Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications
Author(s): A. Ronald Gallant and George Tauchen
Reviewed work(s):
Source: *Econometrica*, Vol. 57, No. 5 (Sep., 1989), pp. 1091-1120
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1913624
Accessed: 05/12/2012 16:46

Econometrica, Vol. 57, No. 5 (September, 1989), 1091–1120

SEMINONPARAMETRIC ESTIMATION OF CONDITIONALLY CONSTRAINED HETEROGENEOUS PROCESSES: ASSET PRICING APPLICATIONS[1]

BY A. RONALD GALLANT AND GEORGE TAUCHEN

The overidentifying restrictions of the intertemporal capital asset pricing model are usually rejected when tested using data on consumption growth and asset returns, particularly when additively separable, constant relative risk utility is attributed to the representative agent. This article investigates the extent to which specification error can explain these rejections. The empirical strategy is limited information maximum likelihood in conjunction with seminonparametric (expanding parameter space) representations for both the law of motion and utility. We find that consumption growth and asset returns display conditional heterogeneity, but this fact does not account for rejection of the overidentifying restrictions as might be anticipated from the work of Hansen, Singleton, and others using generalized method of moments methods. We also find that expansion of the parameter space in the direction of nonseparable utility causes the overidentifying restrictions to be accepted. Our estimation strategy provides information on the manner in which the restrictions distort the law of motion. In particular, imposition of additively separable, constant relative risk aversion utility causes the conditional variance of consumption growth to be overpredicted, the conditional covariance of asset returns with consumption growth to be overpredicted, and an equity premium. Imposition of nonseparable seminonparametric utility causes distortion in these same directions, though the distortions are much smaller which is consistent with the outcomes of the tests of the restrictions.

KEYWORDS: Seminonparametric, nonparametric, Hermite expansions, conditional moment restrictions, asset pricing, utility.

1. INTRODUCTION

THE EMPIRICAL FORCE of the intertemporal capital asset pricing model (I-CAPM) lies in the restrictions it places across the time paths of consumption and asset returns. In stationary environments, the statistical properties of the time paths are completely characterized by their law of motion, which is the one-step ahead joint probability density of consumption and asset returns. The restrictions, in essence, dictate that the series must co-vary in such a way that the product of a suitably defined marginal rate of substitution of consumption and each asset return has conditional mean equal to unity. Restrictions of this form are conditional moment restrictions that constrain the conditional means of nonlinear functions of the data. Equilibrium asset pricing relations leading to such conditional moment restrictions have been deduced in various contexts by Lucas (1978), Breeden (1979), Hansen and Singleton (1982, 1983), and many others. A variety of tests of the restrictions under different auxiliary assumptions about the law of motion or the functional form of the marginal rate of substitution have been reported in the literature. Generally speaking, the tests come out the way

[1] This research was supported by National Science Foundation Grants SES-8207362, SES-8507829, SES-8520244, North Carolina Agricultural Experiment Station Projects NCO-5593, NCO-3879, and the PAMS Foundation. We wish to thank the referees, the co-editor, and many seminar participants for helpful comments.

they did in Hansen and Singleton (1982, 1983), who uncover evidence against the model.

In this paper we address several empirical issues concerning the statistical properties of consumption and asset returns and the restrictions placed on their law of motion by the intertemporal capital asset pricing model. The empirical work begins with an examination of the characteristics of the law of motion itself, with the primary objective being to determine the extent to which it deviates from the Gaussian vector autoregressive (VAR) model. The effort is constructive, in the sense that it provides a consistent estimate of the law of motion when the process is non-Gaussian and conditionally heterogeneous as we find to be the case. The empirical work then proceeds to estimation of the law of motion subject to the conditional moment restrictions from the I-CAPM. The version of the model we test entails minimal assumptions on both the form of the utility function and the stochastic environment. One topic of particular interest is the nature of intertemporal nonseparability of the utility function, which is an aspect of utility that several authors have recently argued is important for understanding the co-movements of consumption and returns. Another is the impact that various assumptions about the intertemporal utility function have on the conditional first and second moment properties of consumption and returns. Overall, our empirical work can be viewed as an effort to understand the sensitivity of empirical conclusions concerning the I-CAPM to assumptions about the dynamic law of motion of the observables and the intertemporal utility function.

The estimation strategy we use is seminonparametric (SNP), which is an approach that applies conventional estimation and testing to models derived from series expansions. The strategy provides a means for making inferences without imposing restrictive auxiliary assumptions that do not follow directly from theory. In our work, the components of the model that are not specified by the theory are the law of motion of the observables and the utility function of the representative agent. For the law of motion, we employ an SNP model, which is a truncated Hermite expansion where the leading term is a Gaussian VAR and higher order terms accommodate both deviations from Gaussianity and conditional heterogeneity. For the utility function, we employ a seminonparametric expansion where the leading term is the familiar constant relative risk aversion (CRR) utility function and the higher order terms capture intertemporal nonseparabilities of nondurable consumption. Our general estimation strategy, then, is the seminonparametric extension of the one used by Hansen and Singleton (1983), who employ a Gaussian VAR model for the law of motion and CRR utility. The impetus for our developing this strategy was the remarks of Hansen (1986).

The remainder of the paper is organized as follows. Section 2 develops the basic theory of SNP models for multiple time series. Section 3 presents the empirical results from fitting SNP models to postwar data on consumption and returns. None of the fits discussed in Section 3 are subject to side constraints. Thus, readers whose main interest is SNP time series models and the mechanics of fitting them, and who might have only a subsidiary interest in asset pricing models, can confine attention to these two sections. Section 4 goes on to develop the conditional moment restrictions from the intertemporal asset pricing model and also presents the numerical methods used for estimation subject to the constraints. Section 5 reports the findings from the constrained estimation. Section 6 contains the concluding remarks, which are mainly an overall summary of our empirical results.

2. SEMINONPARAMETRIC MODELS: THEORY

The prefix semi means half and the term seminonparametric (SNP), coined in Elbadawi, Gallant, and Souza (1983), is intended to convey the notion that SNP procedures are halfway between parametric and nonparametric inference procedures. The method consists of applying classical parametric estimation and inference procedures to models derived from truncated series expansions. Because one states precisely what one will do when specification error is detected, namely increase the truncation point, one can prove that these procedures have nonparametric properties. Nothing is lost, however, if the results are viewed as classical finite-dimensional inferences that have been subjected to a sensitivity analysis. We prefer to regard the empirical results as nonparametric but the reader is at liberty to regard them as parametric.

We obtain our method by applying the results of Gallant and Nychka (1987). To do so, the problem must be structured so that the likelihood is a functional defined over a space of positive valued functions whose domain is a finite dimensional Euclidean space. Denote the space by $\mathscr{H}$ and a typical element by h. Assumptions that will produce this structure are: (i) the data $\{y_t\}_{t=-L+1}^{n}$ are a realization from a stationary time series $\{y_t\}_{t=-\infty}^{\infty}$ where each y_t is a vector of length M; and (ii) the conditional distribution of y_t given the entire past depends only on a finite number L of lagged values of y_t. Denoting these lagged values by

$$x_{t-1} = (y'_{t-L}, y'_{t-L+1}, \dots, y'_{t-1})',$$

which is a vector of length $M \cdot L$, the likelihood can be written in terms of the joint density $h(y_t, x_{t-1})$ of y_t and x_{t-1} as

$$\left[\prod_{t=1}^{n} h(y_t | x_{t-1})\right] \int h(y, x_0)\, dy$$

where

$$h(y_t|x_{t-1}) = h(y_t, x_{t-1}) \Big/ \int h(y, x_{t-1})\, dy.$$

This likelihood has the appropriate form: it is a functional in h and h has a finite number of arguments.

Next, following Gallant and Nychka, h is approximated by a truncated Hermite expansion. The truncated expansion is the SNP model. It replaces h in the likelihood and its parameters are estimated by maximizing the likelihood. Subject to regularity conditions that we will discuss below, the conditional density $h(y|x)$ that corresponds to h in $\mathscr{H}$ is estimated consistently provided the truncation points grows (either adaptively or deterministically) with the sample size.

To follow this recipe exactly, one must set forth the Hermite expansion of h and apply various algebraic simplifications to put it in a tractable form. Doing so will produce the SNP model that we use but will not provide an understanding of what sorts of processes it can approximate well. Instead, we shall present an intuitive derivation of the SNP model that does contribute to understanding.

Observe that in most applications, notably prediction of y_t given the past, it is enough to know the conditional density $h(y_t|x_{t-1})$ and knowledge of the joint density $h(y_t, x_{t-1})$ is not required. This includes determination of the conditional density by the method of maximum likelihood because the maximum likelihood estimate is obtained by minimizing

$$(-1/n)\sum_{t=1}^{n} \ln h(y_t|x_{t-1}) + (-1/n)\ln \int h(y, x_0)\, dy$$

and the term $(-1/n)\ln \int h(y, x_0)\, dy$ is negligible in large samples. Also note that, while the density $h(y_t|x_{t-1})$ is time invariant under the two assumptions above, there is no restriction as to how the function $h(y|x)$ depends on x: the process $\{y_t\}$ can exhibit any sort of conditional heterogeneity under these two assumptions.

A natural approach in modelling such a process is to take a linear function $b_0 + Bx_{t-1}$ of the past as the location parameter of y_t, to scale by an upper triangular matrix R to obtain a standardized residual

$$z_t = R^{-1}(y_t - b_0 - Bx_{t-1}),$$

and then get a conditional density from some parent density $f(z)$ by putting

$$h(y_t|x_{t-1}) = f\left[R^{-1}(y_t - b_0 - Bx_{t-1})\right]/\det(R).$$

The shape characteristics of $f(z)$ will determine the distribution of y_t given x_t. For instance, if $f(z)$ were taken as the standard (multivariate) Gaussian density, denoted hereafter as either $\varphi(z)$ or $n_M(z; 0, I)$, the result would be a Gaussian, vector autoregression (VAR). How can $f(z)$ be chosen without imposing a potentially erroneous a priori shape on the fit? One needs a good general purpose density that can accommodate any shape, especially the Gaussian as it is most likely a priori.

Using a Hermite polynomial as a general approximation to a density function is a long established tradition in statistics; examples are Gram-Charlier and Edgeworth expansions. The form of a Hermite expansion is a polynomial in z times the standard Gaussian density. This is its major strength; the leading term of the expansion is Gaussian; higher order terms accommodate deviations from Gaussianity, if any. To insure positivity we shall square the polynomial part, and to insure that the density integrates to one we shall divide by the integral over

$\mathbb{R}^M$. The result is

$$f(z) = \left[\sum_{|\alpha|=0}^{K_z} a_\alpha z^\alpha\right]^2 \varphi(z) \Big/ \int \left[\sum_{|\alpha|=0}^{K_z} a_\alpha u^\alpha\right]^2 \varphi(u)\, du$$

where z^α maps the multi-index

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)'$$

into the monomial

$$z^\alpha = \prod_{i=1}^{M} (z_i)^{\alpha_i}$$

of degree

$$|\alpha| = \sum_{i=1}^{M} |\alpha_i|.$$

This density will generate a Gaussian VAR if K_z is put to zero and will accommodate arbitrary shape departures from the Gaussian VAR for large enough K_z. However, this form of the density cannot accommodate conditional heteroskedasticity or more general, nonlinear conditional shape variation with x_{t-1}.

The obvious remedy is to let the coefficients a_α be polynomials in x by writing

$$a_\alpha(x) = \sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta$$

where

$$\beta = (\beta_1, \beta_2, \dots, \beta_{ML})',$$

$$|\beta| = \sum_{i=1}^{ML} |\beta_i|,$$

$$x^\beta = \prod_{i=1}^{ML} (x_i)^{\beta_i}.$$

With this modification, the conditional density of a scaled residual given x_{t-1} is

$$f_K(z_t | x_{t-1}) = \frac{\left[\sum_{|\alpha|=0}^{K_z} a_\alpha(x_{t-1}) z_t^\alpha\right]^2 \varphi(z_t)}{\int \left[\sum_{|\alpha|=0}^{K_z} a_\alpha(x_{t-1}) u^\alpha\right]^2 \varphi(u)\, du}$$

with $K = (K_z, K_x)$; the conditional density of y_t given x_{t-1} is

$$h_K(y_t | x_{t-1}) = f_K\left[R^{-1}(y_t - b_0 - Bx_{t-1}) | x_{t-1}\right] / \det(R).$$

When we wish to indicate explicitly that this density depends on the parameters $a_{\alpha\beta}$, b_0, B, and R, we shall write $h_K(y_t|x_{t-1}, \theta)$. To be specific, θ is a vector of length p_θ which has as its leading elements the polynomial coefficients $a_{\alpha\beta}$, the VAR location parameters b_0 and $\text{vec}(B)$ as its next group of elements, and the upper triangle of the VAR scale parameter R stored columnwise as its last group of elements. If p_z is the number of monomials in a polynomial of degree K_z defined on $\mathbb{R}^M$ and p_x is the number of monomials in a polynomial of degree K_x on $\mathbb{R}^{ML}$, then the length of θ is $p_\theta = p_z \cdot p_x + M + M \cdot M \cdot L + M(M+1)/2$.

By increasing K, that is, by increasing K_z and K_x simultaneously, an SNP model can be made to approximate the conditional density corresponding to an h in $\mathscr{H}$ arbitrarily accurately. A mathematically precise description of $\mathscr{H}$ is in Gallant and Nychka (1987). Qualitatively, $\mathscr{H}$ contains distributions with fat t-like tails such as $h(y, x) \propto (1 + y'y + x'x)^{-\delta}$ for $\delta > (M + M \cdot L)/2$, the Gaussian, and distributions with tails that are thinner than the Gaussian such as $h(y, x) \propto \exp[-(y'y + x'x)^{\Delta}]$ for $1 < \Delta < \delta - 1$. Off the tails, that is, on a ball with finite radius that contains all but a small fraction of the probability mass of the density, any sort of skewness, kurtosis, etc. is permitted. What is ruled out are violently oscillatory density functions.

An SNP procedure is an expanding parameter space or sieve method (Grenander (1981), Severini and Wong (1987)). A distribution theory for SNP inference with cross-sectional data has been developed in the following papers: Gallant (1982), Eastwood and Gallant (1987), and Andrews (1987). These papers set forth deterministic and adaptive rules relating the truncation point to the sample size such that classical, finite dimensional hypothesis tests and confidence statements actually achieve their nominal α-level in large samples. Whether or not these results and methods of proof extend to time series data is an open question.

The best known of the various time series models that try to accommodate conditional heterogeneity is the ARCH model (Engle, 1982). Since the stationary distribution of the ARCH model is not known in closed form, we cannot say definitely that the ARCH model is a member of $\mathscr{H}$. It is in a qualitative sense since the stationary distribution of the ARCH model has fat tails and only a finite number of moments which is qualitatively like the t distribution. Conditionally, ARCH models have a variance that is a polynomial in a finite number of lags, the same as the SNP model. These remarks would lead one to expect that we can approximate the conditional density of an ARCH model to within arbitrary accuracy for large K. For large L, the same may be true for GARCH models. GARCH models have a variance that is a polynomial in an infinite number of lags. As yet, we do not know whether it would be possible to let the lag length of an SNP model grow with sample size and retain the consistency result.

3. SEMINONPARAMETRIC MODELS: EMPIRICAL RESULTS

The data used to fit the model are real, monthly: per capita consumption of nondurables and services, value weighted NYSE returns, and Treasury bill returns. We used two data sets: The first covers the years 1959 to 1978 inclusive and is identical to that used by Hansen and Singleton (1982, 1983, 1984); these data are published in Gallant (1987, Chapter 6) together with details regarding the timing conventions and transformations used to get per capita consumption and real returns. The second covers the years 1959 to 1984 inclusive and was constructed from these sources: real ($1972) and current dollar consumption of nondurables and services, Citibase (1983, Series GMCN72, GMCS72, GMCN, GMCS, 1959.01–1983.12) and Department of Commerce (1984, 1984.01–1984.12); the implicit deflator is the ratio of current to real dollar consumption of nondurables and services; population, Citibase (1983, Series POP, 1959.01–1983.12) and Bureau of the Census (1985, 1984.01–1984.12); Treasury bill returns, Ibbotson Associates (1985, U.S. Treasury Bill Returns, Exhibit B-9); total value weighted market returns on NYSE securities, CRSP (1986). Transformations and timing conventions in the second data set are the same as the first. The two data sets differ little in the overlapping years; the main discrepancy is due to a reversion of the monthly population series to reflect new information from the 1980 census.

The per capita consumption series was converted to a consumption growth series by dividing each observation by its predecessor; this reduces series length by one. In addition, the first four initial values were not used in any fits, save those to produce Table I, to insure comparability of likelihoods having differing lag lengths. Fits using the 1959–1978 data set are based on 235 observations and fits using the 1959–1984 data set on 307 observations. The three series, consumption growth, stock returns, and bill returns are denoted as *CG*, *SR*, and *BR* respectively in the text, figures, and tables; natural logarithms of each are denoted as *LCG*, *LSR*, and *LBR*.

Fitting VAR specifications to the series $y_t = (LBR, LCG)$ over 1959–1978 testing at conventional significance levels, a one-lag VAR specification is rejected in favor of a two-lag, a two-lag is not rejected in favor of a three- or four-lag; the same is true for the series $y_t = (LSR, LCG)$. Various descriptive statistics computed from two-lag VAR residuals for $y_t = (LBR, LCG)$ and $y_t = (LSR, LCG)$ are displayed in Table I. There are departures from a Gaussian VAR specification as indicated by the significant Kolmogorov-Smirnov statistics. Normal (or *Q-Q*) plots of these residuals indicate that consumption growth VAR residuals have

TABLE I

DESCRIPTIVE STATISTICS FOR TWO LAG VAR RESIDUALS, 1959–1978

Fitted Series	$y_t = (LBR, LCG)$		$y_t = (LSR, LCG)$	
Residual	*LBR*	*LCG*	*LSR*	*LCG*
Skewness	−0.6824	0.0417	−0.1604	0.0570
Excess Kurtosis	3.2356	0.4697	1.0491	0.4713
Kolmogorov-Smirnov				
statistic	0.0701	0.0420	0.0684	0.0657
p-value	< .01	> .15	< .01	0.014

initially heavy then thin left tails and heavy right tails relative to the Gaussian distribution; both tails of the two returns series are heavy. The cause of these departures from a VAR structure with Gaussian innovations could be a non-Gaussian innovation distribution, a location parameter that depends nonlinearly on the past, conditional heteroskedasticity, some other type of conditional heterogeneity, or a mixture of these factors present at the same time. These are the sorts of departures from a Gaussian VAR specification that an SNP specification is designed to accommodate.

Recall that, for a SNP of dimension M and degree $K=(K_z, K_x)$ in L lags, the conditional density of $y_t=(y_{1t},\ldots, y_{Mt})'$ given $x_{t-1}=(y'_{t-L},\ldots, y'_{t-1})'$ is

$$h_K(y_t|x_{t-1},\theta)$$

$$=\frac{\left[\sum_{|\alpha|=0}^{K_z} a_\alpha(x_{t-1})R^{-1}(y_t-b_0-Bx_{t-1})^\alpha\right]^2 n_M(y_t|b_0+Bx_{t-1}, RR')}{\int\left[\sum_{|\alpha|=0}^{K_z} a_\alpha(x_{t-1})u^\alpha\right]^2 \varphi(u)\,du},$$

$$a_\alpha(x)=\sum_{|\beta|=0}^{K_x} a_{\alpha\beta}x_{t-1}^\beta$$

with θ estimated by $\hat{\theta}$ that minimizes the sample objective function

$$s_n(\theta)=(-1/n)\sum_{t=1}^{n}\ln h_K(y_t|x_{t-1},\theta).$$

In this section, and in Section 5, this minimization was carried out using NPSOL from the Stanford Systems Optimization Laboratory (Gill, Murray, Saunders, and Wright (1983)).[2] NPSOL computes an unconstrained optimum, or a linearly constrained optimum, by minimizing successive quasi-Newton quadratic approximations to the objective function using a quadratic programming routine to find downhill directions; NPSOL computes a nonlinearly constrained optimum by substituting an augmented Lagrangian function for the objective function. We have found that proper scaling is essential in computations to avoid cases where extremely large or small values of the polynomial part of the conditional density are required to compensate for extremely small or large values of the exponential part. Specifically, the computations reported below are based on the transformed data

$$\tilde{y}_t=S^{-1/2}(y_t-\bar{y}),$$

[2] NPSOL is available from the Office of Technology Licensing, 350 Cambridge Avenue, Suite 250, Palo Alto, CA 94306, U.S.A.

where $S=(1/n)\Sigma_t(y_t-\bar{y})(y_t-\bar{y})'$, $\bar{y}=(1/n)\Sigma_t y_t$, and $S^{-1/2}$ denotes the Cholesky factorization of the inverse of S; the sum is over all observations: 239 in the 1959–1978 data set, 311 in the 1959–1984 data set.

A priori our preferred minimal specification is two lags ($L=2$), a two degree polynomial in y ($K_z=2$), and a one degree polynomial in x ($K_x=1$). Our reason for preferring this specification is that it is the minimal specification that can accommodate these two considerations: (i) Experience with linear Gaussian ARCH models suggests that typically at least two lags are needed to pick up ARCH effects (Engle (1982, p. 1002), Engle and Bollerslev (1986, pp. 19–20)); (ii) the polynomial in y must be at least of degree two to accommodate departures from Gaussian tails. To illustrate the number of parameters that this will entail, if one fits this specification to $y_t=(LBR, LCG)$, $M=2$, then there are 13 Gaussian VAR parameters in the term $N_M(y|b_0+Bx, RR')$ and 30 polynomial parameters ($p_z=6$, $p_x=5$) for a total of 43. Allowing for the normalization rule $a_0=1$ there are 42 effective parameters; with a sample size of $n=235$ the saturation ratio is $nM/p_\theta=11$ observations per parameter. If one fits to $y_t=(LBR, LSR, LCG)$, $M=3$, there are 27 VAR parameters, 70 polynomial parameters ($p_z=10$, $p_x=7$), a total of 97, effective 96, and saturation ratio of 10 with a sample of size $n=307$.

A detailed exploration of the likelihood surface of the series $y_t=(LBR, LCG)$ using the 1959–1978 data set is displayed in Table II. The p-values given under, for instance, the column heading K_x are for a comparison of the SNP (L, K_z, K_x) specification to its successor, an SNP (L, K_z, K_x+1) specification, using the asymptotic χ^2 distribution of the likelihood ratio test statistic; the exception is a SNP ($L, 0, K_x$) because its successor is considered to be an SNP ($L, 2, K_x$) specification for reasons discussed above. As an example, comparing a SNP

TABLE II
LIKELIHOOD SURFACE, $y_t=(LBR, LCG)$, 1959–1978

					p-value		
L	K_z	K_x	p_θ	$s_n(\hat{\theta})$	L	K_z	K_x
0	0	0	5	2.825500	.001		
1	0	0	9	2.719224	.01	.001	
1	2	0	14	2.649920	.01		.001
1	2	1	26	2.572404	.01	.04	.04
1	2	2	44	2.508069	.05		
1	3	1	38	2.526017	.01		
2	0	0	13	2.689188	.5	.001	
2	2	0	18	2.622474	.3		.001
2	2	1	42	2.506675	.2	.03	.07
2	2	2	102	2.343508			
2	3	1	62	2.435954		.4	
2	4	1	87	2.379054			
3	0	0	17	2.682065		.001	
3	2	0	22	2.612523			.001
3	2	1	58	2.463787			

(2,2,1) specification with a SNP (2,3,1) specification gives a χ^2 statistic of $\chi^2 = (2)(235)(2.506675 - 2.435954) = 33.23$ on $(62 - 42) = 20$ degrees of freedom which is significant at the 3% level.

A Gaussian VAR specification is overwhelmingly rejected as indicated by the p-values of .001 in Table II under the column heading K_z for rows with $K_z = K_x = 0$. A homogeneous, non-Gaussian, VAR specification, is also overwhelmingly rejected as indicated by the p-values of .001 in Table II under the column heading K_x for rows with $K_z > 0$ and $K_x = 0$. Note also from the p-values in Table II that $L = 2$ is the appropriate lag length.

The heterogeneity implied by these significant test statistics can be assessed graphically using plots of which Figure 1 is representative. At the bottom of Figure 1 is a plot of the contours of the density $h(y|x_{t-1}, \hat{\theta})$ in units of y_t, not units of $\tilde{y}_t$, for an SNP (2,2,1) specification fitted to the 1959–1978 data set with $x_{t-1} = (y_{1973.07}, y_{1973.08})$; $y_{1973.08}$ is an observation at the southwest edge of the

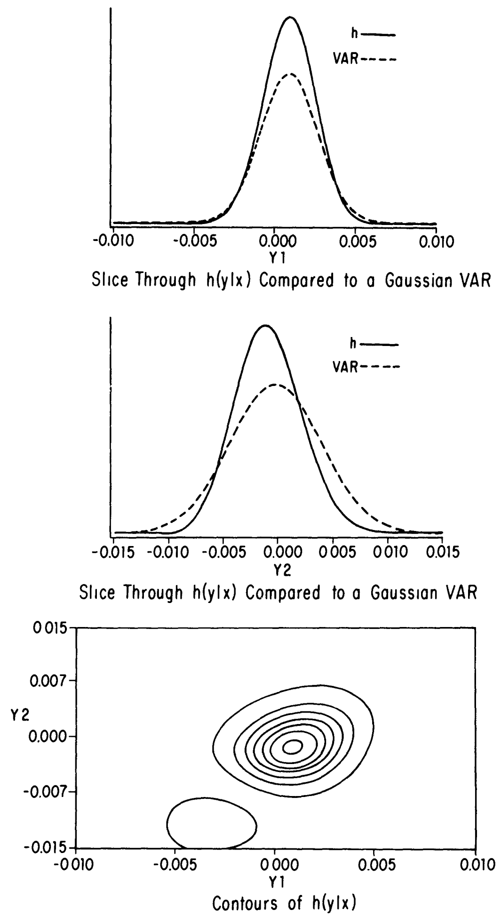

FIGURE 1.—SNP(2,2,1), $y = (LBR, LCG)$, $x = (y_{1973.07}, y_{1973.08})$, 1959–1978.

scatter plot of $\{y_t\}$ while its predecessor $y_{1973.07}$ is at the northeast edge. A comparison with similar plots at other points—specifically, $x_{t-1} = (y_{1965.09}, y_{1965.10})$, $y_{1965.10}$ is at the northeast edge with predecessor at the northeast edge, and $x_{t-1} = (y_{1972.04}, y_{1972.05})$, $y_{1972.05}$ is at the center with predecessor at the center—suggest nonlinearities in excess of that which can be accounted for by conditional heteroskedasticity because the shape of the contours is not elliptical and depends on x_{t-1}. The shape variation could be due to sampling variation or it could be due to significant nonlinearities in the data over and above conditional heteroskedasticity. What is important is that if such nonlinearities are present, our method can accommodate them.

Above the contour plot in Figure 1 are plots of slices taken through the mode of $h(y|x_{t-1}, \hat{\theta})$; over plotted for comparison is a similar slice through the mode of a two-lag VAR conditional density. Figure 1, and plots at other points, indicate that SNP conditional location and scale estimates can differ substantially from VAR location and scale estimates. They also indicate that, for these data, estimates at a saturation ratio of about 10 observations per parameter are relatively free of oscillation due to instability in the polynomial part of the model although some deterioration in the tails is observed.

If one tries to move to a more liberally parameterized model, say an SNP (2,3,1) specification with a saturation ratio of 7.6 observations per parameter in the 1959–1978 data set, this oscillation becomes more severe as seen in Figure 2. However, if one fits the same SNP (2,3,1) specification to the 1959–1984 data set, the saturation ratio is once again about 10 and the oscillation disappears as seen in Figure 3. Also, as one might expect from the discussion in Section 2, estimates of the tails improve as the number of observations increases with the saturation ratio held constant.

We also examined shape and tail behavior by looking at plots of slices, normalized to integrate to one, through the mode of $h(y|x_{t-1}, \hat{\theta})$ (as estimated in the 1959–1984 data set using an SNP (2,3,1) specification, at $x_{t-1} = (y_{1965.09}, y_{1965.10}), (y_{1972.04}, y_{1972.05})$, and $(y_{1973.07}, y_{1973.08})$) and comparing them to plots of a Gaussian density with the same mode and variance. A conditional dependence of shape and tail behavior on the past was observed. Again, this could be due to sampling variation or it could be due to significant nonlinearities in the data over and above conditional heteroskedasticity.

As a result of this specification search, we prefer specifications at a saturation ratio of about ten, the best, in our opinion, being an SNP (2,2,1) for $M = 2$ in the 1959–1978 data set, an SNP (2,3,1) for $M = 2$ in the 1959–1984 data set, and an SNP (2,2,1) for $M = 3$ in the 1959–1984 data set. These fits, together with a few others for comparison, are displayed in Table III.

In Section 5 we find that the bivariate (LBR, LCG) series can discriminate sharply between alternative specifications of utility whereas (LSR, LCG) series cannot. To help determine why, we computed the first and second conditional moments of the SNP (2,2,1), $M = 3$, specification as estimated from the 1959–1984 data set for each of the 307 observations in the data set. We found that the coefficient of variation over the 307 observations for any moment

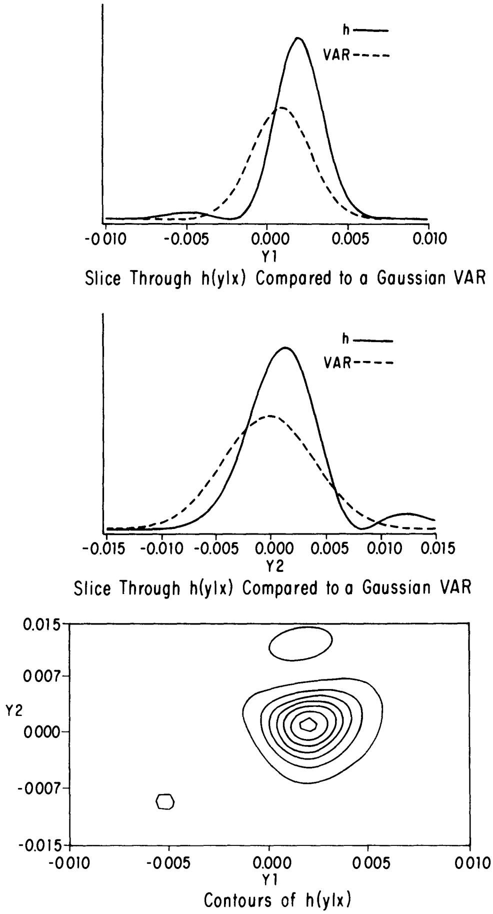

FIGURE 2.—SNP(2,3,1), $y = (LBR, LCG)$, $x = (y_{1973.07}, y_{1973.08})$, 1959–1978.

involving *LSR* was higher than the corresponding moment for *LBR*, ranging from 35% higher for the conditional covariance with *LCG* to 201% higher for the conditional mean. Also, the conditional mean of *LSR*, averaged over the 307 observations, was 282% larger than for *LBR*. As to the pattern of conditional correlations, the conditional correlation of *LBR* and *LSR* with *LCG*, averaged, was nearly the same (0.21) but the conditional correlation between *LSR* and *LBR* was nearly zero. Of the characteristics of the SNP fits that we examined, these seem to be the main differences and might account for the differences observed in Section 5.

4. CONDITIONAL MOMENT RESTRICTIONS FROM ASSET PRICING: THEORY

4.1. *Euler Equations*

We now derive the conditional moment restrictions that the intertemporal asset pricing model places on the SNP time series model. Assume the representative

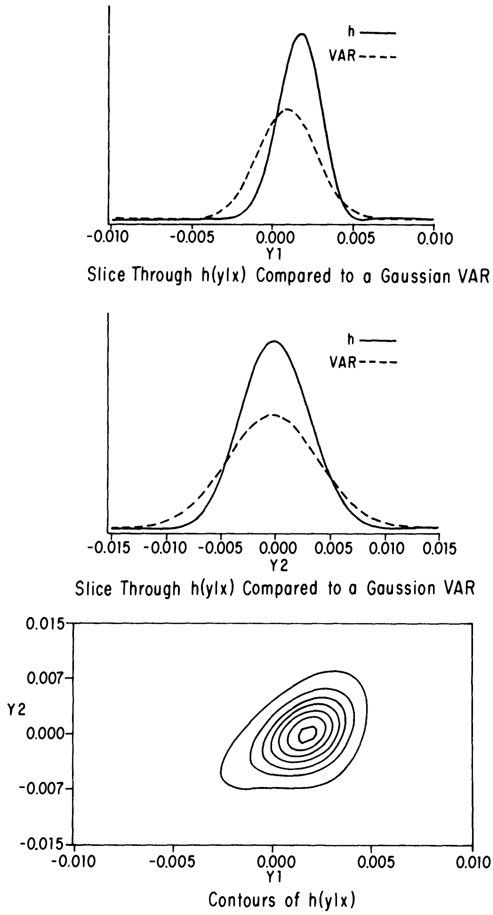

FIGURE 3.—SNP(2,3,1), $y=(LBR, LCG)$, $x=(y_{1973.07}, y_{1973\,08})$, 1959–1984.

agent's intertemporal utility function is

$$v_t = \sum_{k=0}^{\infty} \delta^k u(c^*_{t+k}), \qquad 0<\delta<1,$$

and the agent's task is to maximize $\mathscr{E}_t(v_t)$, where $c^*_t = (c_{t-J}, c_{t-J+1}, \ldots, c_t)$ is a vector comprised of contemporaneous and lagged consumption, $u(\xi) = u(\xi_0, \xi_1, \ldots, \xi_J)$ is a subutility function of $J+1$ arguments, δ is the subjective discount factor, and $\mathscr{E}_t(\cdot)$ is shorthand for $\mathscr{E}(\cdot\,|\,\mathscr{G}_t)$, where $\mathscr{G}_t$ is the agent's information set (a σ-field of events) at time t.

This formulation of the intertemporal utility function permits very general patterns of nonseparabilities across finite stretches of time. Accommodation of intertemporally nonseparable utility is motivated in part by Sims (1980) and Novales (1984) who argue that adjustment cost terms in the utility function play an important role in determining the covariance structure of consumption and real asset returns. It is also motivated by Dunn and Singleton (1986) and

TABLE III

OPTIMIZED LIKELIHOODS FOR THE UNCONSTRAINED LAW OF MOTION

L	K_z	K_x	p_θ	$s_n(\hat{\theta})$	Saturation Ratio (obs/parms)
			$y_t = (LSR, LCG)$, 1959–1978		
1	0	0	9	2.784882	58.8
1	2	1	26	2.697092	18.1
2	0	0	13	2.762652	39.2
2	2	1	42	2.643757	11.2
2	3	1	62	2.588922	7.6
3	0	0	17	2.752573	29.4
3	2	1	58	2.585973	8.1
			$y_t = (LSR, LGC)$, 1959–1984		
1	0	0	9	2.790343	68.2
1	2	1	26	2.726719	23.6
2	0	0	13	2.778216	51.2
2	2	1	42	2.680522	14.6
2	3	1	62	2.624858	9.9
3	0	0	17	2.772228	38.4
3	2	1	58	2.634346	10.6
			$y_t = (LBR, LCG)$, 1959–1978		
1	0	0	9	2.719224	58.8
1	2	1	26	2.572404	18.1
2	0	0	13	2.689188	39.2
2	2	1	42	2.506675	11.2
2	3	1	62	2.435954	7.6
3	0	0	17	2.682065	29.4
3	2	1	58	2.463787	8.1
			$y_t = (LBR, LCG)$, 1959–1984		
1	0	0	9	2.607457	68.2
1	2	1	26	2.510188	23.6
2	0	0	13	2.579602	51.2
2	2	1	42	2.462479	14.6
2	3	1	62	2.413526	9.9
3	0	0	17	2.565491	38.4
3	2	1	58	2.435519	10.6
			$y_t = (LSR, LBR, LCG)$, 1959–1984		
1	0	0	18	4.019321	54.2
1	2	1	57	3.879348	16.2
2	0	0	27	3.970963	35.4
2	2	1	96	3.777279	9.6

Eichenbaum and Hansen (1989) who argue that consumption goods may display local durability by yielding utility flows that extend for one or more periods beyond the period of acquisition. Finally, Constantinides (1988) uses a specific intertemporal utility function of this form to incorporate habit formation for nondurables and services consumption. He shows that, with it, the first and second moments of consumption growth and returns can, in principle, be reconciled given reasonable values of a risk aversion parameter.

In period t the agent's opportunity set for transferring consumption between periods t and $t+1$ is a set of assets, indexed by l. One unit of the consumption good invested in the lth asset yields the stochastic return of $r_{l,t+1}$ units of the consumption good next period. Given this opportunity set, the Euler equations for maximizing intertemporal utility are

$$\mathscr{E}_t[(\partial v_t/\partial c_{t+1})r_{l,t+1} - (\partial v_t/\partial c_t)] = 0$$

for each asset l. The Euler equations can be expressed as

$$\mathscr{E}_t[\delta \Delta u(c_{t+1}^{**})r_{l,t+1} - \Delta u(c_t^{**})] = 0$$

where $\Delta u(\nu_0, \nu_1, \ldots, \nu_{2J})$ is a function of $2J+1$ arguments defined by

$$\Delta u(\nu_0, \nu_1, \ldots, \nu_{2J}) = \sum_{i=0}^{J} \delta^i (\partial/\partial \xi_{J-i}) u(\nu_i, \nu_{i+1}, \ldots, \nu_{i+J})$$

and where

$$c_t^{**} = (c_{t-J}, c_{t-J+1}, \ldots, c_{t+J})$$

is a stretch of the consumption realization centered at c_t and extending for J time units backwards and J periods forward in time. With the Euler equations written this way, $\Delta u(\cdot)$ is interpreted as a generalized marginal utility of consumption.[3]

4.2. *Transformation to Stationarity and Cross-sectional Aggregation*

Asset returns are reasonably modeled as stationary. Consumption, on the other hand, exhibits upward secular drift and is not stationary in levels, although it is reasonably taken as stationary in growth rates. Some device is therefore needed to re-express the Euler equations in such a way that they only depend upon stationary variables.

Our strategy is the natural extension of the one utilized by Hansen and Singleton (1982, 1983) and Mehra and Prescott (1985). Specifically, we assume that the subutility function $u(\cdot)$ is a linear transformation of a function that is homogeneous of degree $1-\gamma$, $\gamma \in \mathbb{R}$, and proceed as follows. This assumption about $u(\cdot)$ implies that its partial derivatives are homogenous of degree $-\gamma$; thus, Δu is likewise homogeneous of degree $-\gamma$. Consequently, multiplication of the Euler equation by c_t^{γ} and bringing this inside the arguments of Δu yields

$$\mathscr{E}_t[\delta \Delta u(c_{t+1}^{**}/c_t)r_{l,t+1} - \Delta u(c_t^{**}/c_t)] = 0$$

where the / is interpreted in the sense of elementwise division. Now let $q_t = c_t/c_{t-1}$ denote consumption growth from $t-1$ to t; the q_t process is assumed to be jointly stationary with returns. Elementary multiplicative identities like $c_{t+2}/c_t = q_{t+2}q_{t+1}$, $c_{t-2}/c_t = 1/(q_t q_{t-1})$, etc., imply that, given the homogeneity assumption, the Euler equations depend on consumption only through q_s for

[3] Hansen and Richard (1987) contains a useful discussion and characterization of the restrictions that Euler equations place on asset pricing relations.

$s = t - J + 1, \ldots, t + J + 1$, which are stationary by assumption, and thus stationarity of the Euler equations is effected.[4]

The assumption of a homogeneous utility function is more than enough to ensure that Gorman's (1953) conditions (see also Barnett, 1981, Appendix B) for cross-sectional aggregation are satisfied for a standard static demand problem, as what is required is that the utility function be an affine translation of a homothetic function. Under a complete contingent markets assumption, however, there is no logical difference between the static demand problem and our intertemporal problem, save for some technical difficulties arising from an infinite dimensional commodity space (Eichenbaum, Hansen, and Richard (1984)). One can expect, then, that under complete markets and further regularity conditions, preferences such as ours will aggregate in the sense of Gorman and a fictitious representative consumer will exist. In particular, it seems likely that there are sufficient conditions under which the economy evolves to a stochastic steady state where consumption/wealth ratios are the same across agents, though exploring this conjecture is well beyond the scope of this paper.

4.3. *Restrictions as Integrals Against the Law of Motion*

To estimate the law of motion $h(y|x)$ subject to the I-CAPM restrictions we need to express the Euler equations as requirements that integrals against h must vanish. To do so, we first note that the observed vector y_t contains the logarithms of a subset of the returns, say M_a returns, and the logarithm of consumption growth. Thus there are mappings

$$r_{l,t} = \exp(y_{l,t}) \qquad (l = 1, 2, \ldots, M_a),$$
$$q_t = \exp(y_{M,t}),$$

through which consumption growth and observed returns can be recovered from y_t. Consequently, since the consumption ratios c_{t+j}/c_t and real returns $r_{l,t+1}$ can be recovered from the elements of y_{t+j}, the Euler equations

$$\mathscr{E}_t[\delta \Delta u(c^{**}_{t+1}/c_t) r_{l,t+1} - \Delta u(c^{**}_t/c_t)] = 0$$

for the M_a observed assets imply that M_a functions of leads and lags of the variable y_t have conditional expectations of zero,

$$\mathscr{E}_t[g_l(y_{t-J+1}, y_{t-J+2}, \ldots, y_{t+J+1})] = 0.$$

The functions g_l are defined by comparison with the previous display.

The integrals for the conditional moment restrictions will be simpler notationally if we move the time index back one period and view the expectations as

[4] Eichenbaum and Hansen (1989) consider two models of growth. Their first model entails logarithmic detrending to induce stationarity while their second model entails logarithmic differencing, and is thus similar to ours. On pure statistical grounds, they could find little reason to choose one model of growth over the other, though an asset pricing model estimated using the same variables and same specification of preferences fit the data somewhat better under their second model. (See their Tables 5.2 and 5.4.)

being taken at time $t-1$. Therefore, we rewrite the Euler equations as

$$\mathscr{E}_{t-1}[g_l(y_{t-J}, y_{t-J+1}, \ldots, y_{t+J})] = 0.$$

Finally, by a familiar iterated expectations argument, conditional moment restrictions are preserved under a reduction of the conditioning set. That is, the information contained in $\{y_{t-j}\}$, $j \geq 1$, is reasonably assumed to be no larger than that available to the agent at time $t-1$, and so the σ-field generated by $(y_{t-L}, \ldots, y_{t-1})$ is a sub-σ-field of that used to form $\mathscr{E}_{t-1}(\cdot)$. Thus, by the law of iterated expectations,

$$\mathscr{E}[g_l(y_{t-J}, y_{t-J+1}, \ldots, y_{t+J}) | y_{t-L}, y_{t-L+1}, \ldots, y_{t-1}] = 0$$

so the basic form of the restrictions remains the same when we condition only on the history of the observed process.

This last conditional expectation translates directly into the restriction that M_a integrals against the law of motion must vanish:

$$\int \cdots \int g_l(y_{-J}, \ldots, y_{-1}, y_0, y_1, \ldots, y_J)$$
$$\times h_J^*(y_0, y_1, \ldots, y_J | y_{-L}, y_{-L+1}, \ldots, y_{-1}) \prod_{i=0}^{J} dy_i = 0$$

where h_J^* is the joint density of $y_t, y_{t+1}, \ldots, y_{t+J}$, conditional on the processes history. In other words, g_l is a function of $2J+1$ vector arguments; the last $J+1$ arguments of g_l are integrated out against h_J^* and the resulting function must vanish identically. Some notational economy is gained by letting x here represent a vector of length $M \cdot \max(J, L)$, putting $y^* = (y_0, y_1, \ldots, y_J)$, and observing that the restriction is equivalent to

$$\int g_l(x, y^*) h_J^*(y^* | x)\, dy^* = 0.$$

There is one integral condition for each observed asset and each one must hold identically in the conditioning variable x.

Note that here we are pricing one-period securities, though multi-period conditional moment restrictions arise because of nonseparable utility. In other applications, for example, term structure problems and pricing pure discount securities, multi-period conditional moment restrictions will arise even with additively separable utility.

4.4. *The Utility Function*

Following Gallant (1982) we employ a seminonparametric strategy for estimation of the subutility function $u(\xi)$. By assumption, the function $u(\xi)$ is a linear transformation of a function that is homogenous of degree $1-\gamma$ in $J+1$ arguments. We can thus characterize it as linear transformation of the product of two functions, one which is a function of one argument and is homogenous of degree $1-\gamma$, while the other is a function that is homogeneous of degree zero.

We then expand the second function in a multivariate polynomial series, which in estimation is truncated at a degree determined empirically.

The characterization of $u(\xi) = u(\xi_0, \xi_1, \dots, \xi_J)$ is obtained by factoring out the argument ξ_J, which corresponds to c_t, and thereby writing

$$u(\xi) = (1-\gamma)^{-1}\left[\xi_J^{1-\gamma}\bar{u}(w) - 1\right]$$

where $\bar{u}(w)$ is a function of J arguments given by

$$\bar{u}(w_1, w_2, \dots, w_J) = u(w_1, w_2, \dots, w_J, 1)$$

with $w_k = \xi_{k-1}/\xi_J$, $k = 1, 2, \dots, J$. Evidently, the function of one argument is the familiar CRR utility function. Thus we represent $u(\xi)$ as being a linear transformation of the product of the CRR utility function and another function, $\bar{u}(w)$, which depends only on the consumption ratios w_k and captures nonseparabilities. For the special case $J = 0$, the function $\bar{u}$ is set identically equal to unity, and then $u(\xi)$ is the CRR utility function. For the other special case $\gamma = 1$, we set $\bar{u}$ to unity and u to $\log(\xi_J)$.

The expansion of $\bar{u}(w)$ is

$$\bar{u}(w) = d_0 + (1-\gamma)^2 \lim_{K_u \to \infty} \sum_{|\lambda|=1}^{K_u} d_\lambda w^\lambda$$

where d_0 is always normalized to equal unity. The multiplication by $(1-\gamma)^2$ has no essential effect when $\gamma \neq 1$, since it amounts to a redefinition of the coefficients of w^λ in the series expansion. But it forces the coefficients to vanish when $\gamma = 1$, so that we effectively make the class of functions generated by our approach to be of the form

$$u(\xi) = \begin{cases} \text{linear transformations of homogeneous functions} \\ \quad \text{of degree } \gamma \neq 1, \\ \log(\xi_J) \quad \text{for } \gamma = 1, \end{cases}$$

and the partial derivatives of the $u(\xi)$ converge continuously to those of $\log(\xi_J)$ as $\gamma \to 1$. In estimation, we truncate the expansion of u at degree K_u, that is, we use the approximating functions

$$\bar{u}_{K_u}(w) = d_0 + (1-\gamma)^2 \sum_{|\lambda|=1}^{K_u} d_\lambda w^\lambda,$$

$$u_{K_u}(\xi) = (1-\gamma)^{-1}\left[\xi_J^{(1-\gamma)}\bar{u}_{K_u}(w) - 1\right],$$

and employ a standard upward testing approach to determine the appropriate degree K_u beyond which the d_λ contribute insignificantly to the likelihood.

Our expansion can reach a wide class of subutility functions as $K_u \to \infty$. The class includes, in particular, the function utilized in Dunn and Singleton (1986) and Eichenbaum and Hansen (1989), which in our notation is written

$$u(\xi) = (1-\gamma)^{-1}\left[\xi_J^{(1-\gamma)}(1+\alpha' w)^{(1-\gamma)} - 1\right]$$

where α is a $J \times 1$ parameter vector.

One potential drawback from introducing as much flexibility as we do is that the implied intertemporal utility function need not be globally regular, that is, have everywhere positive marginal utilities and be everywhere concave for each finite K_u. This difficulty is inherent in most work with flexible functional forms, since, among other things, regularity restrictions typically entail inequality constraints which create difficulties of inference using classical estimation procedures.

4.5. *Numerical Methods for Imposing Conditional Moment Restrictions*

Recall from Subsection 4.3 that the conditional moment restrictions generated by the I-CAPM model require that certain integrals must vanish, where the integrals are taken against the $J+1$ step-ahead conditional density of the observed data:

$$\int g_l(x, y^*) h_J^*(y^*|x)\, dy^* = 0.$$

Here y^* is of length $M(J+1)$, x is of length $M \cdot \max(J, L)$, and $h_J^*(y|x)$ is the J-step ahead density generated by the one-step true density of the data, $h(y|x)$. There is one integral condition for each observed asset $l = 1, 2, \ldots, M_a$.

After replacement of the theoretical conditional density and the utility function by their parametric approximants, then the conditional moment restrictions become parametric restrictions, obtained in the following way. Let $h_K(y|x, \theta_1)$ denote the parametric SNP model for the one-step density where θ_1 contains all of the parameters of the law of motion. Denote by $h_{K,J}^*(y^*|x, \theta_1)$ the $J+1$ step ahead density generated by $h_K(y|x, \theta_1)$. The $J+1$ step density is obtained from the one step density by forming the appropriate products of the one step density. Now let θ_2 denote a vector containing the parameters of the utility function, which are the d_λ from the polynomial part, and δ and γ. Let $g_{l, K_u}(x, y^*, \theta_2)$ denote the function g_l of the conditional moment restriction for asset l when the parameterized utility function u_{Ku} is used for $u(\xi)$. Finally, let $\theta = (\theta_1, \theta_2)$ be a vector containing all of the parameters of the model. Then if we put

$$\tau_l(x, \theta) = \int g_{l, K_u}^*(x, y^*, \theta_2) h_{K,J}^*(y^*|x, \theta_1)\, dy^*,$$

the restrictions take the form

$$\tau_l(x, \theta) = 0$$

identically in x for each asset l.

Unlike linear rational expectations models where the $\tau_l(x, \theta)$ are linear in x and one can determine analytically the parametric restrictions on θ inherent in the requirement that $\tau_l(x, \theta)$ vanish, the functions here are too complicated to determine the parametric restrictions analytically. Indeed, the integrands are sufficiently complicated that we have to use numerical quadrature to do the integration. Nevertheless, we can effectively impose the restrictions by requiring that $\tau_l(x, \theta)$ vanish for x's restricted to lie on a sufficiently fine lattice of discrete points, and then utilizing software that can accommodate side restrictions. Specifically, we take as the lattice a determining set for a multivariate polynomial (see the Appendix) and impose on the estimation the parametric restrictions

$$\tau_l(x_k, \theta) = 0 \qquad (k = 1, 2, \ldots, K^*;\ l = 1, 2, \ldots, M_a)$$

where K^* is the number of x_k in the determining set and, as before, M_a is the number of assets in the observed vector. If the functions $\tau_l(x, \theta)$ are sufficiently smooth in x and θ and if the lattice is fine enough so that $K^* M_a$ exceeds the length of θ, then this has to impose all of the restrictions; we find in estimation that much courser lattices suffice.

We find that this scheme of evaluating the conditional moment restrictions on a determining set does effectively impose the restrictions on the estimation; in fact, in one sense it does too much, as many of the parametric restrictions turn out to be redundant. One can see intuitively why there has to be redundancies among the restrictions by thinking about what would happen if one employed our evaluation scheme to impose the restrictions of a linear rational expectations model. (We do not recommend actually doing this for linear models.) To be concrete, suppose the model dictates that the coefficients of three lags of a variable must vanish, and these three coefficients are possibly nonlinear functions of many more deep parameters. There are only three effective constraints on the deep parameters, but if one imposes the constraints by restricting the function τ to vanish at 50 distinct points in $\mathbb{R}^3$, then 47 restrictions will be redundant.

Mathematically speaking, the redundancy is irrelevant, though in practice it creates some difficulties for us in that the software we use for optimization cannot readily accommodate a Jacobian matrix with linearly dependent rows. We have no direct way of determining analytically which restrictions are redundant, and so we employ numerical methods to reduce the dimensionality of the restrictions and thereby achieve numerical stability in the optimization. Specifically, we stack the restrictions into a vector $\tau(\theta)$ of length $K^* M_a$, find a matrix $\mathscr{H}_0$ with r^* orthonormal rows such that the Jacobian of $\mathscr{H}_0 \tau(\theta)$ has rank r^*, and impose the r^* constraints $\mathscr{H}_0 \tau(\theta) = 0$ on the estimation. These methods are described in more detail in the Appendix.

5. CONDITIONAL MOMENT RESTRICTIONS FROM ASSET PRICING: EMPIRICAL RESULTS

In this section, the seminonparametric specifications of the law of motion identified as reasonable in Section 3 are re-estimated subject to the conditional moment restrictions derived in the previous section. Table IV contains a summary of the results of the estimation. Below, we first give a description of the reporting style used in the table and then give a summary of our findings.

The first column of Table IV contains keys used to reference the rows of Table IV throughout the discussion; a key uniquely identifies the specification of the law of motion, the specification of the utility function, and the data set used in

TABLE IV

OPTIMIZED LIKELIHOODS FOR THE CONSTRAINED LAW OF MOTION

											p-values			
Key	L	K_z	K_x	J	K_u	p_θ	NCNLN	$\tilde{\delta}$	$\tilde{\gamma}$	$s_n(\tilde{\theta})$	Model	γ	J	K_u
								Stocks: $y_t = (LSR, LCG)$, 1959–1978						
(a)	2	2	1	—	—	42	—	—	—	2.643757				
(b)	2	2	1	0	—	43	10/15	0.99852	—	2.663505	.4	.5		
								(0.00360)						
(c)	2	2	1	0	—	44	10/15	0.99390	−1.84262	2.662323	.4			
								(0.00820)	(3.22596)					
								T-bill: $y_t = (LBR, LCG)$, 1959–1978						
(d)	2	2	1	—	—	42	—	—	—	2.506675				
(e)	2	2	1	0	—	43	10/15	0.99958	—	2.581942	.001	.001		
								(0.00035)						
(f)	2	2	1	0	—	44	10/15	1.00117	0.67659	2.544367	.02		.01	
								(0.00070)	(0.23019)					
(g)	2	2	1	1	1	45	10/15	1.01412	6.96761	2.531001	.1		.06	.1
								(0.02540)	(12.27502)					
(h)	2	2	1	1	4	48	10/15	1.00233	1.21624	2.518563	.2			
								(0.00325)	(1.46321)					
(i)	2	2	1	2	1	46	10/15	1.00106	0.60313	2.523651	.2			
								(0.00047)	(0.08623)					
								T-bill: $y_t = (LBR, LCG)$, 1959–1984						
(j)	2	3	1	—	—	62	—	—	—	2.413526				
(k)	2	3	1	0	—	63	10/15	0.99876	—	2.485066	.001	.001		
								(0.00041)						
(l)	2	3	1	0	—	64	10/15	1.00528	3.99358	2.436891	.07		.001	
								(0.00790)	(5.39684)					
(m)	2	3	1	1	1	65	10/15	1.00078	1.05406	2.419048	.8		.3	.6
								(0.00061)	(0.24795)					
(n)	2	3	1	1	4	68	10/15	1.00069	1.00210	2.416341	.8			
								(0.00076)	(0.32540)					
(o)	2	3	1	2	1	66	10/15	1.00087	1.09866	2.417455	.9			
								(0.00071)	(0.29494)					
								Stocks and T-Bills: $y_t = (LSR, LBR, LCG)$, 1959–1984						
(p)	2	2	1	—	—	96	—	—	—	3.777279				
(q)	2	2	1	0	—	97	40/56	0.99816	—	3.916960	.001	.001		
								(0.00063)						
(r)	2	2	1	0	—	98	40/56	0.99398	−2.23949	3.830767	.7		.07	
								(0.00203)	(0.91034)					
(s)	2	2	1	1	1	99	40/56	1.00558	3.75107	3.825561	.8		.1	.4
								(0.00430)	(2.43682)					
(t)	2	2	1	1	4	102	40/56	1.00552	3.64713	3.820434	.8			
								(0.00437)	(2.39127)					
(u)	2	2	1	2	1	100	40/56	1.00586	3.77929	3.821816	.8			
								(0.00506)	(2.78394)					

the estimation. The next five columns of Table IV display the tuning parameters of the law of motion and subutility function that were used for each estimation. The next column, labeled p_θ, indicates the number of free parameters of the fit. The column labeled NCNLN indicates the number of rows of the orthogonal matrix used to reduce the restrictions to full rank; thus, the entry 10/15 means that $\tau(\hat{\theta})$ has fifteen rows with ten effective restrictions and five redundant ones (see the Appendix). The next three columns show, respectively, the estimates of δ and γ, with Wald-type standard errors computed from the inverse of the bordered information matrix, and the optimized value of the objective function.

The p-values given in Table IV require more explanation. They are computed using the asymptotic χ^2 distribution of the likelihood ratio test statistic and are associated to the more restricted of two models being compared. Thus, in reading across a row, if all entries are larger than .05, then that model would be accepted at that significance level. A p-value under the heading Model is a comparison with the unrestricted law of motion. An entry under any other heading is a comparison with the next less restricted model with respect to that heading. For instance, to test whether consumption matters at all for the pricing of assets, that is, to test whether the restrictions of the constant discount factor hypothesis ($J=0$, $K_u=0$, $\gamma=0$) can be imposed on the unrestricted SNP $(2,2,1)$ specification using the series $y_t=(LSR, LCG)$ for the years 1959 to 1978 one has $\chi^2=(2)(235)(2.663505-2.643757)=9.28$ on $42-(43-10)=9$ degrees of freedom which is significant at the 40% level. Similarly, to test the constant discount factor hypothesis against the CRR subutility function ($J=0$, $K_u=0$) with the same data one has a $\chi^2=(2)(235)(2.663505-2.662323)=.56$ on $(44-10)-(43-10)=1$ degree of freedom which is significant at the 50% level.

One apparent conclusion from Table IV is that the stock returns series alone does not have a lot of discriminatory power. In row (c), the estimate of the subjective discount factor δ is less than unity, but the estimate of the curvature parameter γ is imprecise and lies outside the region where utility is concave. The constant discount factor model (risk neutral asset pricing) for stock returns is acceptable at conventional significance levels, row (b). This model is acceptable when tested as a restriction of the law of motion, row (b) versus row (a), and as a restriction on the CRR utility model, row (b) versus row (c).

The T–bill returns series, on the other hand, has considerably more discriminatory power. The constant discount factor model is strongly rejected for T-bills in the 1959–78 period, row (e), and in the 1959–84 period, row (k). It is also rejected for the joint estimation with T-bill returns and stock returns over the 1959–84 period, row (q). For estimations including T-bills, δ tends to exceed unity, though not significantly so in most cases, and γ is generally positive, which is consistent with the pattern discussed by Singleton (1989). The estimations reveal reasonably convincing evidence against the CRR specification of utility, as indicated by the p-values in rows (f), (l), and (r), for the direct tests of the CRR restrictions on the law of motion and the tests that at least one lagged term is needed in our seminonparametric specification of utility. This evidence against the CRR specification complements that presented by Hansen and Singleton

(1983) and Dunn and Singleton (1986), who employ much different estimation strategies for the shorter 1959–78 period, and with that presented by Eichenbaum and Hansen (1989), who work with both the 1959–78 and 1959–85 periods.

We now undertake the specification search using the seminonparametric model for utility. The general form of the subutility function is

$$u(c_{t-J},\ldots,c_t) = (1-\gamma)^{-1}\{c_t^{(1-\gamma)}[1+\text{poly}(c_{t-J}/c_t,\ldots,c_{t-1}/c_t)]-1\}$$

where poly($\cdot$) is a polynomial of degree K_u in J arguments.

With the polynomial restricted to be linear, $K_u = 1$, the test for $J = 1$ versus $J = 2$ reveals some evidence suggesting that $J = 2$ is appropriate for the 1959–78 period, row (g) versus (i), but little or no evidence for $J = 2$ being appropriate for the longer 1959–84 period, rows (m) versus (o) and (s) versus (u). On this basis, we consider $J = 1$ as the appropriate lag length for the subutility function.

With $J = 1$, the test for a linear versus a quartic specification for the polynomial reveals virtually no evidence in favor of the more complex specification, rows (g) versus (h), (m) versus (n), (s) versus (t). A quartic was chosen since it was the highest degree that proved computationally feasible and could be expected to provide an excellent approximation to functions such as $(1+\alpha c_{t-1}/c_t)^{(1-\gamma)}$, which is the adjustment for nonseparability implied by the utility function used in Dunn and Singleton (1986) and Eichenbaum and Hansen (1989). In additional work (not reported in the table) we made an effort to explore more fully the likelihood surface as the degree of the polynomial varies. The computations were expensive, and we did encounter some difficulties with obtaining convergence of the constrained optimization algorithm, which suggests that the data do not discriminate all that well among different functional forms for the nonseparability. In the end, though, this effort did not uncover any further evidence in favor of a more complex specification beyond the linear polynomial. On this basis, and since the tests of the overidentifying restrictions implied by the linear polynomial reveal little evidence against it (see the p-values under Model in rows (g), (m), and (s)), we choose the linear adjustment model as the preferred model from within our nested family of models.

The preceding analysis of the fits obtained using both the CRR and the SNP utility specifications made extensive use of the χ^2 test of the overidentifying restrictions. The χ^2 test is an omnibus test of specification that provides an indication of whether a particular utility specification is consistent with the data. It does not, however, provide much insight about the dimensions along which the utility models do or do not do well in fitting the data. It is of particular interest, then, to explore and characterize the manner in which imposition of the restrictions distorts the law of motion of the data.

We computed the conditional means and variances at each data point using both the unrestricted and restricted laws of motion. This was done for all of the various specifications of utility reported in Table IV. A summary of the calculations, which is representative of what we found, is reported in Table V. The entries in the table are the average restricted conditional moment less the average unrestricted conditional moment expressed as a percentage of the average unre-

TABLE V

PERCENTAGE ERROR IN RESTRICTED CONDITIONAL MOMENTS

		Conditional Moment[b]								
		Mean			Variance			Covariance		
Key[a]	Utility	*LSR*	*LBR*	*LCG*	*LSR*	*LBR*	*LCG*	*LSR* − *LBR*	*LSR* − *LCG*	*LBR* − *LCG*
(f)	CRR		−3.3	2.1		12.3	4.0			24.5
(g)	SNP		−5.0	0.0		0.3	0.9			5.9
(l)	CRR		−0.5	−3.4		0.6	7.7			8.5
(m)	SNP		−2.6	1.5		0.6	1.5			1.4
(r)	CRR	−80.0	6.2	8.6	3.0	5.7	9.9	0.7	14.2	33.2
(s)	SNP	−38.9	6.0	12.3	2.2	0.7	3.4	2.8	−1.2	−15.0

[a] (f) and (g): T-bills, $y_t = (LBR, LCG)$, 1959–1978. (l) and (m): T-bills, $y_t = (LBR, LCG)$, 1959–1984. (r) and (s): Stocks and T-bills, $y_t = (LSR, LBR, LCG)$, 1959–1984.

[b] Conditional moments were computed for each observation in the indicated data set and then averaged. Entries in the table are the average restricted conditional moment less the average unrestricted conditional moment expressed as a percentage of the average unrestricted conditional moment.

stricted conditional moment. Two quite interesting conclusions emerge from Table V. First, imposition of the constraints from CRR utility causes the law of motion to overpredict the conditional variance of consumption growth and the conditional covariance of consumption growth with asset returns. Introduction of the nonseparability adjustment, however, mitigates this overprediction to a large degree. Second, for the estimations using both stock and T-bill returns, rows (r) and (s), the constrained laws of motion systematically underpredict the conditional mean return on stocks, which is a manifestation of the so-called equity premium puzzle. The introduction of the nonseparability adjustment helps shrink the gap, though it does not quite go all the way towards closing it.

We now commence an extensive appraisal of the economic properties of the fitted SNP utility function. In view of the preceding discussion, our preferred specification of the subutility function is

$$u(c_{t-1}, c_t) = (1-\gamma)^{-1}\{c_t^{(1-\gamma)}[1 + d_1'(c_{t-1}/c_t)] - 1\}.$$

Sensible interpretation of the parameter estimates of γ and d_1' requires consideration of the implied intertemporal utility function. From Section 4 this function is

$$v_t = \sum_{j=0}^{\infty} \delta^j u(c_{t-1+j}, c_{t+j})$$

which maps contingent paths for current and future consumption into real numbers.

Insight into the role of nonseparabilities can be obtained by using v_t to compute a fictitious one-period sure return generated by a consumption path (assumed known to the agent) and then examining how this return varies with movements in c_{t-1}. This return is

$$r_t^f = \frac{\partial v_t / \partial c_t}{\partial v_t / \partial c_{t+1}}$$

viewed as a function of the consumption path $(c_{t-1}, c_t, c_{t+1}, c_{t+2}, \ldots)$; r_t^f is the reciprocal of the marginal rate of substitution between c_t and c_{t+1}. If r_t^f goes up with an increase in previous consumption, c_{t-1}, then the preferences between c_t and c_{t+1} are tilted relatively in favor of current consumption, c_t, which is consistent with habit persistence and adjustment cost notions of nonseparabilities. On the other hand, if r_t^f goes down with an increase in c_{t-1}, then preferences are tilted in favor of future consumption, c_{t+1}, which is consistent with a local durability of consumption notion of nonseparability, wherein previous acquisitions of nondurable goods yield a service flow into the subsequent period.

One can show that when $J=1$ and the polynomial is linear, then the sign of $\partial r_t^f/\partial c_{t-1}$ depends both upon the sign of d_1' and upon which side of unity γ lies. Specifically,

$$\operatorname{sign}\left(\partial r_t^f/\partial c_{t-1}\right) = \operatorname{sign}\left(d_1'\right) \qquad \text{if} \qquad \gamma > 1,$$

$$\operatorname{sign}\left(\partial r_t^f/\partial c_{t-1}\right) = -\operatorname{sign}\left(d_1'\right) \qquad \text{if} \qquad \gamma < 1.$$

For each of the three estimations we did with $J=1$ and a linear polynomial the point estimates of d_1' are negative, -0.528 for the row (g) estimation, -0.630, row (m), and -0.410, row (s), and the estimates of γ exceed unity which is consistent with the durability notion of nonseparability. Due to the nature of the parameterization (see Section 4), the estimates of d_1' are functions of other parameter estimates, and Wald type standard errors tend to be large. Invariance with respect to parameterization can be obtained by inverting the likelihood ratio test to get confidence intervals (Gallant (1987, pp. 108–110)). For the row (m) estimation, a 95 percent confidence interval for d_1' obtained in this manner is $(-4.82, -0.067)$, which is wide but lies entirely to the left of zero. Taken together, the evidence appears inconsistent with the adjustment cost notion of nonseparability and consistent with the local durability notion, which is in accordance with the findings of Dunn and Singleton (1986) and Eichenbaum and Hansen (1989) who employ GMM estimation and a different utility function.

As an indication of the plausibility of our parameter estimates, we checked for whether the predicted marginal intertemporal utilities $\partial v_t/\partial c_t$ are positive and whether the implied intertemporal utility function v_t is concave. For this purpose, we computed $\operatorname{sign}(\partial v_t/\partial c_t)$ and checked for negative definiteness of the 3×3 Hessian matrix $[\partial^2 v/\partial c_{t+i}\,\partial c_{t+j}]_{i,j=0,1,2}$ at each available data point for the estimations, save for those lost due to lags or the leads required to compute the Hessian. This calculation yields:

Key	Positive marginal utilities	Hessian negative definite
(g)	228 of 233	228 of 233
(m)	300 of 305	0 of 305
(s)	300 of 305	300 of 305

For all three estimations the marginal utilities are almost always of the right sign. The Hessian also almost always has the correct definiteness properties, except for the row (m) estimation which has been seen to yield point estimates somewhat out of line relative to the other two. With the exception of that estimation, the utility function has the correct curvature properties.

6. CONCLUSION

We close with a summary of our empirical findings. We can say first, and rather emphatically, that for the version of intertemporal asset pricing model with a strictly Gaussian law of motion and time separable utility, both maintained assumptions are wrong. The fact that the law of motion is misspecified is clear from Section 3 above, where we find that SNP models for consumption growth and real asset returns display substantial conditional heterogeneity. Misspecification of the law of motion, though, is not the source of the rejections of this model that were obtained by Hansen and Singleton (1983) using maximum likelihood methods under Gaussian assumptions, but rather it is misspecification of the utility function. The fact that a time separable utility function is misspecified is perhaps implicit in the findings of other work using GMM estimation and CRR utility. However, here we pin the conclusion down by actually estimating the SNP law of motion subject to the restrictions from CRR utility and we uncover evidence against the model. In particular, imposition of the restrictions from CRR utility forces the law of motion to overpredict the conditional variance of consumption growth and the conditional covariance of consumption growth with asset returns. Our findings indicate that one can expect a low payoff on effort directed towards specification and estimation of asset pricing models that take into account complicated higher order dynamics, for example, the multivariate ARCH effects described in Grossman, Melino, and Shiller (1987) and Bollerslev, Engle, and Wooldridge (1988), unless the restrictive assumption of time separable utility is relaxed as well.

Second, taking our results together with those of Dunn and Singleton (1986) and Eichenbaum and Hansen (1989) indicates that the source of the intertemporal nonseparability is not habit persistence or adjustment costs, but rather it is the "temporary" or "local" durability of nondurable consumption that gives rise to the nonseparability. On this issue we get similar qualitative results, but with entirely different functional forms and estimation strategies.

Third, although we ignore durable goods in the analysis, we are generally able to find qualitatively reasonable and quite parsimonious seminonparametric specifications for utility that fit the data, in the sense that the test statistics on the overidentifying restrictions are not rejected. This finding is supportive of the conclusions of Eichenbaum and Hansen (1989), whose work suggests that ignoring durable goods does not seem to represent serious misspecification.

Finally, while the data are unable to resolve the quantitative characteristics of the intertemporal utility function to the degree of precision one might hope for, we are still very encouraged by our experience using SNP techniques for both the law of motion and the utility function. The methods are reasonably practicable and generally give sensible looking results. One extension that could increase precision is to work out a way to follow up on Barnett and Yue (1987) by employing series expansions that always restrict the approximating utility function to be globally concave while keeping the task of making statistical inference manageable. More fundamentally, though, our view is that additional precision can only come from using more assets and data spanning longer time periods with richer historical variation. (We have in mind expanded versions of the long-term data sets employed by Grossman, Melino, and Shiller (1987) and Muoio (1988).) Using such data will entail working out some very difficult problems concerning strategies for controlling dimensionality of parameter spaces and the temporal aggregation of both SNP time series models and nonseparable utility models.

Department of Statistics, Department of Economics, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

and

Department of Economics, Duke University, Durham, NC 27706, U.S.A.

Manuscript received March, 1987; final revision received February, 1989.

APPENDIX

This appendix gives more of the numerical details behind the calculation of the restricted estimations reported in Table IV.

The lattice upon which the constraints are evaluated is generated in such a way that the vectors $x_k \in \mathbb{R}^{L^*}$ are a determining set for a polynomial of degree K_r in $x \in \mathbb{R}^{L^*}$, where $L^* = M \cdot \max(J, L)$. Therefore an evaluation scheme like ours annihilates the coefficients of a polynomial of degree K_r in $x \in \mathbb{R}^{L^*}$. Construction of the determining set follows the algorithm given in Stroud (1971, pp. 54–55). We use as the "basis" numbers a sequence of the form $\{0, .50, -.50, 1.00, -1.00, \dots\}$, which in our case are interpreted as +/− multiples of .50 standard deviations, and generate a determining set $\{z_k\}$ based on these numbers. We then linearly transform the $\{z_k\}$ of this lattice to one in natural units by applying the inverse of the transformation that takes the observed data vectors to vectors with sample mean zero and sample covariance matrix equal to the identity. Since x is a vector of length L^* and the polynomial is of degree K_r, the algorithm generates exactly $K^* = (L^* + K_r)!/(L^*!\,K_r!)$ distinct vectors z_k and x_k. The linear transformation from the $\{z_k\}$ to the $\{x_k\}$ is applied piecewise to the $M \times 1$ sub-blocks of $\{z_k\}$.

The parametric restrictions take the form $\tau(\theta) = 0$ where $\tau(\theta)$ is a vector-valued function comprised of $\tau_l(x_k, \theta)$, with the dependence on x_k suppressed. The dimension of τ is $K^* M_a \times 1$ where K^* is the number of determining x_k and M_a is the number of assets. In all estimations we use a determining set $\{x_k\}$ that annihilates a quadratic, $K_r = 2$, and we use a two-point product Gauss-Hermite numerical quadrature rule to evaluate $\tau_l(x_k, \theta)$. (This rule is exact for polynomials of degree three.)

The redundancy of the restrictions is reflected in the Jacobian matrix

$$\mathscr{J}(\theta) = (\partial/\partial\theta')\tau(\theta)$$

having less than full row rank. To remove the redundancy, we evaluate the Jacobian at the initial value θ_0 for the optimization and calculate an orthogonal matrix $\mathscr{H}$ such that

$$\mathscr{H}\mathscr{J}(\theta_0) = U$$

where U is an upper trapezoidal matrix with nonnegative entries along the diagonal sorted in descending order. The matrix $\mathscr{H}$ is the product of orthogonal (Householder) matrices (Gill, Murray, and Wright (1981)). We then form from $\mathscr{H}$ a matrix $\mathscr{H}_0$ whose rows are consecutive rows of $\mathscr{H}$, starting with the first, such that $\mathscr{H}_0\mathscr{J}(\theta_0)$ is of full row rank. In estimation we impose the restrictions

$$\mathscr{H}_0\tau(\theta) = 0.$$

To implement this reduction in the number of restrictions we need some means of determining the effective number of restrictions r^* at the optimum, so that we know how many rows of $\mathscr{H}$ to use in forming $\mathscr{H}_0$. This is very close to the problem of numerical determination of the rank of a matrix, which is inherently inexact, and is complicated by the fact that we don't know where the optimum is until it is computed.

One estimate of r^* is the number of elements along the diagonal edge of U that exceed machine precision, with the Jacobian $\mathscr{J}(\theta_0)$ computed at the start value for θ. Call this number r_0. Given this putative value r_0 for r^* we can perform a constrained estimation, i.e., maximize the likelihood subject to the restriction $\mathscr{H}_0\mathscr{J}(\theta)=0$, where as before $\mathscr{H}_0$ is the first r_0 rows of $\mathscr{H}$ computed from $\mathscr{J}(\theta_0)$. This gives an estimate $\hat{\theta}_1$.

A second estimate of r^* can be obtained by examining the rank of the bordered information matrix at $\hat{\theta}_1$; this is the matrix that has to be inverted to obtain standard errors. The bordered information matrix is

$$\begin{pmatrix} \mathscr{I}(\hat{\theta}_1) & \mathscr{J}(\hat{\theta}_1)'\mathscr{H}_0' & E' \\ \mathscr{H}_0\mathscr{J}(\hat{\theta}_1) & 0 & 0 \\ E & 0 & 0 \end{pmatrix}$$

where $\mathscr{I}(\hat{\theta}_1)$ is the sample information matrix at $\hat{\theta}_1$, which is computed from the outer product of the gradients while E is a $2\times p_\theta$ matrix that is present because of additional constraints due to the normalizations. The rows of E are unit vectors (that is, have all zeros except for a single one) that reflect the normalizations that the leading term of the density polynomial and the leading term of the utility function polynomial are set to unity; for those estimations reported below where the constraint $\gamma=0$ is imposed, then E has an additional row reflecting that constraint as well. The upper left piece of the inverse of the bordered information at the constrained optimum is the natural estimate of the covariance matrix of $\hat{\theta}$ at a constrained optimum. (See Silvey (1978, Section 4.7); we use the sample information matrix as the estimate of minus the Hessian.) The computed value $\hat{\theta}_1$ may not have yet converged to the constrained optimum, though the rank deficiency provides a second means to calculate an estimate of r^*. That is, r_0 should be reduced by an amount equal to the extent to which the bordered information matrix is not of full rank.

The constrained estimation proceeds iteratively. From a starting estimate $\hat{\theta}_0$, determine $\mathscr{H}_0$, minimize $s_n(\theta)$ subject to $\mathscr{H}_0\tau(\theta)=0$ to get $\hat{\theta}_1$, recompute $\mathscr{H}_0$ and so on, save that at some point in the iterations the number of leading rows from $\mathscr{H}$ that will be chosen to make up $\mathscr{H}_0$ are fixed once and for all. Actually, the computations are remarkably robust to the number of rows that are chosen and, upon termination, the entire vector $\tau(\hat{\theta})$ will have been put to near zero. Various choices of number of rows for $\mathscr{H}_0$ affects how near zero $\tau(\hat{\theta})$ will be upon termination and the stability of the NPSOL algorithm. No statistically significant digits of parameter estimates are much affected. In this discussion, bear in mind that $\mathscr{H}_0\tau(\hat{\theta})$ is always put to zero within a tolerance of 10^{-13} typically and 10^{-9} at worst; this will imply that $\tau(\hat{\theta})$ is zero to within a tolerance of 10^{-8} typically and 10^{-4} at worst. Of the two rank tests, the test on the bordered information matrix is the more sensitive.

We did find that within a class of similar problems the number of rows that ought be in $\mathscr{H}_0$ is not constant. This presents a problem as we wish to compare likelihoods and variation in the number of rows in $\mathscr{H}_0$ would cast doubt on these comparisons. Going to the minimal $\mathscr{H}_0$ within a class is too drastic a step and casts doubt on the claim that the minimization of $s_n(\theta)$ subject to $\tau(\theta)=0$ has actually been carried out. We addressed this difficulty by imposing the median number of constraints within a class. This causes the bordered information matrix to be less than full rank in some instances. As we need to invert it to compute standard errors, we addressed this problem by using the Moore-Penrose g-inverse computed via the Singular Value Decomposition (Kennedy and Gentle (1980, p. 285)). The approach was successful. Computed standard errors vary little with variation in the number of rows in $\mathscr{H}_0$. The chief effect was the precision with which the standard errors of the two polynomial coefficients which had been pegged at one were computed; they ought to be zero exactly.

REFERENCES

ANDREWS, DONALD W. K. (1987): "Asymptotic Normality of Fourier Series Estimators in Nonparametric and Semiparametric Regression," manuscript, Cowles Foundation, Yale University.

BARNETT, W. A. (1981): *Consumer Demand and Labor Supply*. Amsterdam: North Holland.

BARNETT, W. A., AND P. YUE (1987): "Seminonparametric Estimation of the Asymptotically Ideal Model: the AIM Demand System," manuscript, Department of Economics, University of Texas, Austin.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy*, 96, 116–131.

BREEDEN, D. T. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

BUREAU OF THE CENSUS (1985): *Current Population Reports*, Series P-25, No. 975, September 1985. Washington, D.C.: U.S. Government Printing Office.

CRSP (1986): *Monthly Master File: Updated Through 1985*. Chicago, Illinois: Center For Research in Security Prices.

CITIBASE (1983): *Citibank Economic Data Base: 1947–1983*. New York: Citibank, Inc.

CONSTANTINIDES, GEORGE M. (1988): "Habit Formation: A Resolution of the Equity Premium Puzzle," manuscript, Graduate School of Business, University of Chicago.

DEPARTMENT OF COMMERCE (1984): *Business Statistics: 1984*. Washington, D.C.: U.S. Government Printing Office.

DUNN, K., AND K. SINGLETON (1986): "Modeling the Term Structure of Interest Rates Under Non-Separable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

EASTWOOD, BRIAN J., AND A. RONALD GALLANT (1987): "Adaptive Truncation Rules for Seminonparametric Estimators That Achieve Asymptotic Normality," Working Paper 87-53, Graduate School of Business, University of Chicago.

EICHENBAUM, M., AND L. P. HANSEN (1989): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and Economic Statistics*, forthcoming.

EICHENBAUM, M., L. P. HANSEN, AND S. RICHARD (1984): "The Dynamic Equilibrium Pricing of Durable Consumption Goods," manuscript, Department of Economics, University of Chicago.

ELBADAWI, IBRAHIM, A. RONALD GALLANT, AND GERALDO SOUZA (1983): "An Elasticity Can Be Estimated Consistently Without A Priori Knowledge of Functional Form," *Econometrica*, 51, 1731–1752.

ENGLE, ROBERT F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

ENGLE, ROBERT F., AND TIM BOLLERSLEV (1986): "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

GALLANT, A. RONALD (1982): "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20, 285–323.

——— (1987): *Nonlinear Statistical Models*. New York: John Wiley and Sons, Inc.

GALLANT, A. RONALD, AND DOUGLAS W. NYCHKA (1987): "Seminonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

GALLANT, A. RONALD (1987): "Identification and Consistency in Seminonparametric Regression," in *Advances in Econometrics Fifth World Congress*, Volume 1, ed. by Truman F. Bewley. New York: Cambridge University Press, 145–170.

GILL, PHILIP E., WALTER MURRAY, MICHAEL A. SAUNDERS, AND MARGARET H. WHITE (1983): "User's Guide for SOL/NPSOL: A Fortran Package for Nonlinear Programming," Technical Report Number SOL 83-12, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, California.

GILL, PHILIP E., WALTER MURRAY, AND MARGARET H. WHITE (1981): *Practical Optimization*. New York: Academic Press.

GORMAN, W. M. (1953): "Community Preference Fields," *Econometrica*, 21, 63–80.

GRENANDER, ULF (1981): *Abstract Inference*. New York: John Wiley and Sons, Inc.

GROSSMAN, S., A. MELINO, AND R. SHILLER (1987): "Estimation of the Continuous Time Consumption Based Asset Pricing Model," *Journal of Business and Economic Statistics*, 5, 315–327.

HANSEN, LARS PETER (1986): "Comment," *Journal of Business and Economic Statistics*, 4, 418–421.

HANSEN, LARS PETER, AND SCOTT F. RICHARD (1987): "The Role of Conditional Information in Deducing Testable Implications Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–614.

HANSEN, LARS PETER, AND KENNETH J. SINGLETON (1982): "Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–265.

——— (1984): "Errata," *Econometrica*, 52, 267–268.

IBBOTSON ASSOCIATES (1985): *Stocks, Bonds, Bills, and Inflation: 1985 Yearbook*. Chicago, Illinois:

Capital Market Research Center.
KENNEDY, WILLIAM J., JR., AND JAMES E. GENTLE (1980): *Statistical Computing*. New York: Marcel Dekker, Inc.
LUCAS, ROBERT E., JR. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1445.
MEHRA, R., AND E. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–162.
MUOIO, P. (1988): "Estimating the Consumption-Based Asset Pricing Model Using Twentieth Century Temporally Aggregated Data," Ph.D. dissertation, Duke University.
NOVALES, A. (1984): "The Role of Adjustment Costs in Interest Rate Determination," Research Paper No. 284, Department of Economics, State University of New York, Stony Brook.
SEVERINI, THOMAS A., AND WING HUNG WONG (1987): "Convergence Rates of Maximum Likelihood and Related Estimates in General Parameter Spaces," Technical Report No. 207, Department of Statistics, University of Chicago.
SINGLETON, K. J. (1989): "Specification and Estimation of Intertemporal Asset Pricing Models," forthcoming in the *Handbook of Monetary Economics*, ed. by B. Friedman and F. Hahn. Amsterdam: North-Holland.
SILVEY, S. D. (1978): *Statistical Inference*. Cambridge, England: Chapman Hall Ltd.
SIMS, C. A. (1980): "International Evidence on Monetary Factors in Macroeconomic Fluctuations," manuscript, Department of Economics, University of Minnesota.
STROUD, A. H. (1971): *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, New Jersey: Prentice-Hall.