eigenvalues do not appear to cause a failure of identification.

Figures 1 to 4 plot the nonparametric estimates of marginal utility as a function of $\log(C_t)$ corresponding to each bandwidth in Table 1 respectively. For comparison, a parametric estimate is in black. The parametric model is the CRRA power utility as in Hansen and Singleton (1982).[1] In this parametric model the estimated coefficient of risk aversion is 0.5 and the estimated discount factor is 0.9728.

The resulting estimated marginal utility functions are quite choppy, perhaps suggesting that still larger bandwidths might be appropriate. The nonparametric functions show some curvature indicating inadquacy of the parametric model. The estimated marginal utility functions are quite similar across the two assets (particularly for the largest bandwidth), which is consistent with theory that says they should be the same across assets.

7 Extensions

An alternative estimation method would be to first directly estimate $\widehat{f}_j(x,y,z)$ in the form

$$\widehat{f}_j(x,y,z) = \sum_{k=1}^{K^*} \widehat{a}_{jk}(x)\widehat{b}_{jk}(y)\widehat{c}_{jk}(z) \quad \text{for all } x,y,z \in \Omega \tag{23}$$

for some chosen K^* that goes to infinity with the sample size. For example, one might estimate

$$\widehat{f}_j(x,y,z) = \sum_{k_1=1}^{\widetilde{K}} \sum_{k_2=1}^{\widetilde{K}} \sum_{k_3=1}^{\widetilde{K}} \widehat{\beta}_{jk_1k_2k_3} a_{k_1}(x) b_{k_2}(y) c_{k_3}(z)$$

by a sieve maximum likelihood estimator where $K^* = \widetilde{K}^3$ and $a_k(x)$, $b_k(y)$, $c_k(z)$ for $k = 1, ..., \widetilde{K}$ are some relevant space spanning basis functions like polynomials or sines and cosines. Then following the steps of the proof of Theorem 1, for each asset $j = 1, ..., J$ we obtain $\left(I - b\widehat{A}_j\right)\Gamma_j = 0$, where $\widehat{A}_j$ is a K^{*2} by K^{*2} known matrix function of $\widehat{a}_{jk}(x)$, $\widehat{b}_{jk}(y)$ and $\widehat{c}_{jk}(z)$. Given an eigenvalue $\widehat{b}$ (as before, one that satisfies any known equalities or inequalities regarding subjective discount rates) and corresponding eigenvector $\widehat{\Gamma}_j$ of elements $\widehat{\Gamma}_{jkm}$ from the matrix $\widehat{A}_j$, the corresponding estimator of $\widehat{g}$ would then be

$$\widehat{g}(y,z) = \widehat{b} \sum_{m=1}^{K} \sum_{k=1}^{K} \Gamma_{jkm} \widehat{c}_{jk}(z) \widehat{b}_{jm}(y) \widehat{c}_{jm}(z) \;\; \text{for any } y,z \in \Omega. \tag{24}$$

[1] The parametric utility function of consumption considered is $U(C) = \frac{C^{1-\rho}}{1-\rho}$, where ρ^{-1} is the intertemporal substitution of elasticity between consumption in any two periods. The moment conditions are based on $E\left[bR_{jt+1}\left(\frac{C_{t+1}}{C_t}\right)^{-\rho} - 1\right] = 0$, for the risk free and risky assets respectively.

Depending on the application, there may be some additional structure that one would like to incorporate into g on estimation. In particular, models with habit generally assume the structure $g(C_t, V_t) = G(C_t/C_{t-1} + H(V_t))$, where the function $-H(V_t)$ represents habit, with $V_t = C_{t-1}/C_t,...,C_{t-k}/C_t$. See, e.g., Chen and Ludvigson (2009), who estimate a model of this form with parametric G and nonparametric H. The comparable structure $g(C_t, V_t) = G(C_t+H(V_t))$ would arise if C_t is nondurables consumption, V_t is current and lagged expenditures on durables $D_{t-1},...,D_{t-k}$, and $H(V_t)$ is the flow of period t consumption services derived from the current stock of durables, accounting for depreciation over time. We could also consider imposing shape restrictions on utility as in Escanciano and Hoderlein (2010), to aid with identification.

8 Appendix

Consider a single asset for now so suppress the index j on A_j. Let $\mathcal{G}$ be a space of functions, say $L^2(S)$. Recall that $A : \mathcal{G} \to \mathcal{G}$ is a linear operator such that $Ag(y,z) = \int g(x,y) f_j(x,y,z)\, dx$ for any $g \in \mathcal{G}$. We consider the solution to the following linear equation

$$g = bAg,$$

for some known $b \in (0,1)$. Therefore $1/b$ is an eigenvalue of the operator A that lies in Λ.

The approach of the proof is a combination of those found in Darolles, Florens and Gourieroux (2004, hereafter DFG) and Magnus (1985). Magnus' paper provides explicit formulae, in finite dimension, for the differentials of the eigenvalue and eigenvector as a smooth function of a given matrix. We extend this to the infinite dimensional case and show that the estimators for the candidate of the discounting factors has an asymptotically normal distribution, converging at the rate of root$-T$. However, the results for the eigenvector of Magnus rely on less conventional normalization choice. Instead, to get the asymptotic distribution of the nonparametric estimator of the marginal utility of consumption, i.e. the eigenfunction for a particular eigenvalue of the linear operator, we show that the first order conditions of our objective function leads to a countable sequence of an eigenvalue/function identity of a self-adjoint operator despite our initial linear operator not being necessarily self-adjoint. We outline how to obtain the distribution theory of the eigenfunction satisfying these first order conditions.

To establish the large sample properties of our estimators, we first introduce some lemmas. The first two lemmas state the uniform consistency involving the regression functions and joint densities of continuous random variables, which follow from Roussas (1988) and Bosq (1998). The remaining lemmas are results relating some integral transform of $\widehat{f}_j$. Note that all the lemmas only require Assumptions B.1 - B.6.

Lemma 1 *Under Assumption B.1 - B.6:* $\sup_{y,z}\left|\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right|=o(1)$ *a.s.*

Lemma 2 *Under Assumption B.1 - B.6:* $\sup_{x,y,z}\left|\widehat{f}_j(x,y,z)-f_j(x,y,z)\right|=o(1)$ *a.s.*

Proof. This follows from the uniform consistency of the nonparametric estimators of $f(x|y,z)$ and $E[R_{jt+1}|X_t=x,Y_t=y,Z_t=z]$ to their respective truths, which are uniformly bounded. ■

Since we are particularly interested in the functions in $\mathcal{G}$ that have unit norm, the following result is particularly useful.

Lemma 3 *Under Assumption B.1 - B.6:* $\int g(y,z)\left|\widehat{f}(y,z)-f(y,z)\right|dydz$ *converges a.s. uniformly to zero for any function* $g\in\mathcal{G}_0$, *where for some* M_0, $\mathcal{G}_0=\left\{g:\int|g(y,z)|\,f(y,z)\,dydz\leq M_0\right\}$ *is a set of uniformly integrable functions, a subset of* $\mathcal{G}$.

Proof. This follows from Lemma 1, once noticing that since any function $\sup_g\left|\int g(y,z)\left(\widehat{f}(y,z)-f(y,z)\right)\right.$ is bounded above by $M_0\times\sup_{y,z}\left|\frac{\widehat{f}(y,z)-f(y,z)}{f(y,z)}\right|$. ■

Lemma 4 *Under Assumption B.1 - B.6, for each* (x,y,z) *in the interior of the support of* X_t,Y_t,Z_t*:*

$$\sqrt{Th_T^{2\ell_1+\ell_2}}\left(\widehat{f}_j(x,y,z)-f_j(x,y,z)\right)\Rightarrow N(0,\Sigma(x,y,z)),$$

where

$$\Sigma(x,y,z)=\kappa_2(K)^{2\ell_1+\ell_2}f_{x,y,z}(x,y,z)\left(r^2(x,y,z)+\frac{\sigma^2(x,y,z)}{f_{y,z}^2(y,z)}\right),$$

with $\kappa_2(K)=\int K^2(u)\,du$, $\sigma^2(x,y,z)=E[e_t^2|X_t=x,Y_t=y,Z_t=z]$ and $e_t=R_{jt+1}-m(X_t,Y_t,Z_t)$.

Proof.

$$\begin{aligned}\widehat{f}_j(x,y,z)-f_j(x,y,z)&=\widehat{m}(x,y,z)\widehat{f}_{x|y,z}(x|y,z)-m(x,y,z)f_{x|y,z}(x|y,z)\\&\simeq m(x,y,z)\left(\widehat{f}_{x|y,z}(x|y,z)-f_{x|y,z}(x|y,z)\right)\\&\quad+f_{x|y,z}(x|y,z)\left(\widehat{m}(x,y,z)-m(x,y,z)\right).\end{aligned}$$

We can expand uniformly $\widehat{f}_{x|y,z}(x|y,z)-f_{x|y,z}(x|y,z)$:

$$\begin{aligned}\widehat{f}_{x|y,z}(x|y,z)-f_{x|y,z}(x|y,z)&=\frac{\widehat{f}_{x,y,z}(x,y,z)}{\widehat{f}_{y,z}(y,z)}-\frac{f_{x,y,z}(x,y,z)}{f_{y,z}(y,z)}\\&\simeq\frac{\widehat{f}_{x,y,z}(x,y,z)-f_{x,y,z}(x,y,z)}{f_{y,z}(y,z)}-\frac{f_{x,y,z}(x,y,z)}{f_{y,z}^2(y,z)}\left(\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right),\end{aligned}$$

the leading stochastic term comes from the nonparametric density estimator for $f_{x,y,z}$. Next, we do the same for $\widehat{m}(x,y,z)-m(x,y,z)$:

$$\begin{aligned}\widehat{m}(x,y,z)-m(x,y,z) &= \frac{\widehat{r}(x,y,z)}{\widehat{f}_{x,y,z}(x,y,z)}-m(x,y,z)\\ &= \frac{1}{\widehat{f}_{x,y,z}(x,y,z)}\left(\widehat{r}(x,y,z)-\widehat{f}_{x,y,z}(x,y,z)\,m(x,y,z)\right),\end{aligned}$$

since $\sup_{x,y,z}\left|\widehat{f}_{x,y,z}(x,y,z)-f_{x,y,z}(x,y,z)\right|=o_p(1)$, we focus on the numerator

$$\begin{aligned}&\widehat{r}(x,y,z)-\widehat{f}_{x,y,z}(x,y,z)\,m(x,y,z)\\ =\;& \frac{1}{T}\sum\left(R_{jt+1}-m(x,y,z)\right)K_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\\ =\;& \frac{1}{T}\sum\left(m(X_t,Y_t,Z_t)-m(x,y,z)\right)K_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\\ &+\frac{1}{T}\sum e_tK_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t),\end{aligned}$$

where $R_{jt+1}=m(X_t,Y_t,Z_t)+e_t$, and $\{e_t\}$ is a martingale difference sequence (MDS). The first term on the RHS of the second equality is dominated by the bias term that is uniformly $O(h^2)$. In sum

$$\begin{aligned}\widehat{f}_j(x,y,z)-f_j(x,y,z) &= \frac{m(x,y,z)}{f_{y,z}(y,z)}\left(\frac{1}{T}\sum\left[\begin{array}{c}K_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\\ -E\left[K_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\right]\end{array}\right]\right)\\ &+\frac{f_{x|y,z}(x|y,z)}{f_{x,y,z}(x,y,z)}\left(\frac{1}{T}\sum e_tK_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\right)\\ &+O_p\left(h^2+\frac{\log T}{\sqrt{Th_T^{2\ell_1+\ell_2}}}\right).\end{aligned}$$

To reduce the notation, we replace $m(x,y,z)/f_{y,z}(y,z)$ by $r(x,y,z)$ and $f_{x|y,z}(x|y,z)/f_{x,y,z}(x,y,z)$ by $1/f_{y,z}(y,z)$. By the property of the MDS, we can ignore the covariance terms, then it is follows from the usual CLT that $\Sigma(x,y,z)$ has the desired expression. ■

Lemma 5 *Under Assumption B.1 - B.6, for each (y,z) in the interior of S, for any $\varphi\in L^2(S)$:*

$$\sqrt{Th_T^{\ell_1+\ell_2}}\int\varphi(x,y)\left(\widehat{f}_j(x,y,z)-f_j(x,y,z)\right)dx\Rightarrow N\left(0,\Sigma_\varphi(y,z)\right)$$

where

$$\Sigma_\varphi(y,z)=\frac{\kappa_2(K)^{\ell_1+\ell_2}}{f_{y,z}^2(y,z)}\int\sigma^2(x,y,z)\,\varphi(x,y)\,f_{x,y,z}(x,y,z)\,dx,$$

with $\kappa_2(K)=\int K^2(u)\,du, \sigma^2(x,y,z)=E\left[e_t^2|X_t=x,Y_t=y,Z_t=z\right]$ and $e_t=R_{jt+1}-m(X_t,Y_t,Z_t)$.

Proof. To proceed, it will be convenient to express $\widehat{f}_j(x,y,z)$ as the ratio $\frac{\widehat{r}(x,y,z)}{\widehat{f}_{y,z}(y,z)}$,

$$\begin{aligned}\int\varphi(x,y)\left(\widehat{f}_j(x,y,z)-f_j(x,y,z)\right)dx &= \int\varphi(x,y)\left(\frac{\widehat{r}(x,y,z)}{\widehat{f}_{y,z}(y,z)}-\frac{r(x,y,z)}{f_{y,z}(y,z)}\right)dx\\ &= \frac{1}{\widehat{f}_{y,z}(y,z)}\int\varphi(x,y)\left(\widehat{r}(x,y,z)-\widehat{f}_{y,z}(y,z)\frac{r(x,y,z)}{f_{y,z}(y,z)}\right)dx.\end{aligned}$$

We focus on the integral since $\sup_{y,z}\left|\widehat{f}_{y,z}(y,z)-f_{y,z}(,y,z)\right|=o_p(1)$. First, note that

$$\begin{aligned}\int\varphi(x,y)\widehat{r}(x,y,z)\,dx &= \frac{1}{T}\sum R_{jt+1}\int\varphi(x,y)K_h(x-X_t)K_h(y-Y_t)K_h(z-Z_t)\,dx\\ &\simeq \frac{1}{T}\sum R_{jt+1}\varphi(X_t,y)K_h(y-Y_t)K_h(z-Z_t),\end{aligned}$$

therefore

$$\begin{aligned}&\int\varphi(x,y)\left(\widehat{f}_j(x,y,z)-f_j(x,y,z)\right)dx\\ \simeq\ & \frac{1}{f_{y,z}(y,z)}\frac{1}{T}\sum\left(m(X_t,Y_t,Z_t)\varphi(X_t,y)-\int\varphi(x,y)\frac{r(x,y,z)}{f_{y,z}(y,z)}dx\right)K_h(y-Y_t)K_h(z-Z_t)\\ &+\frac{1}{f_{y,z}(y,z)}\frac{1}{T}\sum e_t\varphi(X_t,y)K_h(y-Y_t)K_h(z-Z_t),\end{aligned}$$

where $\{e_t\}$ is the MDS used in the proof Lemma 4. The first term on the RHS is dominated by the deterministic bias that is uniformly $O(h^2)$, and the desired asymptotic distribution is provided by second term. ■

For the integrals w.r.t. A^*, we first define $\widehat{f}_j^*(x,y,z)$ and $f_j^*(x,y,z)$ by $\widehat{m}(x,y,z)\widehat{f}_{z|x,y}(z|x,y)$ and $m(x,y,z)f_{z|x,y}(z|x,y)$ respectively.

Lemma 6 *Under Assumption B.1 - B.6, for each (x,y) in the interior of S, for any $\varphi\in L^2(S)$:*

$$\sqrt{Th_T^{\ell_1+\ell_2}}\int\varphi(y,z)\left(\widehat{f}_j^*(x,y,z)-f_j^*(x,y,z)\right)dz\Rightarrow N\left(0,\Sigma_\varphi^*(x,y)\right)$$

where

$$\Sigma_\varphi^*(x,y)=\frac{\kappa_2(K)^{\ell_1+\ell_2}}{f_{x,y}^2(x,y)}\int\sigma^2(x,y,z)\varphi(y,z)f_{x,y,z}(x,y,z)\,dz,$$

with $\kappa_2(K)=\int K^2(u)\,du$, $\sigma^2(x,y,z)=E\left[e_t^2|X_t=x,Y_t=y,Z_t=z\right]$ and $e_t=R_{jt+1}-m(X_t,Y_t,Z_t)$.

Proof. The proof is almost a symmetry of that found for Lemma 5. We proceed by expressing $\widehat{f}_j^*(x,y,z)$ as the ratio $\frac{\widehat{r}(x,y,z)}{\widehat{f}_{x,y}(x,y)}$,

$$\begin{aligned}\int\varphi(y,z)\left(\widehat{f}_j^*(x,y,z)-f_j^*(x,y,z)\right)dz &= \int\varphi(y,z)\left(\frac{\widehat{r}(x,y,z)}{\widehat{f}_{x,y}(x,y)}-\frac{r(x,y,z)}{f_{x,y}(x,y)}\right)dz\\ &= \frac{1}{\widehat{f}_{x,y}(x,y)}\int\varphi(y,z)\left(\widehat{r}(x,y,z)-\widehat{f}_{x,y}(x,y)\frac{r(x,y,z)}{f_{x,y}(x,y)}\right)dz.\end{aligned}$$

We focus on the integral since $\sup_{x,y}\left|\widehat{f}_{x,y}(x,y)-f_{x,y}(x,y)\right|=o_p(1)$. First, note that

$$\begin{aligned}\int\varphi(y,z)\,\widehat{r}(x,y,z)\,dz &= \frac{1}{T}\sum R_{jt+1}\int\varphi(y,z)\,K_h(x-X_t)\,K_h(y-Y_t)\,K_h(z-Z_t)\,dz\\ &\simeq \frac{1}{T}\sum R_{jt+1}\varphi(y,Z_t)\,K_h(x-X_t)\,K_h(y-Y_t)\,,\end{aligned}$$

therefore

$$\begin{aligned}&\int\varphi(y,z)\left(\widehat{f}_j^*(x,y,z)-f_j^*(x,y,z)\right)dz\\ \simeq\ &\frac{1}{f_{x,y}(x,y)}\frac{1}{T}\sum\left(m(X_t,Y_t,Z_t)\,\varphi(y,Z_t)-\int\varphi(y,z)\frac{r(x,y,z)}{f_{x,y}(x,y)}dz\right)K_h(x-X_t)\,K_h(y-Y_t)\\ &+\frac{1}{f_{x,y}(x,y)}\frac{1}{T}\sum e_t\varphi(y,Z_t)\,K_h(x-X_t)\,K_h(y-Y_t)\,,\end{aligned}$$

where $\{e_t\}$ is the MDS used in the proof Lemma 4 and 5. The first term on the RHS is dominated by the deterministic bias that is uniformly $O(h^2)$, and the desired asymptotic distribution is delivered by second term. ■

Lemma 7 *Under Assumption B.1 - B.6, for each (x,y) in the interior of S, for any $\varphi\in L^2(S)$:*

$$\sqrt{Th_T^{\ell_1+\ell_2}}\int\int\varphi(x',y)\left(\widehat{f}_j(x',y,z')\,\widehat{f}_j^*(x,y,z')-f_j(x',y,z')\,f_j^*(x,y,z')\right)dx'dz'\Rightarrow N\left(0,\Sigma_\varphi^{**}(x,y)\right)$$

where

$$\Sigma_\varphi^{**}(x,y)=\frac{\kappa_2(K)^{\ell_1+\ell_2}}{f_{x,y}^2(x,y)}\int\sigma^2(x,y,z)\left[\int\varphi(x',y)\,f_j(x',y,z)\,dx'\right]^2 f_{x,y,z}(x,y,z)\,dz.$$

with $\kappa_2(K)=\int K^2(u)\,du$, $\sigma^2(x,y,z)=E\left[e_t^2|X_t=x,Y_t=y,Z_t=z\right]$ and $e_t=R_{jt+1}-m(X_t,Y_t,Z_t)$.

Proof.

$$\begin{aligned}&\int\int\varphi(x',y)\left(\widehat{f}_j(x',y,z')\,\widehat{f}_j^*(x,y,z')-f_j(x',y,z')\,f_j^*(x,y,z')\right)dx'dz'\\ \simeq\ &\int\int\varphi(x',y)\,f_j(x',y,z')\left(\widehat{f}_j^*(x,y,z')-f_j^*(x,y,z')\right)dx'dz'\\ &+\int\int\varphi(x',y)\,f_j^*(x,y,z')\left(\widehat{f}_j(x',y,z')-f_j(x',y,z')\right)dx'dz'.\end{aligned}$$

For the first term, as seen in the proof of Lemma 6, it follows immediately that

$$\begin{aligned}&\int\int\varphi(x',y)\,f_j(x',y,z')\left(\widehat{f}_j^*(x,y,z')-f_j^*(x,y,z')\right)dx'dz'\\ =\ &\int\left[\int\varphi(x',y)\,f_j(x',y,z')\,dx'\right]\left(\widehat{f}_j^*(x,y,z')-f_j^*(x,y,z')\right)dz'\\ \simeq\ &\frac{1}{f_{x,y}(x,y)}\frac{1}{T}\sum e_t\left[\int\varphi(x',y)\,f_j(x',y,Z_t)\,dx'\right]K_h(x-X_t)\,K_h(y-Y_t)\,.\end{aligned}$$

For the second term, as seen from proof of Lemma 5, that the inner integral has the following expansion

$$\begin{aligned}
&\int \varphi\left(x', y\right)\left(\widehat{f}_j\left(x', y, z'\right)-f_j\left(x', y, z'\right)\right) dx' \\
\simeq\ & \frac{1}{f_{y,z}(y, z)} \frac{1}{T} \sum e_t \varphi\left(X_t, y\right) K_h\left(y-Y_t\right) K_h\left(z'-Z_t\right),
\end{aligned}$$

therefore

$$\begin{aligned}
&\int\int \varphi\left(x', y\right) f_j^*\left(x, y, z'\right)\left(\widehat{f}_j\left(x', y, z'\right)-f_j\left(x', y, z'\right)\right) dx' dz' \\
\simeq\ & \frac{1}{f_{y,z}(y, z)} \frac{1}{T} \sum e_t \varphi\left(X_t, y\right) K_h\left(y-Y_t\right) \int f_j^*\left(x, y, z'\right) K_h\left(z'-Z_t\right) dz' \\
\simeq\ & \frac{1}{f_{y,z}(y, z)} \frac{1}{T} \sum e_t \varphi\left(X_t, y\right) f_j^*\left(x, y, Z_t\right) K_h\left(y-Y_t\right).
\end{aligned}$$

In sum, we see that the first term is the leading term that determines the pointwise distribution theory. ■

Lemma 8 *Under Assumption B.1 - B.6, for each (y, z) in the interior of S, for any $\varphi, \psi \in L^2(S)$:*

$$\sqrt{T} \int\int\int \varphi(x, y) \psi(y, z)\left(\widehat{f}_j(x, y, z)-f_j(x, y, z)\right) dx dy dz \Rightarrow N\left(0, \Sigma_{\varphi \psi}\right)$$

where

$$\Sigma_{\varphi \psi}=E\left[\sigma^2\left(X_t, Y_t, Z_t\right)\left[\frac{\varphi\left(X_t, Y_t\right) \psi\left(Y_t, Z_t\right)}{f_{y,z}\left(Y_t, Z_t\right)}\right]^2\right].$$

with $\sigma^2(x, y, z)=E\left[e_t^2 | X_t=x, Y_t=y, Z_t=z\right]$ and $e_t=R_{jt+1}-m\left(X_t, Y_t, Z_t\right)$.

Proof. From the proof of Lemma 5, and an additional change of variables and Taylor's expansion, it follows that

$$\begin{aligned}
&\int\int\int \varphi(x, y) \psi(y, z)\left(\widehat{f}_j(x, y, z)-f_j(x, y, z)\right) dx dy dz \\
\simeq\ & \int \psi(y, z) \frac{1}{f_{y,z}(y, z)} \frac{1}{T} \sum e_t \varphi\left(X_t, y\right) K_h\left(y-Y_t\right) K_h\left(z-Z_t\right) dy dz \\
\simeq\ & \frac{1}{T} \sum e_t \frac{\varphi\left(X_t, Y_t\right) \psi\left(Y_t, Z_t\right)}{f_{y,z}\left(Y_t, Z_t\right)},
\end{aligned}$$

where $\{e_t\}$ is a MDS. The proof then follows from the CLT for MDS. ■

Lemma 4 is standard. Lemma 5, 6 and 7 are somewhat similar to Theorem 3.7 of DFG; the intuition these results is that the integrating kernel smoothers can improve their rates of convergence.

Lemma 8 extends the idea further when we integrate out all variables to obtain the parametric rate of convergence.

From the theory of linear operator in Hilbert Spaces (for a review with an econometrics perspective see Carrasco, Florens and Renault, 2007), for a compact operator K:

1. $R(I-K)$ is closed
2. $N(I-K)$ is finite dimensional and $L^2(S) = N(I-K) \oplus R(I-K)$

In our case $K = bA$, and we assume its null space is a one dimensional linear manifold. Define the inner product $\langle \cdot, \cdot \rangle$, inducing the by L^2 norm, $\|\cdot\|_2$, so that $\langle g_1, g_2 \rangle = \int g_1(y,z) g_2(y,z) f(y,z) dydz$, and $\langle \cdot, \cdot \rangle_T$ so that $\langle g_1, g_2 \rangle_T = \int g_1(y,z) g_2(y,z) \widehat{f}(y,z) dydz$ for any $g_1, g_2 \in \mathcal{G}$.

Using implicit function theorem in Banach space, directly extending the results in Magnus (1985), we shall consider the following appropriately smooth mappings for the eigenvalue and eigenfunctions:

$$\begin{aligned} b(A_0) &= b_0, \\ g(A_0) &= g_0, \end{aligned}$$

so that in some neighborhood of A_0

$$\begin{aligned} bAg &= g, \\ \langle g, g \rangle &= 1. \end{aligned} \tag{25}$$

Using a differential argument on the Euler equation in (25), we have

$$dbA_0g_0 + b_0 dAg_0 + b_0A_0dg = dg.$$

Let h_0 be the eigenfunction that corresponds to the unity eigenvalue of the operator $b_0A_0^*$, take an inner product of h_0 with the differentials

$$\begin{aligned} db\langle A_0g_0, h_0\rangle + b_0\langle dAg_0, h_0\rangle + b_0\langle A_0dg, h_0\rangle &= \langle dg, h_0\rangle \\ db\langle g_0, A_0^*h_0\rangle + b_0\langle dAg_0, h_0\rangle + b_0\langle dg, A_0^*h_0\rangle &= \langle dg, h_0\rangle . \end{aligned}$$

Since $b_0A_0^*h_0 = h_0$, we obtain the differential for b as the inner products above reduce to

$$db = -b_0^2 \frac{\langle dAg_0, h_0\rangle}{\langle g_0, h_0\rangle}.$$

This helps us obtain the rate of convergence for $\left(\widehat{b} - b_0\right)$, where $\widehat{b}$ satisfies

$$\begin{aligned} \widehat{b}\widehat{A}\widehat{g} &= \widehat{g}, \\ \langle \widehat{g}, \widehat{g}\rangle_T &= 1, \end{aligned}$$

which satisfies the (theoretical) normalization constraint w.p.a. 1. From Lemma 5, we saw that

$$
\begin{aligned}
&\left(\widehat{A}-A_0\right) g_0 \\
\simeq\ & \frac{1}{f_{y,z}(y,z)}\frac{1}{T}\sum\left(m\left(X_t,Y_t,Z_t\right)g_0\left(X_t,y\right)-\int g_0\left(x,y\right)\frac{r\left(x,y,z\right)}{f_{y,z}\left(y,z\right)}dx\right)K_h\left(y-Y_t\right)K_h\left(z-Z_t\right) \\
&+\frac{1}{f_{y,z}(y,z)}\frac{1}{T}\sum e_t g_0\left(X_t,y\right)K_h\left(y-Y_t\right)K_h\left(z-Z_t\right),
\end{aligned}
$$

therefore the leading terms in $\left(\widehat{b}-b_0\right)$ are

$$
\begin{aligned}
&\left\langle\left(\widehat{A}-A_0\right) g_0, h_0\right\rangle \\
\simeq\ & \int\int\frac{1}{T}\sum\left(m\left(X_t,Y_t,Z_t\right)g_0\left(X_t,y\right)-\int g_0\left(x,y\right)\frac{r\left(x,y,z\right)}{f_{y,z}\left(y,z\right)}dx\right)K_h\left(y-Y_t\right)K_h\left(z-Z_t\right)h_0\left(y,z\right)dydz \\
&+\int\int\frac{1}{T}\sum e_t g_0\left(X_t,y\right)K_h\left(y-Y_t\right)K_h\left(z-Z_t\right)h_0\left(y,z\right)dydz.
\end{aligned}
$$

By the change of variable, ignoring the bias terms, the following sums will converge at a regular parametric rate of $1/\sqrt{T}$ once demeaned:

$$
\begin{aligned}
\left\langle\left(\widehat{A}-A_0\right) g_0, h_0\right\rangle \simeq\ & \frac{1}{T}\sum\left(m\left(X_t,Y_t,Z_t\right)g_0\left(Y_t,Z_t\right)-\int g_0\left(x,Y_t\right)\frac{r\left(x,Y_t,Z_t\right)}{f_{y,z}\left(Y_t,Z_t\right)}dx\right)h_0\left(Y_t,Z_t\right) \\
&+\frac{1}{T}\sum e_t g_0\left(X_t,Y_t\right)h_0\left(Y_t,Z_t\right).
\end{aligned}
$$

Since $|\langle g_0,h_0\rangle|$ is bounded above by 1, due to the triangle inequality, and bounded below by 0 since h_0 cannot lie in the null space of A^*; we conclude that the rate of convergence for $\left(\widehat{b}-b_0\right)$ is $1/\sqrt{T}$, and folloing Lemma 8, with the asymptotic variance $\Sigma_{g_0h_0}$.

Instead of generalizing Magnus' result (Theorem 2) to deal with the eigenfunctions. We obtain the pointwise distribution theory for the eigenfunctions directly from the first order condition of a spectral problem as described below.[2]

For any given eigenvalue b of the operator A, the theoretical problem solves $\min_g \langle (I-bA)g,(I-bA)g\rangle$ subject to $\langle g,g\rangle=1$. The first order condition of the Lagrangean yields a countable number of stationary points:[3]

$$
(I-B)\varphi_j=\eta_j\varphi_j, \tag{26}
$$

[2] Magnus' result relies on a less common normalization condition, requiring the inner products between the eigenvector and any near-by vectors to be 1.

[3] A note on the more formal derivation of Equation (26):

To calculate the Gateux derivative in the direction of h, we need look at

$$
\langle (I-bA)g,(I-bA)g\rangle=\langle g,g\rangle-b\langle g,Ag\rangle-b\langle Ag,g\rangle+b^2\langle Ag,Ag\rangle,
$$

for each eivenvalue-function pair $\left(\eta_j, \varphi_j\right)$ of the linear operator $(I - bA)^* (I - bA) = I - B$, where $B = b(A + A^*) - b^2 A^* A$; we let K^* denote the adjoint of the operator K. Some comments on the operator B. Firstly, B is a self-adjoint compact operator. One particular eigenfunction is of B is g_0, corresponding to the *isolated eigenvalue* 1 (this corresponds to the Lagrange multiplier $\eta = 0$), of which is of our interest. We denote the orthornormal basis from the spectral decomposition of B by $\left\{\varphi_j\right\}$, and their eigenvalues by $\left\{\rho_j\right\}$; w.l.o.g. we let $\varphi_1 = g_0$ and $\rho_1 = 1$. Note that g_0 is the basis for the null space of A^b that coincides with the zero spectral value of A^b. For the specifics of the operators A^* and A^*A, by definition of the adjoint:

$$
\begin{aligned}
\langle Ag, h \rangle &= E\left[E\left[g(X_t, Y_t) R_{jt+1} | Y_t, Z_t\right] h(Y_t, Z_t)\right] \\
&= E\left[g(X_t, Y_t) h(Y_t, Z_t) R_{jt+1}\right] \\
&= E\left[g(X_t, Y_t) E\left[h(Y_t, Z_t) R_{jt+1} | X_t, Y_t\right]\right] \\
&= \langle g, A^* h \rangle,
\end{aligned}
$$

so we have that $A\varphi = E\left[\varphi(X_t, Y_t) R_{jt+1} | Y_t, Z_t\right], A^*\varphi = E\left[\varphi(Y_t, Z_t) R_{jt+1} | X_t, Y_t\right]$ and $A^*A\varphi = E\left[E\left[\varphi(X_t, Y_t) R_{jt+1} | Y_t, Z_t\right] R_{jt+1} | X_t, Y_t\right]$. In our parsimonious notation, recall that

$$
A\varphi(y, z) = \int \varphi(x, y)\, m(x, y, z)\, f_{x|y,z}(x|y, z)\, dx,
$$

we can write the other terms analogously

$$
\begin{aligned}
A^*\varphi(x, y) &= \int \varphi(y, z)\, m(x, y, z)\, f_{z|x,y}(z|x, y)\, dz, \\
A^*A\varphi(x, y) &= \int\int \varphi(x', y)\, m(x', y, z')\, f_{x|y,z}(x'|y, z')\, m(x, y, z')\, f_{z|x,y}(z'|x, y)\, dx' dz'.
\end{aligned}
$$

---

and a similar inner product with $(I - bA)(g + th)$

$$
\begin{aligned}
\langle (I - bA)(g + th), (I - bA)(g + th) \rangle = \; & \langle g, g \rangle - b\langle g, Ag \rangle + t\langle g, h \rangle - bt\langle g, Ah \rangle \\
& -b\langle Ag, g \rangle + b^2\langle Ag, Ag \rangle - bt\langle Ag, h \rangle + b^2 t\langle Ag, Ah \rangle \\
& +t\langle h, g \rangle - bt\langle h, Ag \rangle + t^2\langle h, h \rangle - bt^2\langle h, Ah \rangle \\
& -bt\langle Ah, g \rangle + b^2 t\langle Ah, Ag \rangle - bt^2\langle Ah, h \rangle + b^2 t^2\langle Ah, Ah \rangle.
\end{aligned}
$$

So the Gâteaux derivative of the objective function, taking into account of the normalization constrait that $\langle g, g \rangle = 1$ and denoting the Lagrange multiplier by η, is

$$
\begin{aligned}
& \lim_{t\to 0} \frac{\langle (I - bA)(g + th), (I - bA)(g + th) \rangle - \langle (I - bA) g, (I - bA) g \rangle}{t} - \eta \lim_{t\to 0} \frac{\langle g + th, g + th \rangle - \langle g, g \rangle}{t} \\
= \; & 2\left(\langle g, h \rangle - b(\langle g, Ah \rangle + \langle Ag, h \rangle) + b^2\langle Ag, Ah \rangle - \eta\langle g, h \rangle\right) \\
= \; & 2\left(\langle g, h \rangle - b(\langle A^* g, h \rangle + \langle Ag, h \rangle) + b^2\langle A^*Ag, h \rangle - \eta\langle g, h \rangle\right)
\end{aligned}
$$

The feasible estimator satisfies the analogous conditions when the operators above are replaced by their empirical counterparts; defined by replacing $m, f_{x|y,z}$ and $f_{z|x,y}$ with their nonparametric counterparts. By definition, $\widehat{g}$ satisfies the following conditions

$$\left(I-\widehat{B}\right)\widehat{g} = \widehat{\eta}\widehat{g}, \tag{27}$$

$$\langle\widehat{g},\widehat{g}\rangle_T = 1. \tag{28}$$

Ignoring higher order terms, the main determinants of these equations are

$$(I-B)(\widehat{g}-g_0)-\left(\widehat{B}-B\right)g_0 = \widehat{\eta}g_0 \tag{29}$$

$$-\frac{1}{2}\int g_0^2(y,z)\left(\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right)dydz = \int g_0(y,z)\,f_{y,z}(y,z)\,(\widehat{g}(y,z)-g_0(y,z))\,dydz.$$

Let's denote $-\left(\widehat{B}-B\right)g_0$ by C_T and $\int g_0(y,z)\,f_0(y,z)\,(\widehat{g}(y,z)-g_0(y,z))\,dydz$ by D_T. $\widehat{g}-g_0$ also admits a Fourier expansion with the bases $\{\varphi_j\}$, we have

$$\widehat{g}(y,z)-g_0(y,z)\simeq\sum_{j=1}^{\infty}\gamma_{jT}\varphi_j(y,z),$$

where $\gamma_{jT}=\langle\widehat{g}-g_0,\varphi_j\rangle$ and we can ignore the constant term since $\widehat{g}-g_0$ is asymptotically zero mean. We then write (29) as

$$\sum_{j=2}^{\infty}\gamma_{jT}\left(1-\rho_j\right)\varphi_j+C_T\simeq\widehat{\eta}g_0, \tag{30}$$

since $(I-B)\varphi_j=\left(1-\rho_j\right)\varphi_j$ for $j\geq 1$. Note that we can also express D_T as a Fourier coefficient, $D_T=\langle\widehat{g}-g_0,g_0\rangle=\gamma_{1T}$. Using the definition of the orthornormal basis, by taking inner products of equation (30) w.r.t. $\{\varphi_j\}$, we can then write

$$\widehat{\eta} \simeq \langle C_T,g_0\rangle, \tag{31}$$

$$\widehat{g}(y,z)-g_0(y,z) \simeq D_T g_0(y,z)-\sum_{j=2}^{\infty}\frac{\langle C_T,\varphi_j\rangle}{1-\rho_j}\varphi_j(y,z). \tag{32}$$

This expression for $\widehat{g}-g_0$, in particular, has incorporated the normalization constraint, so we should be able to derive the distribution theory based on the distribution theory of C_T and D_T. We proceed to provide the asymptotic expansion for D_T and $\langle C_T,\varphi_j\rangle$.

Here we outline the derivation of this expansion.

1) $D_T=-\frac{1}{2}\int g_0^2(y,z)\left(\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right)dydz$

$$\begin{aligned}\int g_0^2(y,z)\left(\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right)dydz &= \int g_0^2(y,z)\left(\widehat{f}_{y,z}(y,z)-E\widehat{f}_{y,z}(y,z)\right)dydz\\ &\quad+\int g_0^2(y,z)\left(E\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right)dydz,\end{aligned}$$

under some smoothness assumptions, the usual change of variables and Taylor's expansion yield that $\int g_0^2(y,z)\left(E\widehat{f}_{y,z}(y,z)-f_{y,z}(y,z)\right)dydz = O(h^2)$ uniformly on any compact subset of S (i.e. the support of (Y_t,Z_t)). Similarly we have the following uniform expansion for the stochastic term

$$\begin{aligned}\int g_0^2(y,z)\left(\widehat{f}_{y,z}(y,z)-E\widehat{f}_{y,z}(y,z)\right)dydz &= \int g_0^2(y,z)\left(\frac{1}{T}\sum\begin{pmatrix} K_h(y-Y_t)K_h(z-Z_t) \\ -E\left(K_h(y-Y_t)K_h(z-Z_t)\right)\end{pmatrix}\right)dydz \\ &= \frac{1}{T}\sum\left(g_0^2(Y_t,Z_t)-E\left(g_0^2(Y_t,Z_t)\right)\right)+O\left(h^2\right).\end{aligned}$$

In sum we expect that $D_T = O_p\left(\frac{1}{\sqrt{T}}+h^2\right)$.

2) $C_T = -\left(\widehat{B}-B\right)g_0$, from Lemma 5, 6 and 7, it follows that

$$\begin{aligned}C_T(x,y) \simeq\ & -\frac{b}{f_{x,y}(x,y)}\frac{1}{T}\sum e_t g_0(X_t,\widetilde{y})K_h(\widetilde{y}-Y_t)K_h(\widetilde{z}-Z_t) \\ & -\frac{b}{f_{x,y}(x,y)}\frac{1}{T}\sum e_t g_0(y,Z_t)K_h(x-X_t)K_h(y-Y_t) \\ & +\frac{b^2}{f_{x,y}(x,y)}\frac{1}{T}\sum e_t\left[\int g_0(x',y)f_j(x',y,Z_t)dx'\right]K_h(x-X_t)K_h(y-Y_t),\end{aligned} \tag{33}$$

where $(\widetilde{y},\widetilde{z}) = (x,y)$ such that $(\widetilde{y},\widetilde{z}) \in \mathbb{R}^{\ell_2}\times\mathbb{R}^{\ell_1}$ to coincide with the indexing of an element representing (Y_t,Z_t). Therefore we have that

$$\sqrt{Th_T^{\ell_1+\ell_2}}C_T(x,y) \Rightarrow N\left(0,\Sigma_0(x,y)\right).$$

3) For any φ_j:

$$\begin{aligned}\left\langle C_T,\varphi_j\right\rangle \simeq\ & -b\frac{1}{T}\sum e_t\int\varphi_j(\widetilde{y},\widetilde{z})g_0(X_t,\widetilde{y})K_h(\widetilde{y}-Y_t)K_h(\widetilde{z}-Z_t)d\widetilde{y}d\widetilde{z} \\ & -b\frac{1}{T}\sum e_t\int\varphi_j(x,y)g_0(y,Z_t)K_h(x-X_t)K_h(y-Y_t)dxdy \\ & +b^2\frac{1}{T}\sum e_t\int\varphi_j(x,y)\left[\int g_0(x',y)f_j(x',y,Z_t)dx'\right]K_h(x-X_t)K_h(y-Y_t)dxdy \\ \simeq\ & -b\frac{1}{T}\sum e_t\varphi_j(Y_t,Z_t)g_0(X_t,Y_t)-b\frac{1}{T}\sum e_t\varphi_j(X_t,Y_t)g_0(Y_t,Z_t) \\ & +b^2\frac{1}{T}\sum e_t\varphi_j(X_t,Y_t)\left[\int g_0(x',Y_t)f_j(x',Y_t,Z_t)dx'\right]+O_p\left(h^2\right),\end{aligned}$$

so we can deduce that $\left\langle C_T,\varphi_j\right\rangle = O_p\left(\frac{1}{\sqrt{T}}+h^2\right)$. In particular, this means that $\left\langle C_T,\varphi_1\right\rangle = \widehat{\eta} = O_p\left(\frac{1}{\sqrt{T}}+h^2\right)$.

4) Now, ignoring the smaller order terms from equation (29)

$$(I-B)(\widehat{g}-g) \simeq \left(\widehat{B}-B\right)g_0,$$

and transform the equation by the Moore-Penrose pseudoinverse of $(I-B)$, denoted by $(I-B)^{\dagger}$,

$$(I-B)^{\dagger}(I-B)(\widehat{g}-g) \simeq (I-B)^{\dagger}\left(\widehat{B}-B\right) g_0.$$

By construction of the pseudoinverse, from the singular value decomposition, $(I-B)^{\dagger}(I-B)\varphi_j = \varphi_j$ when $j>1$, and it takes value 0 when $j=1$ since $\varphi_1 (= g_0)$ spans $N(I-B)$. Using the Fourier expansion

$$\begin{aligned} \widehat{g}-g_0 &\simeq \gamma_{1T} g_0 + (I-B)^{\dagger}(I-B)(\widehat{g}-g_0) \\ &\simeq (I-B)^{\dagger}\left(\widehat{B}-B\right) g_0, \end{aligned}$$

where we drop $\gamma_{1T} (= D_T)$ in the last approximation as we showed previously that $D_T = O_p\left(\frac{1}{\sqrt{T}} + h^2\right)$. First consider $\left(\widehat{B}-B\right) g_0 = -C_T$, recall that

$$\sqrt{T h_T^{\ell_1+\ell_2}}\, C_T(x,y) \Rightarrow N\left(0, \Sigma_0(x,y)\right).$$

In particular $\Sigma_0(x,y)$ takes the following form

$$\Sigma_0(x,y) = \frac{b_0^2 \kappa (K)^{\ell_1+\ell_2}}{f_{x,y}^2(x,y)} \begin{pmatrix} \int \sigma^2(x',\widetilde{y},\widetilde{z})\, g_0^2(x',\widetilde{y})\, f_{x,y,z}(x',\widetilde{y},\widetilde{z})\, dx' \\ +\int \sigma^2(x,y,z')\, g_0^2(y,z')\, f_{x,y,z}(x,y,z')\, dz' \\ +b^2 \int \sigma^2(x,y,z')\left[\int g_0(x'',y)\, f_j(x'',y,z')\, dx''\right]^2 f_{x,y,z}(x,y,z')\, dz \end{pmatrix}. \tag{34}$$

This follows from computing the variance of

$$\sqrt{T h_T^{\ell_1+\ell_2}}\, C_T(x,y) \simeq -\frac{b_0}{f_{x,y}(x,y)} \frac{\sqrt{T h_T^{\ell_1+\ell_2}}}{T} \sum e_t \begin{pmatrix} g_0(X_t,\widetilde{y})\, K_h(\widetilde{y}-Y_t)\, K_h(\widetilde{z}-Z_t) \\ +g_0(y,Z_t)\, K_h(x-X_t)\, K_h(y-Y_t) \\ -b\left[\int g_0(x',y)\, f_j(x',y,Z_t)\, dx'\right] K_h(x-X_t)\, K_h(y-Y_t) \end{pmatrix}$$

by the MDS property, we only need to focus on

$$E\left[\sigma^2(X_t,Y_t,Z_t)\begin{pmatrix} g_0(X_t,\widetilde{y})\, K_h(\widetilde{y}-Y_t)\, K_h(\widetilde{z}-Z_t) \\ +g_0(y,Z_t)\, K_h(x-X_t)\, K_h(y-Y_t) \\ -b_0\left[\int g_0(x',y)\, f_j(x',y,Z_t)\, dx'\right] K_h(x-X_t)\, K_h(y-Y_t) \end{pmatrix}^2\right].$$

we multiply out the square in the bracket and compute the following integrals seperately:

$$\int\int\int \sigma^2(x',y',z') \begin{pmatrix} g_0(x',\widetilde{y}) K_h(\widetilde{y}-y') K_h(\widetilde{z}-z') \\ +g_0(y,z') K_h(x-x') K_h(y-y') \\ -b_0\left[\int g_0(x'',y) f_j(x'',y,z') dx''\right] K_h(x-x') K_h(y-y') \end{pmatrix}^2 f_{x,y,z}(x',y',z') dx'd$$

$$\begin{aligned}
= & \int\int\int \sigma^2(x',y',z') g_0^2(x',\widetilde{y}) K_h^2(\widetilde{y}-y') K_h^2(\widetilde{z}-z') f_{x,y,z}(x',y',z') dx'dy' \\
& +\int\int\int \sigma^2(x',y',z') g_0^2(y,z') K_h^2(x-x') K_h^2(y-y') f_{x,y,z}(x',y',z') dx'dy'dz' \\
& +b_0^2\int\int\int \sigma^2(x',y',z') \left[\int g_0(x'',y) f_j(x'',y,z') dx''\right]^2 K_h^2(x-x') K_h^2(y-y') f_{x,y,z}(x',y',z') dx'dy'dz' \\
& +2\int\int\int \sigma^2(x',y',z') \left(\begin{array}{c} g_0(x',\widetilde{y}) g_0(y,z') K_h(\widetilde{y}-y') K_h(\widetilde{z}-z') K_h(x-x') K_h(y-y') \\ -b_0 g_0(x',\widetilde{y}) \left[\int g_0(x'',y) f_j(x'',y,z') dx''\right] K_h(\widetilde{y}-y') K_h(\widetilde{z}-z') K_h(x-x') \\ -b_0 g_0(y,z') \left[\int g_0(x'',y) f_j(x'',y,z') dx''\right] K_h(x-x') K_h(y-y') K_h(x-x') \end{array} \right.
\end{aligned}$$

Note that we collect the cross-terms together. It is easy to see that the integral of these terms will be negligible since it involves integration over more variables than the square terms, and the leading terms of these integrals is the following sum:

$$\begin{aligned}
= & \frac{\kappa(K)^{\ell_1+\ell_2}}{h_T^{\ell_1+\ell_2}} \int \sigma^2(x',\widetilde{y},\widetilde{z}) g_0^2(x',\widetilde{y}) f_{x,y,z}(x',\widetilde{y},\widetilde{z}) dx' \\
& +\frac{\kappa(K)^{\ell_1+\ell_2}}{h_T^{\ell_1+\ell_2}} \int \sigma^2(x,y,z') g_0^2(y,z') f_{x,y,z}(x,y,z') dz' \\
& +b_0^2\frac{\kappa(K)^{\ell_1+\ell_2}}{h_T^{\ell_1+\ell_2}} \int \sigma^2(x,y,z') \left[\int g_0(x'',y) f_j(x'',y,z') dx''\right]^2 f_{x,y,z}(x,y,z') dz + o\left(\frac{1}{h_T^{\ell_1+\ell_2}}\right).
\end{aligned}$$

So we expect that

$$\sqrt{Th_T^{\ell_1+\ell_2}}\left(\widehat{g}(x,y)-g_0(x,y)\right) = (I-B)^\dagger\left[\sqrt{Th_T^{\ell_1+\ell_2}} C_T(x,y)\right] + o_p(1),$$

as $T\to\infty$.

However, in practice, we do not know what is the true value of b. We replace the unknown b by an estimated one. In such case, all of the previous arguments still hold with little modifications since the eigenfunctions will necessarily satisfy the following feasible first order condition

$$\left(I-\widetilde{B}\right)\widehat{g} = \widehat{\eta}\widehat{g},$$

and the fact that $\widehat{b}$ converge to the true discounting factor at a parametric rate. So that $(\widehat{\eta},\widehat{g})$ is an eivenvalue-function pair of the linear operator $\left(I-\widetilde{B}\right)$, where $\widetilde{B} = \widehat{b}\left(\widehat{A}+\widehat{A}^*\right) - \widehat{b}^2\widehat{A}^*\widehat{A}$. The distribution theory for functions satisfying such condition is outlined above for fixed b. When b is random, we have

$$\begin{aligned}\left(I-\widetilde{B}\right)\widehat{g} &= \left(I-\widehat{b}\left(\widehat{A}+\widehat{A}^{*}\right)-\widehat{b}^{2}\widehat{A}^{*}\widehat{A}\right)\widehat{g} \\ &= \left(I-\widehat{b}\left(\widehat{A}+\widehat{A}^{*}\right)-\widehat{b}^{2}\widehat{A}^{*}\widehat{A}\right)\widehat{g}+\left(\left(\widehat{b}-b_{0}\right)\left(\widehat{A}+\widehat{A}^{*}\right)-\left(\widehat{b}^{2}-b_{0}^{2}\right)\widehat{A}^{*}\widehat{A}\right)\widehat{g} \\ &= \left(I-\widehat{B}\right)\widehat{g}+O_{p}\left(\left|\widehat{b}-b_{0}\right|\right),\end{aligned}$$

where $\widehat{B}$ is defined previously (for $b = b_0$). It then follows that (assume we undersmooth and ignore the biases), the asymptotic distribution is determined by

$$\widehat{g}-g_{0}\simeq(I-B)^{\dagger}\left(\widehat{B}-B\right)g_{0},$$

which also coincides with the finite dimensional formula of Magnus (1985).

References

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

[2] Anatolyev, S. (1999), "Nonparametric estimation of nonlinear rational expectation models," Economics Letters, 62, 1-6.

[3] 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.

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

[5] 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.

[6] Chen, X., V. Chernozhukov, S. Lee, and W. Newey (2010), "Identification in Semiparametric and Nonparametric Conditional Moment Models," unpublished manuscript.

[7] 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.

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

[9] Chen, X. and M. Reiss (2010), "On rate optimality for ill-posed inverse problems in econometrics," Econometric Theory, forthcoming.

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

[11] Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2010), "Nonparametric Instrumental Regression," unpublished manuscript.

[12] 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.

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

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

[15] Gallant, A. R. and G. Tauchen (1989), "Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications," Econometrica, 57, 1091-1120.

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

[17] Hall, P., and , J. L. Horowitz, (2005) "Nonparametric methods for inference in the presence of instrumental variables" Annals of Statistics, 33, 2904–2929.

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

[19] Hansen, L. P. and K. J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," 50, 1269-1286.

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

[21] Magnus, J. R. (1985): "On Differentiating Eigenvalues and Eigenvectors," *Econometric Theory*, 1, 179-191.

[22] Manski, C.F. (2003), "Identification Problems in the Social Sciences and Everyday Life," Southern Economic Journal, 70, 11-21.

[23] McManus, D. A. (1992), "How Common is Identification in Parametric Models?" Journal of Econometrics, 53, 5-23.

[24] Rothenberg, T. J. (1971), "Identification in Parametric Models," Econometrica, 39, 577-591.

[25] Wang, Q. and Phillips, P. C. B. (2009): "Structural Nonparametric Cointegrating Regression," *Econometrica*, **77**, 1901–1948

Table 1: Eigenvalue discount factors by bandwidth

Risk Free Asset $j=1$				Risky Asset $J=2$			
h	$h/2$	$2h$	$4h$	h	$h/2$	$2h$	$4h$
0.9693	**0.7650**	**0.9544**	**0.8824**	**0.8962**	**0.7938**	0.9954	**0.8349**
0.9107	**0.7650**	**0.9544**	**0.8824**	**0.8962**	**0.7938**	0.9954	**0.8349**
0.9107	0.6909	0.9076	0.7477	0.8809	0.6493	**0.9455**	0.6743
0.8268	0.5229	0.9076	0.7477	0.8809	0.5943	**0.9455**	0.5570
0.8268	0.5229	0.8660	0.6330	0.8085	0.5943	0.8937	0.5570
0.7679	0.4760	0.8660	0.4250	0.8085	0.4881	0.8937	0.4389
0.7679	0.4760	0.8016	0.4250	0.7758	0.4881	0.8209	0.4389
0.6261	0.4503	0.8016	0.4096	0.7758	0.4505	0.8209	0.4339
0.6261	0.4503	0.7716	0.4096	0.6076	0.4505	0.7871	0.4339
0.5868	0.3185	0.7716	0.3988	0.6076	0.408	0.7871	0.2737

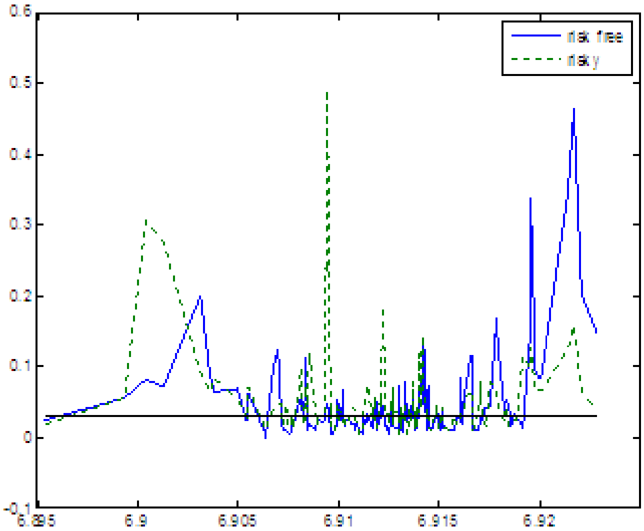

Figure 1: Estimates of the Marginal Utilities (bandwidth = h)

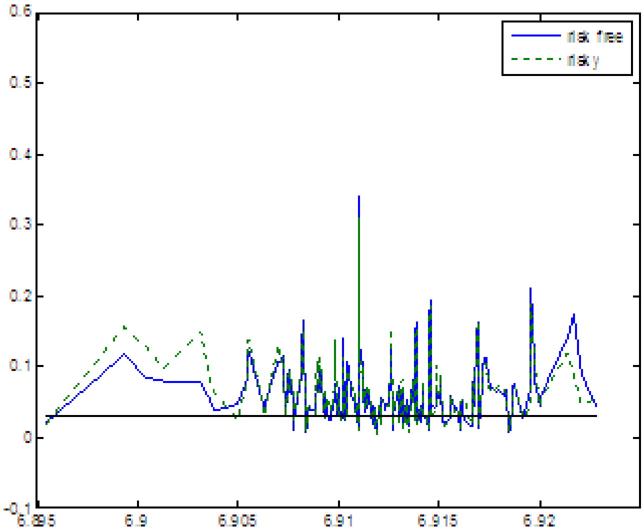

Figure 2: Estimates of the Marginal Utilities (bandwidth = h/2)

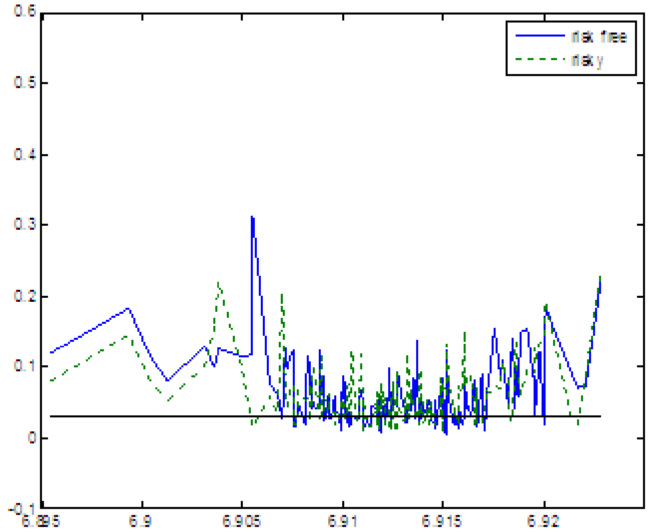

Figure 3: Estimates of the Marginal Utilities (bandwidth = 2h)

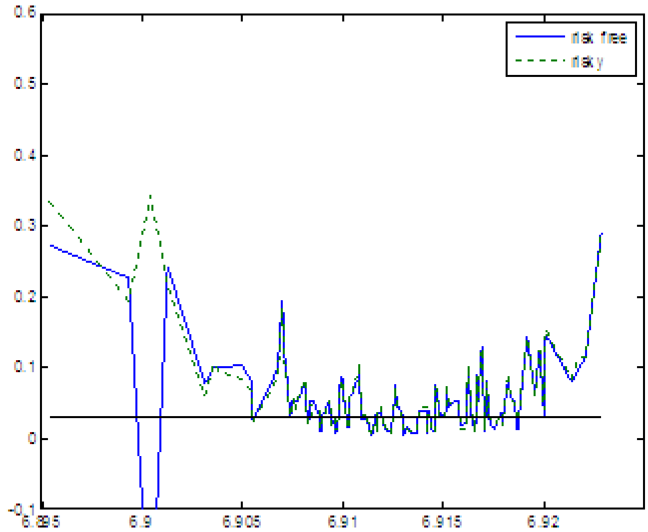

Figure 4: Estimates of the Marginal Utilities (bandwidth = 4h)