GMM WITH WEAK IDENTIFICATION

By James H. Stock and Jonathan H. Wright[1]

This paper develops asymptotic distribution theory for GMM estimators and test statistics when some or all of the parameters are weakly identified. General results are obtained and are specialized to two important cases: linear instrumental variables regression and Euler equations estimation of the CCAPM. Numerical results for the CCAPM demonstrate that weak-identification asymptotics explains the breakdown of conventional GMM procedures documented in previous Monte Carlo studies. Confidence sets immune to weak identification are proposed. We use these results to inform an empirical investigation of various CCAPM specifications; the substantive conclusions reached differ from those obtained using conventional methods.

KEYWORDS: Instrumental variables, empirical processes, Euler equation estimation, asset pricing.

1. INTRODUCTION

THERE IS CONSIDERABLE EVIDENCE that asymptotic normality often provides a poor approximation to the sampling distributions of generalized method of moments (GMM) estimators and test statistics in designs and sample sizes of empirical relevance in economics. Examples of this discrepancy in estimation of stochastic Euler equations are investigated by Tauchen (1986), Kocherlakota (1990), Neeley (1994), West and Wilcox (1994), Fuhrer, Moore, and Schuh (1995), and Hansen, Heaton, and Yaron (1996); also see the articles in the 1996 special issue of the *Journal of Business and Economic Statistics* on GMM estimation. Depending on the design, the sampling distributions of GMM estimators can be skewed and can have heavy tails, and likelihood ratio tests of the parameter values and tests of overidentifying restrictions can exhibit substantial size distortions. Although these problems are well documented, their source is not well understood.

This paper investigates one possible source of these problems in GMM with instrumental variables: that the instruments are, loosely speaking, only weakly correlated with the relevant first order condition so that the parameters are poorly identified. In the linear simultaneous equations model, it is well known that when the instruments are weak in the sense that they have a low correlation with the included endogenous variables, then the large-sample normal approximations work poorly; see, for example, Anderson and Sawa (1979), Nelson and

Startz (1990), and Maddala and Jeong (1992). Intuition suggests that a similar phenomenon could be important in nonlinear GMM problems. For example, because lagged asset returns and consumption growth have low correlations with current returns and consumption growth in postwar U.S. data, similar problems might arise in nonlinear asset pricing models in which lagged consumption and asset returns are used as instruments for a function of current returns and consumption growth.

We therefore address four questions. First, is it possible to develop an asymptotic theory for nonlinear GMM estimation when some or all of the parameters are weakly identified? Second, does this theory explain the puzzling failures found in Monte Carlo studies of conventional GMM asymptotics? Third, if so, are there alternative econometric procedures that perform reliably even if there is weak identification? Fourth, do these alternative procedures produce different results than the conventional methods in empirical applications?

We find affirmative answers to all four questions. We develop nonstandard asymptotic approximations to the distributions of GMM estimators and test statistics when some or all of the parameters are weakly identified, in a sense made precise in Section 2. In Section 4 we present a Monte Carlo study, modeled on that of Hansen, Heaton, and Yaron (1996), of GMM estimation of the intertemporal consumption capital asset pricing model (CCAPM). In this study, the nonstandard asymptotic approximations generally are found to match closely the finite sample distributions, although the usual normal approximations do not.[2]

The weak-identification asymptotic approximations to the distributions of the GMM estimators depend on nuisance parameters that are typically unknown in empirical applications. Thus these approximating distributions cannot be used directly for inference. Nonetheless, our asymptotic theory does lead to feasible methods for hypothesis testing and for the construction of confidence sets. These methods do not require knowledge of the nuisance parameters and yield asymptotically valid tests and confidence sets even if there are weakly identified parameters. These confidence sets are constructed by direct comparison of an objective function, evaluated over the entire parameter space, to a chi-squared critical value.

These findings are used to guide an empirical investigation of the CCAPM, using aggregate data from the United States, under three specifications of preferences: constant relative risk aversion (CRRA) utility, habit formation/ durability (Dunn and Singleton (1986)), and Epstein-Zin (1989, 1991) preferences. As predicted by our Monte Carlo study, there is considerable evidence of

---

[2] Some Monte Carlo studies have suggested that another possible source of the poor performance of the conventional normal approximation is finite sample discrepancies between the GMM weighting matrix and its population value; see Pagan and Robertson (1997) for a discussion. The alternative asymptotic theory developed in this paper ignores this possibility and simply assumes that this weight matrix is consistent (the details are given in Section 2), so as to focus solely on issues related to weak identification.

weak identification, and the new confidence sets we propose in this paper typically differ from conventional GMM confidence sets.

Although there is a growing literature on instrumental variables estimation with weak identification, it almost exclusively considers linear models. In the linear case, the most closely related paper is Staiger and Stock (1997). They consider single equation estimators and tests, and their main results obtain as a special case of ours. However, the technical approach in this paper is quite different than in Staiger and Stock (1997): their method of taking limits of first order conditions seems not to generalize to the nonlinear case, so we consider instead limits of the GMM objective function directly using empirical process methods. Other recent papers that study weak instruments in the linear case include Hall, Rudebusch, and Wilcox (1996), Pagan and Robertson (1997, 1998), Chamberlain and Imbens (1996), Nelson, Startz, and Zivot (1998), Wang and Zivot (1998), and Shea (1997). Sargan (1983) considered models that are linear in the variables but nonlinear in the parameters, in which the derivative of the population objective function with respect to the parameter vector is not of full rank but the parameters are still locally identifiable in the sense of Fisher (1966). He argued that in this circumstance estimators are consistent but not asymptotically normal, and he used local asymptotic expansions to approximate their distributions. None of these treatments handles GMM Euler equation estimation in the general nonlinear case.

The main theoretical results are laid out in Section 2. These results rely on high level assumptions that accommodate applications to either time series or cross-sectional data. In Section 3, explicit formulas are provided for the special case of single equation estimation in the linear simultaneous equations model. Section 4 reports on a Monte Carlo study of GMM estimation of the parameters of the power utility function in a representative agent model of consumption. An empirical investigation of the CCAPM using U.S. data is reported in Section 5. Section 6 concludes.

2. ASYMPTOTIC REPRESENTATIONS: GENERAL RESULTS

This section provides limiting representations of a GMM estimator with a general weighting matrix when some of the parameters are weakly identified. These general results are then used to obtain somewhat simpler expressions for some specific estimators and test statistics, in particular the one-step and two-step estimators and associated tests and what Hansen, Heaton, and Yaron (1996) term the "continuous updating" estimator.

2.1. *The GMM Estimator*

Let θ be an n-dimensional parameter vector with true value θ_0, which is assumed to be in the interior of the compact parameter space Θ. The true

parameter value satisfies the G equations,

(2.1) $E[h(Y_t, \theta_0) | F_t] = 0,$

where F_t is the information set at time t. Let Z_t be a K-dimensional vector of instruments contained in F_t. The data are $\{(Y_t, Z_t), t = 1, \ldots, T\}$.

The GMM estimator $\hat{\theta}$ minimizes the objective function $S_T(\theta; \bar{\theta}_T(\theta))$ over $\theta \in \Theta$, where

(2.2) $$S_T(\theta; \bar{\theta}_T(\theta)) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]' W_T(\bar{\theta}_T(\theta)) \left[T^{-1/2} \sum_{s=1}^{T} \phi_s(\theta) \right],$$

where $\phi_t(\theta) = h(Y_t, \theta) \otimes Z_t$ and where $W_T(\bar{\theta}_T(\theta))$ is an $O_p(1)$ positive definite $GK \times GK$ weighting matrix. The somewhat cumbersome notation for the weighting matrix allows for various special cases. For the one-step GMM estimator, W_T does not depend on θ and typically does not depend on the data; for example it might be the identity matrix. For the efficient two-step estimator, W_T is computed using a preliminary estimator of θ, in which case $\bar{\theta}_T$ does not depend on θ. For the efficient continuous updating estimator, W_T is continuously evaluated at the parameter values used for the moments, in which case $\bar{\theta}_T(\theta) = \theta$. For some of the test statistics considered below, W_T is evaluated at a fixed hypothesized value of θ, say θ_H; in this case $\bar{\theta}_T(\theta) = \theta_H$. For notational convenience, $\bar{\theta}_T(\theta)$ will simply be denoted $\bar{\theta}_T$ unless the explicit notation is necessary.

We adopt the following additional notation. As is discussed in Section 2.3, some expectation operators E depend on T, but this dependence is suppressed for notational convenience. Let

$$\Psi_T(\theta) = T^{-1/2} \sum_{t=1}^{T} [\phi_t(\theta) - E\phi_t(\theta)],$$

$$\Omega(\theta_1, \theta_2) = \lim_{T \to \infty} E\Psi_T(\theta_1) \Psi_T(\theta_2)',$$

$$Q_{ZZ} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} EZ_t Z_t',$$

$$\Sigma_{hh}(\theta) = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\{[h(Y_t, \theta) - Eh(Y_t, \theta)]$$

$$\times [h(Y_t, \theta) - Eh(Y_t, \theta)]'\}, \quad \text{and}$$

$$\hat{Q}_{ZZ} = T^{-1} \sum_{t=1}^{T} Z_t Z_t';$$

Ω, Σ_{hh}, and Q_{ZZ} are assumed to be finite. If $\text{cov}[h(Y_t, \theta), h(Y_s, \theta) | Z_t, Z_s]$ does not depend on Z_t and Z_s, then the errors are conditionally homoskedastic and

(2.3) $\Omega(\theta_0, \theta_0) = \Sigma_{hh}(\theta_0) \otimes Q_{ZZ}$ (conditional homoskedasticity).

2.2. *Moment Assumptions*

Our approach is to make so-called "high level" assumptions about the properties of the moments that enter the GMM first order conditions. The advantage of making high-level assumptions is that the results cover a wide range of special cases. The disadvantage is that the assumptions must be interpreted, and their plausibility checked, on a case by case basis. This process of interpreting the assumptions in two leading cases (linear IV estimation and the CCAPM) is undertaken in Sections 3 and 4.

The first assumption is that $\Psi_T(\theta_0)$ obeys a central limit theorem:

ASSUMPTION A: $\Psi_T(\theta_0) \overset{d}{\to} N(0, \Omega(\theta_0, \theta_0))$.

This assumption is local in the sense that it pertains to the properties of Ψ_T only at θ_0. This assumption typically will not be satisfied if the instruments are integrated of order one or higher.

The next, stronger assumption is that Ψ_T obeys a functional central limit theorem, so that Ψ_T treated as a function of θ converges to a Gaussian empirical process. Functional central limit theory and the related empirical process literature in econometrics are surveyed by Andrews (1994). Let " $\Rightarrow$ " denote weak convergence of random functions on Θ with respect to the sup norm; see Andrews (1994, Section 2).

ASSUMPTION B: $\Psi_T \Rightarrow \Psi$, *where* $\Psi(\theta)$ *is a Gaussian stochastic process on* Θ *with mean zero and covariance function* $E\Psi(\theta_1)\Psi(\theta_2)' = \Omega(\theta_1, \theta_2)$.

Assumption B implies Assumption A.

It is of course possible to provide primitive conditions that in turn imply Assumption B. One such set of conditions applies to time series applications in which $\phi_t(\theta)$ is m-dependent:

ASSUMPTION B':
  (i) $\phi_t(\theta)$ *is* m-*dependent*;
  (ii) $|\phi_t(\theta_1) - \phi_t(\theta_2)| \le B_t|\theta_1 - \theta_2|$, *where* $\lim_{T \to \infty} T^{-1}\sum_{t=1}^{T} E(B_t^{2+\delta}) < \infty$ *for some* $\delta > 0$;
  (iii) $\sup_{\theta \in \Theta} E|\phi_t(\theta)|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Weak convergence (Assumption B) follows from the convergence of the finite dimensional distributions of $\Psi_T(\theta)$, stochastic equicontinuity, and the total boundedness of Θ (Andrews (1994)). Assumptions B'(i) and B'(ii) imply stochastic equicontinuity (Andrews (1994, Theorems 1 and 2)). Assumptions B'(i) and B'(iii) imply the convergence of the finite dimensional distributions. The boundedness of Θ and Assumption B'(ii) imply that $\phi_t(\theta)$ is totally bounded.

2.3. *Identification, Lack of Identification, and Weak Identification*

Identification in GMM is a combined property of the function $h(Y_t, \theta)$, the instruments Z_t, and the weighting matrix W_T. Identification hinges on whether the population GMM moment conditions are satisfied uniquely. Assume (as we shall below) that W_T is positive definite. Then θ is identified at θ_0 if $\theta = \theta_0$ is the unique solution of $E\phi_t(\theta) = 0$ for $\theta \in \Theta$, where the expectation is, as usual, taken with respect to the true distribution (for which $\theta = \theta_0$), cf. Newey and McFadden (1994, Sec. 2.2.3). An extreme version of lack of identification arises when $E\phi_t(\theta) = 0$ for all $\theta \in \Theta$, in which case we shall say that θ is completely unidentified at θ_0.

Our primary interest is not in whether θ is strictly identified or unidentified, but in the intermediate case in which θ, or a subset of θ, is weakly identified. Because identification is a feature of $E\phi_t(\theta)$, it is natural to characterize weak identification in terms of $E\phi_t(\theta)$. In finite samples, $E\phi_t(\theta)$, while nonzero for $\theta \neq \theta_0$, might be small for a large set of θ, so that the population objective function has large regions of plateaus or ridges; thus the population objective function provides only limited ability to discriminate among a large set of parameter values. If so, it is useful to think of θ (or a subset of θ) as being weakly identified.

Our formal characterization of weak identification starts with an identity. First adopt some additional notation. Partition θ as $\theta = (\alpha', \beta')'$, where $\alpha \in A$ is $n_1 \times 1$ and will be treated as weakly identified, while $\beta \in B$ is $n_2 \times 1$ and will be treated as strongly identified. With this notation, it will at times be convenient to write functions of θ interchangeably as functions of α and β; for example $\Psi_T(\theta)$ and $\Psi_T(\alpha, \beta)$ are equivalent. Also, let $ET^{-1}\sum_{t=1}^{T}\phi_t(\alpha, \beta) = \tilde{m}_T(\alpha, \beta)$. Now write the identity

$$(2.4) \qquad \tilde{m}_T(\alpha, \beta) = \tilde{m}_T(\alpha_0, \beta_0) + \tilde{m}_{1T}(\alpha, \beta) + \tilde{m}_{2T}(\beta)$$

where $\tilde{m}_{1T}(\alpha, \beta) = \tilde{m}_T(\alpha, \beta) - \tilde{m}_T(\alpha_0, \beta)$ and $\tilde{m}_{2T}(\beta) = \tilde{m}_T(\alpha_0, \beta) - \tilde{m}_T(\alpha_0, \beta_0)$. Because $E\phi_t(\alpha_0, \beta_0) = 0$, $\tilde{m}_T(\alpha_0, \beta_0) = 0$, $\tilde{m}_{1T}(\alpha_0, \beta_0) = 0$, and $\tilde{m}_{2T}(\beta_0) = 0$.

The key idea in this paper, made precise in Assumption C below, is to treat $\tilde{m}_{2T}(\beta)$ as large for β outside a neighborhood of β_0, but $\tilde{m}_{1T}(\alpha, \beta)$ as small for all α and β. Thus β can be thought of as well identified, whereas α is weakly identified in the sense that the population moment conditions are zero at (α_0, β_0) but are very nearly zero for $\alpha \neq \alpha_0$. In other words, the population objective function is steep in β around β_0, but is nearly flat in α.

The notion that β is well identified is implemented by assuming that $\tilde{m}_{2T}(\beta)$ satisfies conventional identification conditions in the GMM literature. Specifically, we suppose that $\tilde{m}_{2T}(\beta)$ does not depend on T and (to simplify notation) write $\tilde{m}_{2T}(\beta) = m_2(\beta)$, where $m_2(\beta_0) = 0$, $m_2(\beta) \neq 0$ for $\beta \neq \beta_0$, and $\partial m_2(\beta)/\partial\beta'|_{\beta_0}$ has full column rank. If all the parameters are well identified so that $\theta = \beta$, this becomes the standard set of GMM identification conditions that, with additional conditions on moments and dependence, lead to $T^{1/2}$-con-

sistency and asymptotic normality of the GMM estimator of β (cf. Newey and McFadden (1994)).

If, like $\tilde{m}_{2T}(\beta)$, $\tilde{m}_{1T}(\theta)$ is nonzero for $\theta \neq \theta_0$ and does not depend on T, then the usual asymptotic approximation would obtain for the distribution of the GMM estimator of α. However, the purpose of this paper is to develop a nonstandard distribution theory for α that embodies its weak identification. Whereas conventional asymptotics for estimators of well-identified parameters rests on the assumption that it suffices to consider only those values of β close to β_0, we question whether this is adequate to approximate the distributions of estimators of weakly identified parameters. We therefore adopt the device of linking the expectation of the moment condition to the sample size so that, even asymptotically, the population objective function is finite globally in α, although it is finite in β only in a local neighborhood of β_0. This entails considering a sequence of models in which $E\phi_t(\theta)$ depends on T for $\theta \neq \theta_0$, but the true value θ_0 does not depend on T. Specifically, we adopt the nesting, $\tilde{m}_{1T}(\theta) = T^{-1/2} m_1(\theta) + o(T^{-1/2})$, where $m_1(\theta)$ is bounded. The choice of the $T^{-1/2}$ rate in this nesting yields tractable asymptotic approximations to the sampling distributions of GMM estimators and test statistics that reflect the fact that weak identification results in an objective function that is nearly flat in α.[3]

In the limiting special case that $m_{1T}(\theta)$ does not depend on θ (and thus equals zero), $ET^{-1}\sum_{t=1}^{T}\phi_t(\theta) = m_2(\beta)$, so that α is completely unidentified. This would occur if, for example, $Eh(Y_t, \alpha, \beta)$ does not depend on α and Y_t and Z_t are independent, so that the instruments are irrelevant. This is the so-called partially identified model studied in the linear case by Phillips (1989) and Choi and Phillips (1992).

We thus have the following assumption:

ASSUMPTION C: $ET^{-1}\sum_{t=1}^{T}\phi_t(\theta) = m_{1T}(\theta)/\sqrt{T} + m_2(\beta)$, where:

(i) $m_{1T}(\theta) \to m_1(\theta)$ uniformly in $\theta \in \Theta$, $m_1(\theta_0) = 0$, and $m_1(\theta)$ is continuous in θ and is bounded on Θ;

(ii) $m_2(\beta_0) = 0$, $m_2(\beta) \neq 0$ for $\beta \neq \beta_0$, $R(\beta)$ is continuous, and $R(\beta_0)$ has full column rank, where $R(\beta) = \partial m_2(\beta)/\partial\beta'$ is $GK \times n_2$.

Identification also depends on the weighting matrix W_T, which, as was mentioned above, is assumed to have a positive definite uniform limit.

---

[3] In the terminology of Davidson and MacKinnon (1993), this is a drifting DGP, conceptually akin to the sequence of models used to study local asymptotic power of a test against a Pitman drift. Such local nestings can be a useful device for approximating sampling distributions in the region of knife-edge special cases. An example is the so-called local to unity model of an autoregression with a nearly unit root (Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987)). An unusual feature here is that the local parameter is in general infinite dimensional in the sense that it is the function $m_1(\theta)$.

ASSUMPTION D: W_T is positive definite and $W_T(\theta) \overset{p}{\to} W(\theta)$ uniformly in θ, where $W(\theta)$ is a symmetric nonrandom $GK \times GK$ matrix that is continuous in θ and is positive definite for all $\theta \in \Theta$.

2.4. General Results

Theorem 1 provides a limiting representation for the GMM estimator by first obtaining a limiting empirical process representation for the GMM objective function; let $\hat{\beta}(\alpha)$ solve $\text{argmin}_{\beta \in B} S_T(\alpha, \beta; \bar{\theta}_T(\alpha, \beta))$, let $\hat{\alpha}$ solve $\text{argmin}_{\alpha \in A} S_T(\alpha, \hat{\beta}(\alpha); \bar{\theta}_T(\alpha, \hat{\beta}(\alpha)))$, and let $\hat{\beta} = \hat{\beta}(\hat{\alpha})$.

THEOREM 1: Suppose that Assumptions B, C, and D hold, and that $\bar{\theta}_T(\theta) \Rightarrow \bar{\theta}(\theta)$ uniformly in θ. Then:

$$S_T\big(\alpha, \beta_0 + b/T^{1/2}; \bar{\theta}_T(\alpha, \beta_0)\big)$$
$$\Rightarrow [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0)b]'$$
$$\times W\big(\bar{\theta}(\alpha, \beta_0)\big)[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0)b]$$
$$\equiv \bar{S}\big(\alpha, b; \bar{\theta}(\alpha, \beta_0)\big).$$

(ii) If $\bar{S}(\alpha, b; \bar{\theta}(\alpha, \beta_0))$ has a unique minimum, then $(\hat{\alpha}', T^{1/2}(\hat{\beta} - \beta_0)') \overset{d}{\to} (\alpha^{*\prime}, b^{*\prime})$, where $\alpha^* = \text{argmin}_{\alpha \in A} S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$ and where

$$b^* = -\big[R(\beta_0)' W\big(\bar{\theta}(\alpha^*, \beta_0)\big) R(\beta_0)\big]^{-1}$$
$$\times R(\beta_0)' W\big(\bar{\theta}(\alpha^*, \beta_0)\big)[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0)],$$
$$S^*\big(\alpha; \bar{\theta}(\alpha, \beta_0)\big)$$
$$= [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0)]'$$
$$\times M\big(\alpha, \beta_0; \bar{\theta}(\alpha, \beta_0)\big)[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0)], \quad and$$
$$M\big(\alpha, \beta_0; \bar{\theta}(\alpha, \beta_0)\big) = W\big(\bar{\theta}(\alpha, \beta_0)\big) - W\big(\bar{\theta}(\alpha, \beta_0)\big)$$
$$\times R(\beta_0)\big[R(\beta_0)' W\big(\bar{\theta}(\alpha, \beta_0)\big) R(\beta_0)\big]^{-1}$$
$$\times R(\beta_0)' W\big(\bar{\theta}(\alpha, \beta_0)\big).$$

Proofs are given in Appendix A.

Several remarks are in order. First, although $\hat{\beta}$ is $\sqrt{T}$-consistent, $\hat{\alpha}$ is not consistent but rather is $O_p(1)$. Because $m_1(\theta)$ is finite on Θ, the objective function $S_T(\alpha, \beta_0; \bar{\theta}_T)$ is uniformly $O_p(1)$, so α could not be consistently estimated even if β_0 were known. The lack of consistency of $\hat{\alpha}$ stems from $m_1(\theta)$ being finite under Assumption C.

Second, in general the limiting distributions of $\hat{\alpha}$ and $T^{1/2}(\hat{\beta} - \beta_0)$ are nonstandard. It is not surprising that $\hat{\alpha}$ has a nonnormal limiting distribution in this setting because its limiting representation is the solution to a global, generally nonquadratic rather than a local quadratic minimization problem. The limiting nonnormality of $T^{1/2}(\hat{\beta} - \beta_0)$, which is perhaps more surprising because β is well identified, arises from the inconsistent estimation of α. If $\hat{\alpha}$ were consistent for α_0, then the term $\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0)$ in the limiting expression for b^* would simplify to $\Psi(\alpha_0, \beta_0)$, $\bar{\theta}$ would have a nonrandom probability limit, and b^* would be normally distributed with mean zero and the usual GMM covariance matrix. However, the inconsistent estimation of α implies that in general the population moments are not evaluated within a local neighborhood of α and so impart a nonzero bias to the limiting representation. In the special case that W_T does not depend on $\bar{\theta}_T$, the extent of the asymptotic bias depends on $Em_1(\alpha^*, \beta_0)$, where the expectation is taken over α^*. In general this expectation need not be zero even if α^* is symmetrically distributed around α_0, and in any event the distribution of α^* need not be centered around α_0, so in general this contribution to the bias is nonzero.

Third, as a special case, these results provide limiting representations of the estimators when α is completely unidentified in the sense that $E\phi_t(\alpha, \beta_0)$ does not depend on α, so $m_1(\alpha, \beta_0) = m_1(\alpha_0, \beta_0) = 0$. Then

$$S^*\left(\alpha; \bar{\theta}(\alpha, \beta_0)\right) = \Psi(\alpha, \beta_0)' M\left(\alpha, \beta_0; \bar{\theta}(\alpha, \beta_0)\right) \Psi(\alpha, \beta_0)$$

and $\hat{\alpha} \Rightarrow \alpha^* = \text{argmin}_{\alpha \in A} S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$. Complete characterizations of these distributions depend on Ω, W, and $R(\beta_0)$, which are specific to a given application.

Because the limiting distributions are nonstandard, confidence intervals for β constructed by inverting the quasi-likelihood ratio (LR) statistic $S_T(\theta_0, \bar{\theta}_T(\theta_0)) - S_T(\hat{\theta}, \bar{\theta}_T(\hat{\theta}))$ or the conventional Wald statistic will not in general be valid. However, under weak conditions (Assumption A) confidence intervals can be constructed directly from the objective function. This is a consequence of the following theorems:

THEOREM 2: *Suppose that Assumption A holds and $W_T(\theta_0) \xrightarrow{p} W(\theta_0) = \Omega(\theta_0, \theta_0)^{-1}$. Then*

$$S_T(\theta_0; \theta_0) \xrightarrow{d} \chi^2_{GK}.$$

THEOREM 3: *Suppose that Assumptions B, C, and D hold and that $W(\theta_0) = \Omega(\theta_0, \theta_0)^{-1}$. Then*

$$S_T\left(\alpha_0, \hat{\beta}(\alpha_0); \alpha_0, \hat{\beta}(\alpha_0)\right) \xrightarrow{d} \chi^2_{GK - n_2}.$$

Thus, despite the weak identification, at the true values of the parameters the objective function has a standard asymptotic χ^2 distribution if an efficient weighting matrix is used (efficient in the usual sense that $W_T(\theta_0)$ consistently estimates the inverse of the covariance matrix of $\phi_t(\theta_0)$). Theorem 2 holds under quite weak conditions and does not involve any assumptions about instrument validity except that the moment orthogonality condition $E\phi_t(\theta_0) = 0$ holds; the only assumptions on the properties of sample moments needed for Theorem 2 are ones at the true parameter value. Theorem 3 does not require $m_1(\theta)$ to be nonzero for $\theta \neq \theta_0$, but it does require that β be well identified in the sense of Assumption C. Under these stronger conditions, the concentrated objective function has an asymptotic χ^2 distribution.

Theorem 2 provides a straightforward method for constructing asymptotically valid hypothesis tests and confidence sets, that is, tests and confidence sets with asymptotic size and coverage equal to their respective nominal levels uniformly over the parameter space. To perform an asymptotically valid test of the hypothesis $\theta = \theta_0$, reject if $S_T(\theta_0; \theta_0)$ exceeds the appropriate χ^2_{GK} critical value. To construct an asymptotically valid confidence set, invert the test based on $S_T(\theta_0; \theta_0)$. That is, $\{\theta_0 : S_T(\theta_0, \theta_0) \leq \chi^2_{GK, r}\}$ is an asymptotic $100(1 - r)\%$ confidence set, where $\chi^2_{GK, r}$ is the $100r\%$ critical value of the χ^2_{GK} distribution. Alternatively, a confidence set for α alone can be constructed by inverting the test of $\alpha = \alpha_0$ based on $S_T(\alpha_0, \hat{\beta}(\alpha_0); \alpha_0, \hat{\beta}(\alpha_0))$. Because they are based directly on the objective function S_T, we refer to these confidence sets as S-sets. Note that Theorems 2 and 3 apply only to the continuous updating objective function, and in particular the S-sets cannot be formed using the two-step objective function.

Construction of asymptotically valid confidence sets for subvectors of θ (other than α) or subvectors of α is somewhat more difficult. One approach is to construct a valid $100(1 - r)\%$ set for θ (or α) and to project out the other elements; see, for example, Dufour (1997, Section 5.2). A confidence set thus constructed will be asymptotically conservative, with asymptotic coverage rate of at least $100(1 - r)\%$.

The S-sets are related to standard GMM test statistics. Under conventional asymptotics, $S_T(\theta_0; \theta_0)$ is asymptotically the sum of the LR statistic testing $\theta = \theta_0$ and Hansen's (1982) J statistic testing the overidentifying conditions. Under weak-identification asymptotics, in the special case of the linear simultaneous equations model, S-sets are asymptotically equivalent to confidence sets constructed by inverting the Anderson-Rubin (1949) test statistic.

The S-sets consist of parameter values at which one fails to reject the joint hypothesis that $\theta = \theta_0$ *and* that the overidentifying conditions are valid, that is, $E[h(Y_t, \theta_0)|Z_t] = 0$. This has some appealing consequences, but also requires care in interpretation. If the model is misspecified so that the overidentifying conditions are invalid, S-sets can be null. If the instruments are weak or irrelevant, it is possible that no parameter value will be rejected, that is, the S-sets can contain the entire parameter space. The case of a small but nonempty

S-set is, however, more ambiguous. The S-set could be small either because the model is correctly specified and precisely estimated or because the model is misspecified but the evidence is too weak to reject it entirely.[4]

2.5. *Results for Specific GMM Estimators*

We now provide explicit expressions for some common GMM estimators and their associated test statistics. The estimators differ in their choice of the weighting matrix W_T. Weighting matrices that are asymptotically equivalent under conventional assumptions need not be with weak identification, and indeed can produce substantially different inferences.

The two-step and continuous updating estimators entail construction of an efficient weighting matrix. We consider both heteroskedasticity robust and nonrobust versions of the weighting matrix, respectively V_T^{-1} and $(V_T^N)^{-1}$, where

$$(2.5) \qquad V_T(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} \left[\phi_t(\tilde{\theta}) - \overline{\phi}(\tilde{\theta}) \right] \left[\phi_t(\tilde{\theta}) - \overline{\phi}(\tilde{\theta}) \right]',$$

$$(2.6) \qquad V_T^N(\tilde{\theta}) = \hat{\Sigma}_{hh}(\tilde{\theta}) \otimes \hat{Q}_{ZZ}, \quad \text{where}$$

$$\hat{\Sigma}_{hh}(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} \left[h(Y_t, \tilde{\theta}) - \overline{h}(\tilde{\theta}) \right] \left[h(Y_t, \tilde{\theta}) - \overline{h}(\tilde{\theta}) \right]',$$

$$\overline{\phi}(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} \phi_t(\tilde{\theta}), \quad \text{and}$$

$$\overline{h}(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} h(Y_t, \tilde{\theta}).$$

The one-step estimator, $\hat{\theta}_1$, is computed using $W_T = I_{GK}$. The efficient two-step estimator, $\hat{\theta}_2$, minimizes the objective function with the efficient weight matrix evaluated at the one-step estimator, so in the heteroskedasticity-robust case $W_T(\overline{\theta}(\theta)) = V_T(\hat{\theta}_1)^{-1}$. The efficient continuous updating estimator, $\hat{\theta}_c$, minimizes the objective function with the efficient weight matrix evaluated at the same point as the moments themselves, so $W_T(\overline{\theta}(\theta)) = V_T(\theta)^{-1}$. Accordingly, the one-step, efficient two-step, and efficient continuous updating estimators

---

[4] The S-sets are consistent with the recommendations in Dufour's (1997) study of confidence sets with locally almost unidentified parameters. He provided finite-sample results and did not consider GMM estimation in the general case. In linear instrumental variables estimation with weak instruments and an unbounded parameter space, he showed that Wald-type confidence ellipsoids are bounded with probability one and hence cannot be valid confidence sets. He also pointed out that Anderson-Rubin (1949) sets are unbounded with positive probability and, with fixed instruments and Gaussian errors, they constitute valid confidence sets in finite samples.

respectively are the minimizers of the three objective functions,

$$(2.7) \qquad S_{1T}(\theta) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]' \left[T^{-1/2} \sum_{s=1}^{T} \phi_s(\theta) \right],$$

$$(2.8) \qquad S_{2T}(\theta) = S_T(\theta; \hat{\theta}_1) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]' V_T(\hat{\theta}_1)^{-1} \left[T^{-1/2} \sum_{s=1}^{T} \phi_s(\theta) \right],$$

$$(2.9) \qquad S_{cT}(\theta) = S_T(\theta; \theta) = \left[T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \right]' V_T(\theta)^{-1} \left[T^{-1/2} \sum_{s=1}^{T} \phi_s(\theta) \right].$$

Either for computational convenience or because heteroskedasticity is considered negligible, the two-step and continuous updating estimators could alternatively be computed using the nonrobust covariance matrix V_T^N. These will be referred to as the nonheteroskedasticity robust versions of these estimators; they will be denoted $\hat{\theta}_2^N$ and $\hat{\theta}_c^N$, and their objective functions $S_{2T}^N(\theta)$ and $S_{cT}^N(\theta)$ correspond to (2.8) and (2.9) with V_T replaced by V_T^N.

The quasi likelihood ratio statistics, which test the hypothesis $\theta = \theta_0$, based on the two-step and continuous updating estimators respectively, are

$$(2.10a) \quad LR_2 = S_{2T}(\theta_0) - S_{2T}(\hat{\theta}_2),$$

$$(2.10b) \quad LR_c = S_{cT}(\theta_0) - S_{cT}(\hat{\theta}_c).$$

The J-tests of overidentifying restrictions based on these two estimators reject for large values of the statistics,

$$(2.11a) \quad J_2 = S_{2T}(\hat{\theta}_2),$$

$$(2.11b) \quad J_c = S_{cT}(\hat{\theta}_c).$$

We assume that the weighting matrices in the objective functions are consistent. For some purposes, pointwise consistency is sufficient, while for others, uniform (on Θ) consistency is used. These assumptions are as follows.

ASSUMPTION D': $\hat{Q}_{ZZ} \xrightarrow{p} Q_{ZZ}$, $\hat{\Sigma}_{hh}(\theta_0) \xrightarrow{p} \Sigma_{hh}(\theta_0)$, and $V_T(\theta_0) \xrightarrow{p} \Omega(\theta_0, \theta_0)$.

ASSUMPTION D'': $\hat{Q}_{ZZ} \xrightarrow{p} Q_{ZZ}$, $\hat{\Sigma}_{hh}(\theta) \xrightarrow{p} \Sigma_{hh}(\theta)$, and $V_T(\theta) \xrightarrow{p} \Omega(\theta, \theta)$ uniformly in $\theta \in \Theta$.

The limiting behavior of the objective functions S_{1T}, S_{2T}, and S_{cT} and the associated estimators and test statistics now follow from Theorem 1. To simplify notation, let Ω_θ denote $\Omega(\theta, \theta)$ and let Ω_{α, β_0} denote $\Omega(\theta, \theta)$ evaluated at $\theta = (\alpha', \beta_0')'$. Let $\mu(\alpha) = \Omega_{\alpha, \beta_0}^{-1/2'} m_1(\alpha, \beta_0)$ and let $z(\alpha) = \Omega_{\alpha, \beta_0}^{-1/2'} \Psi(\alpha, \beta_0)$, so that $z(\alpha)$ is a mean-zero, GK-dimensional Gaussian process in α with covariance function $Ez(\alpha_1)z(\alpha_2)' = \Omega_{\alpha_1, \beta_0}^{-1/2'} \Omega((\alpha_1', \beta_0')', (\alpha_2', \beta_0')') \Omega_{\alpha_2, \beta_0}^{-1/2}$ (we adopt the notational convention that $B = B^{1/2'} B^{1/2}$ and $B^{-1} = B^{-1/2} B^{-1/2'}$, where B is any nonsingular symmetric matrix).

COROLLARY 4: *Under Assumptions B, C, and D′, the following representations hold jointly:*

(a) *One-step objective function:*

$$S_{1T}(\alpha, \hat{\beta}_1) \Rightarrow S_1^*(\alpha) = [z(\alpha) + \mu(\alpha)]' Q_1(\alpha)[z(\alpha) + \mu(\alpha)],$$

uniformly in $\alpha \in A$, *where*

$$Q_1(\alpha) = \Omega_{\alpha, \beta_0}^{1/2} \{ I - R(\beta_0)[R(\beta_0)' R(\beta_0)]^{-1} R(\beta_0)' \} \Omega_{\alpha, \beta_0}^{1/2'}.$$

(b) *One-step estimator:*

$$\left(\hat{\alpha}_1', T^{1/2}(\hat{\beta}_1 - \beta_0)' \right) \Rightarrow (\alpha_1^{*\prime}, b_1^{*\prime}), \quad where$$

$$\alpha_1^* = argmin_{\alpha \in A} S_1^*(\alpha) \quad and$$

$$b_1^* = -[R(\beta_0)' R(\beta_0)]^{-1} R(\beta_0)' \Omega_{\alpha_1^*, \beta_0}^{1/2'}[z(\alpha_1^*) + \mu(\alpha_1^*)].$$

(c) *Two-step objective function:*

$$S_{2T}(\alpha, \hat{\beta}_2) \Rightarrow S_2^*(\alpha), \quad where$$

$$S_2^*(\alpha) = [z(\alpha) + \mu(\alpha)]' Q_2(\alpha)[z(\alpha) + \mu(\alpha)], \quad where$$

$$Q_2(\alpha) = \Omega_{\alpha, \beta_0}^{1/2} \{ \Omega_{\alpha_1^*, \beta_0}^{-1} - \Omega_{\alpha_1^*, \beta_0}^{-1} R(\beta_0) [R(\beta_0)' \Omega_{\alpha_1^*, \beta_0}^{-1} R(\beta_0)]^{-1}$$
$$\times R(\beta_0)' \Omega_{\alpha_1^*, \beta_0}^{-1} \} \Omega_{\alpha, \beta_0}^{1/2'}.$$

(d) *Two-step estimator:*

$$\left(\hat{\alpha}_2', T^{1/2}(\hat{\beta}_2 - \beta_0)' \right) \Rightarrow (\alpha_2^{*\prime}, b_2^{*\prime}), \quad where$$

$$\alpha_2^* = argmin_{\alpha \in A} S_2^*(\alpha) \quad and$$

$$b_2^* = -\left[R(\beta_0)' \Omega_{\alpha_1^*, \beta_0}^{-1} R(\beta_0) \right]^{-1} R(\beta_0)' \Omega_{\alpha_1^*, \beta_0}^{-1} \Omega_{\alpha_2^*, \beta_0}^{1/2'}[z(\alpha_2^*) + \mu(\alpha_2^*)].$$

(e) *Continuous updating objective function:*

$$S_{cT}(\alpha, \hat{\beta}_c) \Rightarrow S_c^*(\alpha) = [z(\alpha) + \mu(\alpha)]'$$
$$\times [I - F(\alpha)(F(\alpha)' F(\alpha))^{-1} F(\alpha)']$$
$$\times [z(\alpha) + \mu(\alpha)], \quad where$$

$$F(\alpha) = \Omega_{\alpha, \beta_0}^{-1/2} R(\beta_0).$$

(f) *Continuous updating estimator:*

$$\left(\hat{\alpha}_c', T^{1/2}(\hat{\beta}_c - \beta_0)' \right) \Rightarrow (\alpha_c^{*\prime}, b_c^{*\prime}), \quad where$$

$$\alpha_c^* = \underset{\alpha \in A}{argmin} \, S_c^*(\alpha) \quad and$$

$$b_c^* = -\left[R(\beta_0)' \Omega_{\alpha_c^*, \beta_0}^{-1} R(\beta_0) \right]^{-1} R(\beta_0)' \Omega_{\alpha_c^*, \beta_0}^{-1/2}[z(\alpha_c^*) + \mu(\alpha_c^*)].$$

(g) $LR_2 \Rightarrow \tilde{S}(\alpha_0, 0; \alpha_1^*) - \tilde{S}(\alpha_2^*, b_2^*; \alpha_1^*)$, where

$$\tilde{S}(\alpha, b; a) = [\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0)b]'$$
$$\times \Omega_{a, \beta_0}^{-1}[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0) + R(\beta_0)b];$$

(h) $LR_c \Rightarrow \tilde{S}(\alpha_0, 0; \alpha_0) - \tilde{S}(\alpha_c^*, b_c^*; \alpha_c^*)$;

(i) $J_2 \Rightarrow \tilde{S}(\alpha_2^*, b_2^*; \alpha_1^*)$;

(j) $J_c \Rightarrow \tilde{S}(\alpha_c^*, b_c^*; \alpha_c^*)$.

Limiting representations for the two-step and continuous updating GMM estimators based on the nonheteroskedasticity robust objective function are also readily obtained from Theorem 1. One that will be used in Section 4 is the nonrobust concentrated continuous updating objective function, $S_{cT}^N(\alpha, \hat{\beta}_c)$, which has the limit,

$$(2.12) \quad S_{cT}^N(\alpha, \hat{\beta}_c)$$
$$\Rightarrow [z(\alpha) + \mu(\alpha)]'$$
$$\times \Omega_{\alpha, \beta_0}^{1/2}\left[W^* - W^*R(\beta_0)(R(\beta_0)'W^*R(\beta_0))^{-1}R(\beta_0)'W^*\right]$$
$$\times \Omega_{\alpha, \beta_0}^{1/2'}[z(\alpha) + \mu(\alpha)],$$

where $W^* = \Sigma_{hh}(\alpha, \beta_0) \otimes Q_{ZZ}$.

2.6. *The Unidentified Case and Measures of Identification*

If $E\phi_t(\alpha, \beta_0) = 0$ for all α, then α is completely unidentified and $m_1(\alpha, \beta_0) = 0$, so $\mu(\alpha) = 0$ for all α. In this case, the expressions above simplify and it becomes possible to make some general comments about the behavior of these estimators. First consider the concentrated continuous updating objective function, $S_{cT}(\alpha, \hat{\beta}_c)$. In the unidentified case, this has the limit, $S_c^*(\alpha) = z(\alpha)'[I - F(\alpha)(F(\alpha)'F(\alpha))^{-1}F(\alpha)']z(\alpha)$. Evidently, for fixed α, $S_c^*(\alpha)$ is distributed $\chi^2_{GK-n_2}$, so S_c^* may be considered a chi-squared process indexed by α. If α is not unidentified but rather is weakly identified, then $\mu(\alpha)$ is nonzero for $\alpha \neq \alpha_0$, and for fixed α, $S_c^*(\alpha)$ is distributed as a noncentral $\chi^2_{GK-n_2}$ random variable with noncentrality parameter $\mu(\alpha)'\mu(\alpha)$. Thus S_c^* can be thought of as following a noncentral $\chi^2_{GK-n_2}$ process.

Consideration of the unidentified case suggests that the two-step estimator of α will be biased towards the probability limit of the nonlinear least squares (NLS) estimator, with the bias increasing as $\mu(\alpha)'\mu(\alpha)$ decreases. This parallels the linear simultaneous equations case, in which TSLS is biased towards the probability limit of the OLS estimator. To see this for GMM, consider the nonrobust estimator with $G = 1$ when all the coefficients are weakly identified, so $\theta = \alpha$, and suppose that $Eh(Y_t, \alpha) = 0$. The NLS objective function is $S_{nls}(\alpha) = T^{-1}\sum_{t=1}^{T} h(Y_t, \alpha)^2$, and $S_{nls}(\alpha) \overset{p}{\to} \Sigma_{hh}(\alpha)$ uniformly in α. The coun-

terpart to the result in Corollary 4(c) for the nonrobust estimator simplifies in this case because the terms in $R(\beta_0)$ vanish, and

$$S_{2T}^N(\alpha) \Rightarrow [z(\alpha) + \mu(\alpha)]'\Omega_\alpha^{1/2}[\Sigma_{hh}(\alpha_1^*)Q_{ZZ}]^{-1}\Omega_\alpha^{1/2\prime}[z(\alpha) + \mu(\alpha)]$$

where $\Omega_\alpha = \Sigma_{hh}(\alpha)Q_{ZZ}$. If α is unidentified, $\mu(\alpha) = 0$ so $S_{2T}^N(\alpha) \Rightarrow z(\alpha)'z(\alpha)(\Sigma_{hh}(\alpha)/\Sigma_{hh}(\alpha_1^*))$. Because $\Sigma_{hh}(\alpha_1^*)$ is a scalar that does not depend on α, this factor can be ignored for the minimization. Because $Ez(\alpha)'z(\alpha) = K$, in expectation the limiting objective function is proportional to the probability limit of $S_{nls}(\alpha)$. This suggests that the minimizer of $S_{2T}^N(\alpha)$ will be biased towards the probability limit of the NLS estimator.

This discussion also suggests that the function $\mu(\alpha)'\mu(\alpha)$ is a population measure of the strength of identification. In single equation estimation in the linear simultaneous equations model (examined in the next section) when $n = 1$, $\mu(\alpha)'\mu(\alpha)$ is quadratic in $\alpha - \alpha_0$ and $\int_A \mu(\alpha)'\mu(\alpha)\, d\alpha / \int_A (\alpha - \alpha_0)'(\alpha - \alpha_0)\, d\alpha$ (where A is symmetric around α_0) equals the so-called concentration parameter which governs the rate of convergence of the finite sample distributions of the two stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators to their Gaussian limits (e.g. Anderson (1977)). This provides a simple one dimensional summary of the quality of identification in this case. In general, however, the dependence of $\mu(\alpha)$ on α is complicated, and $\mu(\alpha)'\mu(\alpha)$ need not be monotone increasing in $|\alpha - \alpha_0|$. This introduces the possibility of multiple peaks in the pdf of the continuous updating estimator even if $\mu(\alpha)'\mu(\alpha)$ is steep for α close to α_0. This suggests that, in general, a full characterization of the extent of weak identification requires global knowledge of $\mu(\alpha)'\mu(\alpha)$.

3. SINGLE-EQUATION LINEAR INSTRUMENTAL VARIABLES ESTIMATION

In this section the results of Section 2 are specialized to the estimation of a single equation in the linear simultaneous equations model. In this case, the two-step estimator is TSLS and the continuous updating estimator is LIML.

There is a large literature on exact distribution theory of instrumental variables estimators in the linear model; see Phillips (1983) for a review. Let y_t be the dependent variable in the equation of interest and let $\tilde{Y}_t$ denote the n other endogenous variables included in that equation. Suppose that n_1 of these variables, $\tilde{Y}_A$, have a small correlation with the instruments, while the remaining n_2 variables, $\tilde{Y}_B$, have a large correlation. The coefficients on $\tilde{Y}_A$ will be treated as weakly identified, while the coefficients on $\tilde{Y}_B$ will be treated as well identified in the sense of Section 2. Suppose, for notational convenience, that the equation of interest contains no exogenous variables (this is readily relaxed using standard projection arguments). The equation of interest and the equation relating the instruments Z_t to $\tilde{Y}_t$ are, in matrix form,

(3.1) $\qquad y = \tilde{Y}\theta + u = \tilde{Y}_A\alpha_0 + \tilde{Y}_B\beta_0 + u,$

(3.2) $\qquad \tilde{Y} = Z\Pi + V = Z\begin{bmatrix} \Pi_A & \Pi_B \end{bmatrix} + \begin{bmatrix} V_A & V_B \end{bmatrix},$

where $U_t = [u_t V_t']'$ satisfies $E(U_t|Z_t) = 0$, and Π and V are partitioned conformably with $\tilde{Y}$. We follow the convention in that literature and assume that U_t is serially uncorrelated and homoskedastic, so $E(U_t U_s'|Z_t, Z_s) = \Sigma_{UU}$ if $t = s$, else $= 0$.

In the notation of Section 2, $h(Y_t, \theta) = y_t - \tilde{Y}_t'\theta$, $\phi_t(\theta) = (y_t - \tilde{Y}_t'\theta)Z_t$, $\Sigma_{hh}(\theta) = \lim_{T \to \infty} T^{-1}\sum_{t=1}^{T} \text{var}(y_t - \tilde{Y}_t'\theta)$, $\hat{\Sigma}_{hh}(\theta) = T^{-1}\sum_{t=1}^{T}[h(Y_t, \theta) - \bar{h}(\theta)]^2$, and $\Omega(\theta_0, \theta_0) = \Sigma_{hh}(\theta_0)Q_{ZZ}$. The objective functions $S_{2T}^N(\theta)$ and $S_{cT}^N(\theta)$ can be written

$$(3.3) \qquad S_{2T}^N(\theta) = (y - \tilde{Y}\theta)' P_Z (y - \tilde{Y}\theta) / \hat{\Sigma}_{hh}(\hat{\theta}_1),$$

$$(3.4) \qquad S_{cT}^N(\theta) = T\{1 + A_T(\theta)^{-1}\}^{-1},$$

where $A_T(\theta) = (y - \tilde{Y}\theta)' P_Z (y - \tilde{Y}\theta)/(y - \tilde{Y}\theta)' M_Z (y - \tilde{Y}\theta)$, $P_Q = Q(Q'Q)^{-1}Q'$, and $M_Q = I - P_Q$ for any full rank $a \times b$ matrix, Q, $a \geq b$. Evidently $\hat{\theta}_2 = (\tilde{Y}'P_Z\tilde{Y})^{-1}\tilde{Y}'P_Z y$ is the TSLS estimator. Since the LIML estimator minimizes $A_T(\theta)$ and $S_{cT}^N(\theta)$ is an increasing function of $A_T(\theta)$, $\hat{\theta}_c$ is the LIML estimator.

It is useful to translate Assumptions A–C and D″ into more transparent assumptions which are tailored to this model. Accordingly, suppose that sample moments involving u, V, and Z converge in probability to their expectations and that $T^{-1/2}\sum_{t=1}^{T} U_t \otimes Q_{ZZ}^{-1/2'} Z_t \xrightarrow{d} \xi \sim N(0, \Sigma_{UU} \otimes I_K)$. Then, by direct calculation, $\hat{\Sigma}_{hh}(\theta) \xrightarrow{p} \Sigma_{hh}(\theta)$, $V_T^N(\theta) \xrightarrow{p} \Sigma_{hh}(\theta)Q_{ZZ}$, and

$$(3.5) \qquad \Psi_T(\theta) = T^{-1/2} \sum_{t=1}^{T} (\phi_t(\theta) - E\phi_t(\theta))$$

$$= T^{-1/2} \sum_{t=1}^{T} \left(\begin{bmatrix} 1 & (\theta_0 - \theta)' \end{bmatrix} U_t \right) Z_t$$

$$\Rightarrow \left(\begin{bmatrix} 1 & (\theta_0 - \theta)' \end{bmatrix} \otimes Q_{ZZ}^{1/2'} \right) \xi.$$

Because the primitive moments do not involve θ and the various functionals are continuous in θ, all limits are automatically uniform in θ on Θ, and Assumptions A, B, and D″ follow.

Translating Assumption C requires making the notion of weakly correlated asymptotics concrete in this model. Direct calculation reveals that

$$(3.6) \qquad ET^{-1} \sum_{t=1}^{T} \phi_t(\theta)$$

$$= Q_{ZZ}\Pi(\theta_0 - \theta) = Q_{ZZ}\Pi_A(\alpha_0 - \alpha) + Q_{ZZ}\Pi_B(\beta_0 - \beta).$$

Assumption C is satisfied by setting $\Pi_A = T^{-1/2}C_A$ and $\Pi_B = C_B$, where C_A and C_B are fixed matrices with dimensions $K \times n_1$ and $K \times n_2$, respectively; then $m_1(\theta) = Q_{ZZ}C_A(\alpha_0 - \alpha)$ and $m_2(\beta) = Q_{ZZ}C_B(\beta_0 - \beta)$. In the special case that all parameters are weakly identified so that $\theta = \alpha$, then the term in Π_B is not present in (3.6) and Assumption C reduces to $\Pi = T^{-1/2}C_A$, which is the nesting used in Staiger and Stock (1997, Assumption L_{Π}).

The linearity of this model permits considerable simplification of the formal limits in Theorem 1 and Corollary 4. Consider the TSLS estimator. Partition ξ as $(\xi_u' \; \text{vec}(\xi_{V_A})' \; \text{vec}(\xi_{V_B})')' = (\xi_u' \; \text{vec}(\xi_V)')'$, where ξ_u is $K \times 1$, ξ_{V_A} is $K \times n_1$, and ξ_{V_B} is $K \times n_2$, so ξ_V is $K \times n$. For the TSLS estimator, $W(\theta) = Q_{ZZ}^{-1}$, $R(\beta) = -Q_{ZZ}C_B$, and

$$\Psi_T(\theta) = T^{-1/2} \sum_{t=1}^{T} [u_t + (\theta_0 - \theta)' V_t] Z_t \quad \text{so}$$

$$\Psi_T(\theta) \Rightarrow \Psi(\theta) = Q_{ZZ}^{1/2\prime}[\xi_u + \xi_V(\theta_0 - \theta)] \quad \text{and}$$

$$\Psi(\alpha, \beta_0) = Q_{ZZ}^{1/2\prime}[\xi_u + \xi_{V_A}(\alpha_0 - \alpha)].$$

Define $\lambda_A = Q_{ZZ}^{1/2} C_A$ and $\lambda_B = Q_{ZZ}^{1/2} C_B$. Substituting these expressions into the formulas in Theorem 1, we obtain

$$(3.7) \qquad S_2^*(\alpha) = \left[\xi_u + (\lambda_A + \xi_{V_A})(\alpha_0 - \alpha) \right]'$$
$$\times M_{\lambda_B}\left[\xi_u + (\lambda_A + \xi_{V_A})(\alpha_0 - \alpha) \right] / \Sigma_{hh}(\alpha_1^*, \beta_0).$$

Thus $\hat{\alpha}_{TSLS} \Rightarrow \alpha_2^* = \text{argmin}_\alpha S_2^*(\alpha)$. Because $S_2^*(\alpha)$ is quadratic in α, this minimization can be carried out analytically; this yields

$$(3.8) \qquad \hat{\alpha}_{TSLS} \Rightarrow \alpha_{TSLS}^* = \alpha_0 + \left[(\lambda_A + \xi_{V_A})' M_{\lambda_B}(\lambda_A + \xi_{V_A})\right]^{-1}(\lambda_A + \xi_{V_A})'$$
$$\times M_{\lambda_B}\xi_u,$$

$$(3.9) \qquad T^{1/2}\left(\hat{\beta}_{TSLS} - \beta_0\right) \Rightarrow (\lambda_B'\lambda_B)^{-1}\lambda_B'\left[\xi_u - (\lambda_A + \xi_{V_A})(\alpha_{TSLS}^* - \alpha_0)\right].$$

Two special cases of (3.8) and (3.9) can be found in the literature. First, when $\Pi_A = 0$, α is unidentified and the model reduces to the partially identified case considered by Choi and Phillips (1992), and (3.8) and (3.9) reduce to Corollary 3.1 in that paper. Second, in the special case that all coefficients are weakly identified,

$$\hat{\alpha}_{TSLS} - \alpha_0 \Rightarrow \left[(\lambda_A + \xi_{V_A})'(\lambda_A + \xi_{V_A})\right]^{-1}(\lambda_A + \xi_{V_A})' \xi_u,$$

which, upon setting $\lambda_A = \lambda$ and $\xi_{V_A} = \xi_V$, is the limiting representation in Staiger and Stock (1997, Thm. 1).

Clearly the linear model permits a substantial simplification, relative to the general results in Section 2. The proof of uniform convergence (the verification of Assumption B) is straightforward because θ does not enter the primitive sample moments, so uniform convergence follows from finite dimensional convergence and the continuous mapping theorem. For the same reason, the stochastic process $z(\alpha)$ is linear in α, which in turn leads to the relatively simple expressions (3.8) and (3.9).

4. THE INTERTEMPORALLY SEPARABLE CCAPM: MONTE CARLO EVIDENCE

This section reports numerical results for the prototype consumption-based asset pricing model, the representative agent intertemporally separable CCAPM model with CRRA preferences. Two sets of questions are addressed. Does this new asymptotic theory explain the puzzling failures of conventional GMM asymptotics found in previous Monte Carlo studies, and does it provide better approximations to the finite sample distributions than the usual Gaussian distributions?

4.1. *Theoretical Considerations*

With CRRA preferences, the G Euler equations are (2.1) with

$$(4.1) \qquad h(Y_t, \theta) = \delta(C_{t+1}/C_t)^{-\gamma} R_{t+1} - \iota_G \quad \text{(CRRA)}$$

where δ is a discount factor, C_t is consumption, R_t is a $G \times 1$ vector of asset returns, and ι_G is the $G \times 1$ vector of ones (cf. Hansen and Singleton (1982)). Moment conditions are available using a constant and lagged variables as instruments. In the notation of Section 2, $\phi_t(\theta) = [\delta R_{t+1}(C_{t+1}/C_t)^{-\gamma} - \iota_G] \otimes Z_t$, where $\theta = (\gamma, \delta)'$. The parameters are assumed to be bounded by $\delta_{\min} \leq \delta \leq \delta_{\max}$ and $\gamma_{\min} \leq \gamma \leq \gamma_{\max}$.

The first step is to provide primitive assumptions that imply the high-level Assumptions A, B, C, and D″. Assumption A holds under standard conditions in the GMM literature; cf. Newey and McFadden (1994). If $((C_{t+1}/C_t), R_{t+1}, Z_t)$ are m-dependent, then Assumption B′(i) is satisfied. Also, $E|R_{t+1} \otimes Z_t|^5 < \infty$ and $E\exp(5(\gamma_{\max} + 1)|c_{t+1}|) < \infty$, where $c_{t+1} \equiv \ln(C_{t+1}/C_t)$, imply B′(ii) and B′(iii), so Assumption B′ is satisfied, which implies Assumption B. Assumption D″, which implies Assumption D, holds if $((C_{t+1}/C_t), R_{t+1}, Z_t)$ also have enough moments.

Assumption C is satisfied by treating γ as weakly identified and δ as strongly identified; in the notation of Section 2, $\alpha = \gamma$ and $\beta = \delta$. Specific formulas for implementing this assumption are given in Appendix B. The motivation for the different treatment of δ and γ comes from the structure of the first order conditions. First suppose $G = 1$. Given γ, δ can be estimated precisely from the sample mean, $\hat{\delta}(\gamma) = [T^{-1}\sum_{t=1}^{T}(C_{t+1}/C_t)^{-\gamma} R_{t+1}]^{-1}$. Under the assumptions in the previous paragraph, $\hat{\delta}(\gamma)$ is $T^{1/2}$-consistent for any fixed γ. In this sense, δ is strongly identified by a constant, which is one of the instruments. When $G > 1$, if $\text{cov}((C_{t+1}/C_t)^{-\gamma}, R_{t+1})$ is nearly zero, then the additional first order conditions with a constant as the instrument arguably will not result in improved estimation. If the covariance between $(C_{t+1}/C_t)^{-\gamma} R_{t+1}$ and a stochastic instrument is small, things will not be improved by adding that instrument. It should be stressed that the success of this weak/strong treatment of γ and δ can and will be ascertained numerically. For example, if both δ and γ are appropriately modeled as weakly identified, the approximations will be less satisfactory than

they could be. On the other hand, if both δ and γ are appropriately modeled as strongly identified, they will have the usual joint normal distribution and the weak-identification approximation will offer no improvement.

4.2. *Data Generation and Estimation Equations*

The Monte Carlo design follows Tauchen (1986), Kocherlakota (1990), and Hansen, Heaton, and Yaron (1996). The artificial data were generated by the method discussed by Tauchen and Hussey (1991), which was used by Kocherlakota (1990) and Hansen, Heaton, and Yaron (1996). This method fits a 16 state Markov chain to the law of motion of consumption and dividend growth, calibrated so as to approximate a Gaussian VAR(1).[5] The consumption CAPM with CRRA preferences is then used to price stocks and a risk-free bond in each time period, thereby yielding a time series of asset returns.

Four combinations of true parameters values and estimation equations are studied. Let c_t, r_t^f, and r_t^s denote consumption growth, the risk-free rate, and the stock return. These combinations, or models, are as shown in Table I. Models M1b, M2, and M3 were selected as representative of models that previously have been found to produce nonnormal estimator distributions. Kocherlakota (1990) studied models M1a and M1b. Hansen, Heaton, and Yaron (1996) studied models M1a, M2, and M3.

Preliminary simulations indicated that whether a heteroskedasticity robust or nonrobust covariance matrix is used makes only a small difference for the distribution of the estimator and test statistics. All the results reported here are based on the nonrobust covariance matrix, which is faster to compute. In this design errors are martingale difference sequences at the true values (there are no overlapping data) so a correction for autocorrelation is not used. Each Monte Carlo draw from the finite-sample distribution required numerical optimization over (γ, δ).[6]

TABLE I

Model	(γ_0, δ_0)	Interest Rate(s) in First Order Condition(s)	Instruments
M1a	(1.3, .97)	r_t^s	$1, r_{t-1}^s, c_{t-1}$
M1b	(13.7, 1.139)	r_t^s	$1, r_{t-1}^s, c_{t-1}$
M2	(1.3, .97)	r_t^s, r_t^f	$1, r_{t-1}^s, r_{t-1}^f, c_{t-1}$
M3	(1.3, .97)	r_t^s, r_t^f	$1, c_{t-1}$

[5] Let A denote the VAR matrix (with A_{dc} the coefficient on consumption growth in the dividend growth equation, etc.), and let f and H be the intercept vector and error variance-covariance matrix in the VAR. Following Hansen, Heaton, and Yaron (1996), the values of the parameters used are $(A_{dd}, A_{dc}, A_{cd}, A_{cc}) = (.117, .414, .017, -.161)$, $(f_d, f_c) = (.004, .021)$, and $(H_{dd}, H_{dc}, H_{cc}) = (.014, .00177, .0012)$.

[6] Numerical optimization details and the results of numerical sensitivity checks are available from the authors upon request.

TABLE II

Summary Measures of Estimator and Test Statistic Distributions: Monte Carlo, Weak Instrument Asymptotic, and Normal Asymptotic Distributions:

	γ				δ				J	LR	$S_{cT}(\gamma_0, \delta_0)$	$S_{cT}(\gamma_0, \hat{\delta}(\gamma_0))$
	10%	Median	90%	$KS(\gamma)$	10%	Median	90%	$KS(\delta)$				
A. Model M1a: $\delta_0 = 0.97$, $\gamma_0 = 1.3$												
Two Step												
Monte Carlo	-1.284	1.646	4.359	—	0.917	0.976	1.028	—	3.2%	5.2%	—	—
Weak Iden.	-1.139	1.750	4.538	0.03	0.918	0.978	1.024	0.02	3.3%	4.5%	—	—
Normal	-1.800	1.300	4.400	0.09	0.910	0.970	1.030	0.09	10.0%	10.0%	—	—
Continuous Updating												
Monte Carlo	-4.685	1.368	5.041	—	0.836	0.969	1.033	—	3.5%	14.2%	10.1%	9.3%
Weak Iden.	-5.718	1.325	4.912	0.02	0.791	0.968	1.026	0.02	3.3%	14.0%	10.0%	10.0%
Normal	-1.800	1.300	4.400	0.10	0.910	0.970	1.030	0.12	10.0%	10.0%	10.0%	10.0%
B. Model M1b: $\delta_0 = 1.139$, $\gamma_0 = 13.7$												
Two Step												
Monte Carlo	5.664	9.470	16.052	—	1.029	1.091	1.164	—	21.3%	40.1%	—	—
Weak Iden.	5.996	9.968	16.377	0.07	1.030	1.095	1.164	0.05	23.4%	36.8%	—	—
Normal	3.852	13.700	23.542	0.34	1.052	1.139	1.226	0.34	10.0%	10.0%	—	—
Continuous Updating												
Monte Carlo	8.374	12.930	42.315	—	0.867	1.104	1.188	—	7.2%	10.8%	10.3%	9.4%
Weak Iden.	8.510	13.702	51.858	0.06	0.305	1.102	1.187	0.09	6.8%	11.3%	10.0%	10.0%
Normal	3.852	13.700	23.542	0.14	1.052	1.139	1.226	0.25	10.0%	10.0%	10.0%	10.0%

C. Model M2: $\delta_0 = 0.97$, $\gamma_0 = 1.3$

Two Step

Monte Carlo	−0.904	0.814	3.611	—	0.924	0.960	1.001	—	10.3%	25.2%	—	—
Weak Iden.	−0.481	0.937	3.899	0.05	0.932	0.961	1.003	0.04	16.1%	29.0%	—	—
Normal	0.348	1.300	2.252	0.24	0.954	0.970	0.986	0.30	10.0%	10.0%	—	—

Continuous Updating

Monte Carlo	0.756	1.308	4.651	—	0.960	0.969	1.025	—	9.3%	11.0%	9.8%	9.2%
Weak Iden.	0.687	1.286	4.315	0.05	0.959	0.970	1.015	0.05	10.5%	12.5%	10.0%	10.0%
Normal	0.348	1.300	2.252	0.16	0.954	0.970	0.986	0.14	10.0%	10.0%	10.0%	10.0%

D. Model M3: $\delta_0 = 0.97$, $\gamma_0 = 1.3$

Two Step

Monte Carlo	−2.125	1.256	5.406	—	0.905	0.966	1.030	—	4.9%	16.3	—	—
Weak Iden.	−1.581	1.361	5.364	0.04	0.912	0.967	1.023	0.02	5.9%	21.2%	—	—
Normal	0.292	1.300	2.308	0.23	0.953	0.970	0.987	0.19	10.0%	10.0%	—	—

Continuous Updating

Monte Carlo	0.728	1.297	4.809	—	0.960	0.969	1.026	—	9.7%	11.2%	10.6%	10.5%
Weak Iden.	0.586	1.296	4.595	0.08	0.957	0.970	1.018	0.07	10.1%	10.9%	10.0%	10.0%
Normal	0.292	1.300	2.308	0.16	0.953	0.970	0.987	0.15	10.0%	10.0%	10.0%	10.0%

Notes: δ is treated as strongly identified and γ is treated as weakly identified. The columns headed "γ" and "δ" summarize the distributions of the estimators of these parameters. Kolmogorov-Smirnov statistics compare the Monte Carlo distribution with the asymptotic approximation in the relevant row. The columns labeled "J," "LR," "$S_{cT}(\gamma_0, \delta_0)$," and "$S_{cT}(\gamma_0, \hat{\delta}(\gamma_0))$" report rejection rates of these four test statistics at the nominal (standard asymptotic) 10% level, where the test statistics are described in the text.

Finite sample distributions were computed using 5000 Monte Carlo replications. Computation of the weak-identification asymptotic approximation is discussed in Appendix B.

4.3. *Results*

The results are summarized in Table II for $T = 100$. The finite-sample distributions diverge substantially from the asymptotic normal approximation for models M1b, M2, and M3. In almost all cases, the weak-instrument asymptotics provides a much better approximation than the normal approximation, as measured by the quantiles and the Kolmogorov-Smirnov statistic (the maximum absolute difference between the two empirical cdfs). The weak-identification asymptotic approximations also match the rejection rates of the J and LR statistics.

The final two columns in Table II present the rejection rates of the test statistics used to form the S-sets (the S-set coverage rate is one minus this rejection rate). The first statistic tests the joint hypothesis that γ and δ take on their true values, according to Theorem 2, and the second statistic tests the hypothesis that γ takes on its true value based on the concentrated (over δ) continuous updating objective function, according to Theorem 3. In each of the designs the finite-sample size of both these test statistics is very close to the nominal size of 10%.

Cumulative distribution functions for the two-step and continuous updating estimators of δ and γ are presented in Figure 1 for model M1b. Evidently, the weak-identification asymptotic approximation captures the main qualitative features of the finite-sample distribution, while the normal approximation typically does not.[7]

In Section 2 it was predicted that, as $\mu(\alpha)'\mu(\alpha)$ gets large, the weak-instrument asymptotic distribution will approach the usual Gaussian limit, and the LR and J statistics will approach their usual χ^2 distributions. In contrast, as $\mu(\alpha)'\mu(\alpha)$ decreases, $\hat{\theta}_2$ was predicted to be biased towards the probability limit of the NLS estimator, and the distribution of $\hat{\theta}_2$ was predicted to be tighter than that of $\hat{\theta}_c$. These predictions are explored here by exploiting a scaling property of the weak-identification approximation that permits ready computa-

---

[7] Results were also computed for $T = 50$, although these are not tabulated here to save space. Relative to the $T = 100$ results, the performance of both the conventional normal and the weak-identification asymptotic approximations deteriorates when $T = 50$. Although the weak instrument approximation generally provides a good approximation to the central tendency of the distributions, the Monte Carlo distributions generally have heavier tails than the weak instrument approximations, and the Kolmogorov-Smirnov statistics for the weak-identification approximation are greater for $T = 50$ than $T = 100$ by between .01 and .03. The weak instrument approximation nonetheless does considerably better than the normal approximation in all cases considered.
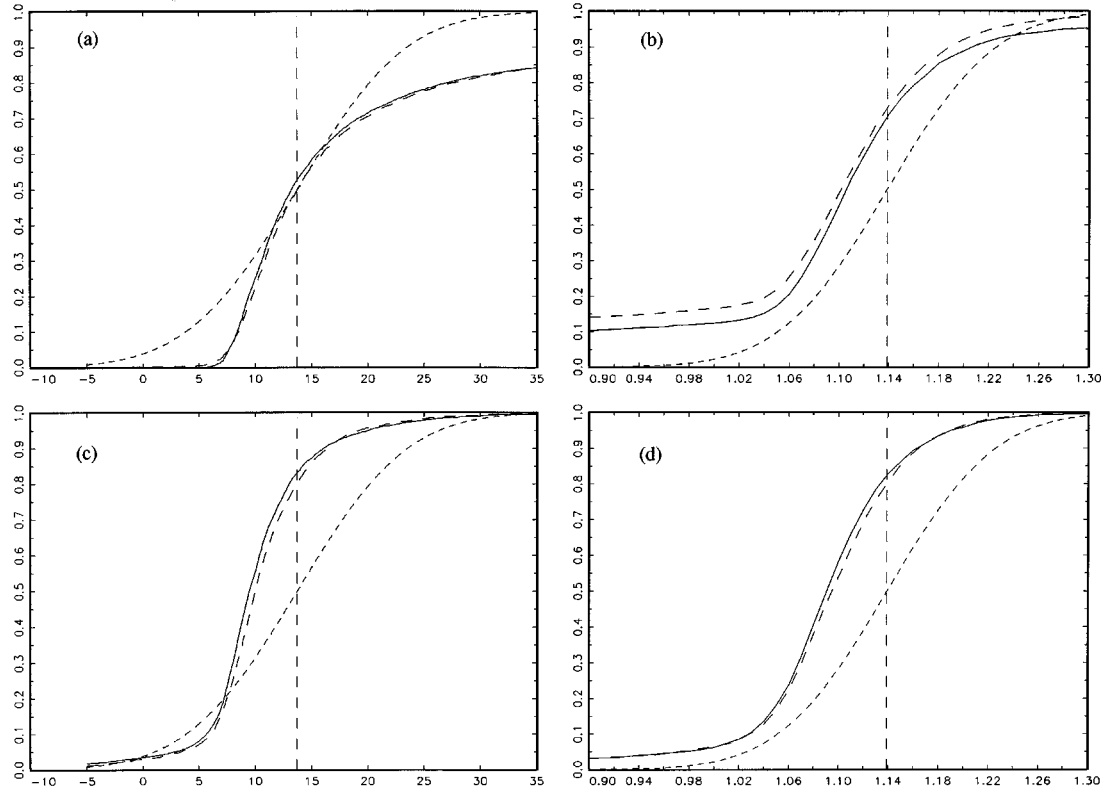
FIGURE 1.—Estimator cdf's for model M1b: Finite sample monte carlo (solid line), standard normal asymptotic (short dashes), and weak instrument asymptotic (long dashes). *Vertical dashed line denotes true value.* (a) $\hat{\gamma}$, continuous updating; (b) $\hat{\delta}$, continuous updating; (c) $\hat{\gamma}$, two-step; (d) $\hat{\delta}$, two-step.

tion of the approximation for alternative sample sizes once it has been computed for an initial sample size.[8]

The results, summarized in Table III, are consistent with these predictions. For small T, and thus small $\mu(\alpha)/\mu(\alpha)$, the distribution of $\hat{\gamma}_2$ is tighter than that of $\hat{\gamma}_c$. In models M1a and M1b the median of $\hat{\gamma}_2$ is strongly biased towards the probability limit of the NLS estimator, which is 2.39 for M1a and 3.91 for M1b. As T increases this median shifts towards γ_0. For small T, the J and LR statistics can have major size distortions, but as T increases their sizes approach the desired 10% level. For the weak-identification and normal approximations to converge, as measured by a Kolmogorov-Smirnov statistic of .05 or less, requires $T \cong 1000$ in M1a and 10,000 in M1b, M2, and M3. Because the weak-identification approximation was found to work well in these designs, this suggests that approximately a century of monthly data are needed before conventional normal asymptotics provides a good approximation to the finite sample distributions of these estimators.

5. EMPIRICAL RESULTS FOR U.S. DATA

5.1. *Models*

In this section, the tools developed in the previous section are applied to an empirical investigation of several consumption-based asset pricing models. Three sets of preferences are considered. The first is CRRA, with $h(Y_t, \theta)$ given in (4.1). The second allows for time nonseparability in the form of durability and habit formation. In this model the representative agent maximizes $E_t \sum_{s=0}^{\infty} \delta^s [(C_s + \rho C_{s-1})^{1-\gamma} - 1]/(1-\gamma)$, and

$$(5.1) \qquad h(Y_t, \theta) = \delta R_{t+1} \left[(C_{t+1} + \rho C_t)^{-\gamma} + \delta\rho(C_{t+2} + \rho C_{t+1})^{-\gamma} \right] \Big/$$

$$(C_t + \rho C_{t-1})^{-\gamma}$$

$$- \left[(C_t + \rho C_{t-1})^{-\gamma} + \delta\rho(C_{t+1} + \rho C_t)^{-\gamma} \right] / (C_t + \rho C_{t-1})^{-\gamma}$$

$$\text{(habit formation/durability)}.$$

---

[8] Let $m_1^0(\theta)$, $m_2^0(\delta)$, and $R^0(\delta)$ denote these quantities computed for some $T = T_0$ so that $E\phi_t = T_0^{-1/2} m_1^0(\theta) + m_2^0(\delta)$. Because the functions $E\phi_t$ and $\Omega(\theta, \theta)$ determine the weak-identification asymptotic approximation, holding the design fixed and changing T amounts to holding these functions fixed and changing T. Thus for general T,

$$ET^{-1/2} \sum_{t=1}^{T} \phi_t = (T/T_0)^{1/2} m_1^0(\theta) + T_0^{1/2} [(T/T_0)^{1/2} m_2^0(\delta)],$$

so the weak-identification asymptotic approximation for general T is obtained by making the transformation $m_1(\theta) = (T/T_0)^{1/2} m_1^0(\theta)$, $m_2(\delta) = (T/T_0)^{1/2} m_2^0(\delta)$, and $R(\delta) = (T/T_0)^{1/2} R^0(\delta)$. Note that $\Omega(\theta, \theta)$ does not depend on T_0 so no adjustment is needed for general T. Because the mean and covariance functions need to be computed only once (for T_0), the computational burden of recomputing the asymptotic distribution does not depend on T.

Positive values of ρ imply durability of consumption, negative values of ρ imply habit formation, and $\rho = 0$ corresponds to CRRA preferences. For additional discussion see Dunn and Singleton (1986), Eichenbaum and Hansen (1990), Ferson and Constantinides (1991), and Hansen, Heaton, and Yaron (1996).

TABLE III

WEAK IDENTIFICATION ASYMPTOTIC APPROXIMATIONS, VARIOUS SAMPLE SIZES

T	γ				δ				J	LR
	10%	Median	90%	$KS(\gamma)$	10%	Median	90%	$KS(\delta)$		
A. Model M1a										
Two Step										
10	−1.482	2.267	6.527	0.26	0.891	0.984	1.064	0.20	2.8%	3.6%
100	−1.139	1.750	4.538	0.12	0.918	0.978	1.024	0.10	3.3%	4.5%
1000	0.338	1.354	2.265	0.03	0.951	0.971	0.988	0.03	5.5%	7.7%
10,000	0.998	1.310	1.604	0.02	0.964	0.970	0.976	0.01	8.0%	9.1%
Continuous Updating										
10	−12.983	1.262	6.961	0.14	0.513	0.965	1.065	0.16	1.6%	15.2%
100	−5.718	1.325	4.912	0.11	0.791	0.968	1.026	0.13	3.3%	14.0%
1000	0.220	1.315	2.259	0.02	0.948	0.970	0.988	0.03	8.3%	9.9%
10,000	0.992	1.303	1.600	0.01	0.964	0.970	0.976	0.01	9.1%	9.5%
B. Model M1b										
Two Step										
10	1.137	6.014	11.931	0.43	0.914	1.029	1.152	0.38	12.8%	57.9%
100	5.996	9.968	16.377	0.30	1.030	1.095	1.164	0.31	23.4%	36.8%
1000	11.092	13.258	16.826	0.08	1.110	1.134	1.160	0.11	11.9%	14.7%
10,000	12.806	13.659	14.672	0.03	1.130	1.138	1.147	0.04	8.9%	10.6%
Continuous Updating										
10	5.440	13.494	71.562	0.30	−2.017	1.058	1.371	0.20	5.4%	12.7%
100	8.510	13.702	51.858	0.17	0.305	1.102	1.187	0.24	6.8%	11.3%
1000	11.385	13.712	18.109	0.08	1.111	1.135	1.162	0.07	8.8%	10.1%
10,000	12.843	13.705	14.734	0.04	1.131	1.139	1.147	0.02	9.2%	9.9%
C. Model M2										
Two Step										
10	−1.041	0.740	5.990	0.11	0.897	0.951	1.024	0.23	12.0%	27.5%
100	−0.481	0.937	3.899	0.20	0.932	0.961	1.003	0.26	16.1%	29.0%
1000	0.942	1.273	1.833	0.11	0.963	0.969	0.978	0.12	29.0%	35.1%
10,000	1.194	1.300	1.415	0.04	0.968	0.970	0.972	0.06	15.5%	18.6%
Continuous Updating										
10	0.426	1.316	17.665	0.29	0.949	0.968	1.095	0.21	8.6%	15.2%
100	0.687	1.286	4.315	0.15	0.959	0.970	1.015	0.14	10.5%	12.5%
1000	1.026	1.299	1.770	0.09	0.965	0.970	0.978	0.08	11.0%	10.7%
10,000	1.198	1.300	1.412	0.04	0.968	0.970	0.972	0.03	10.3%	10.7%

TABLE III—(*Continued*)

T	γ				δ				J	LR
	10%	Median	90%	KS(γ)	10%	Median	90%	KS(δ)		
					D. Model M3					
Two Step										
10	−3.975	1.316	11.338	0.21	0.832	0.950	1.098	0.19	5.3%	10.8%
100	−1.581	1.361	5.364	0.24	0.912	0.967	1.023	0.19	5.9%	21.2%
1000	0.985	1.364	2.021	0.17	0.964	0.970	0.980	0.13	10.0%	36.1%
10,000	1.198	1.307	1.437	0.08	0.968	0.970	0.972	0.06	9.6%	18.6%
Continuous Updating										
10	0.144	1.275	20.430	0.26	0.946	0.966	1.129	0.21	9.3%	12.8%
100	0.586	1.296	4.595	0.16	0.957	0.970	1.018	0.15	10.1%	10.9%
1000	1.004	1.294	1.766	0.07	0.965	0.970	0.978	0.07	11.0%	10.9%
10,000	1.193	1.298	1.420	0.04	0.968	0.970	0.972	0.04	10.0%	10.7%

Note: Kolmogorov-Smirnov statistics compare the weak identification and normal asymptotic approximations. See the notes to Table II.

The third set of preferences considered are the time-separable Kreps-Porteus (1978) preferences as developed by Epstein and Zin (1989, 1991). With CRRA preferences, the coefficient of relative risk aversion is the reciprocal of the intertemporal elasticity of substitution. With Epstein-Zin preferences, this link is broken, and

$$(5.2) \qquad h(Y_t, \theta) = \delta^\lambda (C_{t+1}/C_t)^{-\lambda\gamma} R_{o,t+1}^{\lambda-1} R_{1,t+1} - \iota_G \quad \text{(Epstein-Zin)}$$

where γ now denotes the reciprocal of the intertemporal elasticity of substitution, $R_{o,t}$ denotes the return on the optimal portfolio, and $R_{1,t+1}$ denotes a G-vector of returns on arbitrary asset portfolios. When $\lambda = 1$ this reduces to the CRRA case.

5.2. *The Data*

Two data sets are used. The first is an updated version of the long annual data set used by Campbell and Shiller (1987), and consists of annual U.S. data on stock returns, bond returns, and consumption covering the period 1871 to 1993. The stock returns are based on the Cowles Commission index, followed by the annual average price of the Standard & Poors monthly composite index. The interest rate is the nominal rate for prime 4−6 month commercial paper. The spread, used as an instrument in some models, is the difference between the yield on long term U.S. Treasury bonds and the commercial paper rate. Asset returns were put on a real basis using the producer price index. The consumption series is the real consumption of nondurables and services per capita. Sources and construction of the data are detailed in Campbell and Shiller (1987) (asset data) and Shiller (1982) (consumption data).

The second data set consists of Campbell's (1996) monthly U.S. data from 1959:1 to 1990:12 on twelve assets: returns on 11 portfolios of stocks sorted by sector (petroleum, finance/real estate, consumer durables, basic industry, food/tobacco, construction, capital goods, transportation, textiles and services, and utilities) and the one-month U.S. Treasury bill rate. For details and data sources, see Campbell (1996). We augmented these data by real per capita consumption, constructed as nominal personal consumption expenditures on nondurables (taken from CITIBASE, mnemonic *GMCN*), divided by the adult population (*POP*) and deflated by its implicit deflator (*GMDCN*). This deflator was also used to convert all the monthly nominal asset returns into real returns. Following Epstein and Zin (1991), the optimal portfolio in (5.2) was proxied by the market portfolio as measured by the NYSE value-weighted index of stock returns, and the *GMM* statistics were computed with $G = 13$ (eleven sectors, the NYSE value-weighted return, and the Treasury bill rate).

5.3. *Results*

Conventional two-step GMM results are reported in Tables IV (CRRA preferences), V (habit formation/durability), and VI (Epstein-Zin). Most instrument sets include only the first lag. Because of concerns about temporal aggregation bias, however, in some cases only second lags were used as instruments. The details are given in the tables.

These results, when analyzed using conventional normal asymptotics, generally accord with the existing literature. When stocks and bonds are both priced and the full set of instruments is used (first lags; CRRA-4 and CRRA-10), the overidentifying restrictions implied by the CRRA model are rejected at the 5% level by the J statistic in both data sets. For the habit formation/durability model with the annual data, a moderate positive value of ρ is estimated, indicating durability, but all estimates of γ are, nonsensically, negative, and the J statistic rejects three of the four specifications at the 10% level. In the monthly data, conflicting results are obtained for the two models not rejected at the 10% level by the J statistic. With first lags as instruments (H/D-5), ρ is precisely estimated as positive, but with second lags as instruments (H/D-7), ρ is imprecisely estimated as negative. For Epstein-Zin preferences, when lagged returns on the market and consumption growth are used as instruments (EZ-1 and EZ-2), the J statistic fails to reject and the hypothesis $\lambda = 1$ (CRRA preferences) is rejected at the 1% level, but the estimates of γ are negative. When the spread and the dividend yield are added as instruments, the J statistic rejects at the 5% level (EZ-3 and EZ-4).

Because of the possibility of weak identification, we computed S-sets for these models, both for all parameters jointly (based on Theorem 2) and for the weakly identified parameters (based on Theorem 3). In the CRRA model, δ is treated as strongly identified as discussed in the previous section. In the habit formation/durability and Epstein-Zin models, it remains reasonable to treat the function of the parameters describing the unconditional mean as well identified

TABLE IV

GMM ESTIMATION RESULTS, CRRA PREFERENCES

Model	Assets	GK	Instruments	δ	γ	J	90% S-set for γ	S-set
			A. Annual Data, 1871−1993					
CRRA-1	SR	3	SR, CG	0.903	0.132	0.69	$(-2.0, 2.3)$	Fig. 2
				(0.022)	(1.037)	[0.69]		
CRRA-2	SR, BR	6	SR, CG	0.958	−0.507	22.93***	∅	∅
				(0.007)	(0.298)	[25.27]***		
CRRA-3	SR, BR	8	SR, BR, CG	0.952	−0.819	32.73***	∅	∅
				(0.007)	(0.306)	[39.10]***		
CRRA-4	SR, BR	10	SR, BR, CG,	0.953	−0.745	37.81***	∅	∅
			Spread, DY	(0.007)	(0.277)	[46.82]***		
CRRA-5	SR, BR	8	As CRRA-3,	0.971	6.436	10.11	$(18.8, 40.8)$	Fig. 3
			second lag	(0.037)	(2.994)	[33.00]***		
CRRA-6	SR, BR	10	As CRRA-4,	0.822	−3.333	12.02	$(16.0, 58.8)$	†
			second lag	(0.042)	(2.269)	[70.81]***		
			B. Monthly Data, 1959:1−1990:12					
CRRA-7	SR	3	SR, CG	0.999	0.641	2.79	$(-0.1, 1.5)$	†
				(0.002)	(0.743)	[2.79]		
CRRA-8	SR, BR	6	SR, CG	0.999	0.122	3.07	∅	∅
				(0.000)	(0.068)	[20.36]***		
CRRA-9	SR, BR	8	SR, BR, CG	0.999	0.210	4.38	∅	∅
				(0.000)	(0.071)	[81.78]***		
CRRA-10	SR, BR	10	SR, BR, CG,	0.998	0.035	20.43***	∅	∅
			Spread, DY	(0.000)	(0.064)	[113.58]***		
CRRA-11	SR, BR	8	As CRRA-9,	0.999	1.148	1.53	$(0.6, \infty)$	†
			second lag	(0.001)	(0.396)	[3.23]		
CRRA-12	SR, BR	10	As CRRA-10,	0.999	0.289	13.88*	∅	∅
			second lag	(0.000)	(0.160)	[54.59]***		

Notes: J statistics are significant at *10%, **5%, ***1% significance levels, based on the standard chi-squared critical value. ∅ denotes an empty *S*-set, and † denotes a nonempty *S*-set, which is not presented graphically to save space; these figures are available upon request from the authors. Point estimates, standard errors (in parentheses), the first *J* statistic in each row are two-step estimates. The *J* statistics in square brackets were computed from the continuous updating objective function. The *S*-set for γ is based on the objective function, concentrated with respect to δ. The instruments are the indicated variables, lagged once, except for the models in which the instruments are stated as lagged twice. The instruments always include a constant term. Variable definitions: *SR* = stock returns (returns on a market portfolio); *BR* = bond returns; *CG* = consumption growth, spread = long bond rate minus short interest rate, *DY* = dividend yield. See the text for a discussion of the data.

(the constant term is a strong instrument). This function depends on δ and ρ (habit formation-durability) or δ and λ (Epstein-Zin); as in the CRRA case, γ enters this function only weakly when movements in consumption are small, as they are in the data. In those models, we therefore report two concentrated *S*-sets, one in which δ is concentrated out and one in which either ρ or λ is concentrated out.

The results for the CRRA models are summarized in the final columns of Table IV, and selected 90% *S*-sets (along with conventional two-step GMM confidence ellipses) are graphed in Figures 2 and 3. When only stocks are priced

TABLE V

GMM Estimation Results, Habit Formation / Durability Preferences

Model	Assets	GK	Instruments	δ	γ	ρ	J	S-set
A. Annual Data, 1871−1993								
H/D-1	SR, BR	8	SR, BR, CG	0.944 (0.009)	−0.897 (0.438)	0.116 (0.187)	24.77*** [42.86]***	Fig. 4
H/D-2	SR, BR	10	SR, BR, CG, Spread, DY	0.944 (0.008)	−0.895 (0.421)	0.112 (0.187)	28.08*** [50.21]***	†
H/D-3	SR, BR	8	As H/D-1, second lag	0.866 (0.037)	−2.154 (2.152)	0.082 (0.318)	11.17* [76.09]***	†
H/D-4	SR, BR	10	As H/D-2, second lag	0.790 (0.067)	−6.480 (3.864)	0.286 (0.258)	13.09 [59.86]***	†
B. Monthly Data, 1959:1−1990:12								
H/D-5	SR, BR	8	SR, BR, CG	1.004 (0.002)	5.182 (1.604)	0.443 (0.074)	3.87 [6.83]	Fig. 5
H/D-6	SR, BR	10	SR, BR, CG, Spread & DY	0.999 (0.000)	1.111 (0.299)	2.111 (0.482)	19.89*** [64.17]***	∅
H/D-7	SR, BR	8	As H/D-5, second lag	0.999 (0.002)	1.174 (1.709)	−0.336 (0.359)	1.43 [3.98]	†
H/D-8	SR, BR	10	As H/D-6, second lag	0.992 (0.004)	−1.885 (2.061)	−0.332 (0.275)	12.99* [25.83]***	†

Notes: J statistics are significant at *10%, **5%, ***1% significance levels, based on the standard chi-squared critical value. ∅ denotes an empty *S*-set, and † denotes a nonempty *S*-set, which is not presented graphically to save space. See the notes to Table IV.

TABLE VI

GMM Estimation Results, Epstein-Zin Preferences
Monthly Data, 1959:1−1990:12

Model	Assets	GK	Instruments	δ	γ	λ	J	S-set
EZ-1	BR, MR, returns on 11 sector portfolios	39	MR, CG	0.999 (0.000)	−0.025 (0.065)	0.710 (0.079)	45.87 [52.92]**	∅
EZ-2	BR, BR returns on 11 sector portfolios	39	As EZ-1, second lag	0.995 (0.005)	−4.082 (4.597)	0.055 (0.041)	29.42 [33.68]	†
EZ-3	BR, MR, returns on 11 sector portfolios	65	MR, CG, Spread & DY	0.999 (0.000)	−0.031 (0.062)	0.738 (0.031)	93.44*** [114.37]***	∅
EZ-4	BR, MR, returns on 11 sector portfolios	65	As EZ-2, second lag	0.997 (0.001)	−1.892 (0.906)	0.738 (0.038)	82.70** [81.09]*	Fig. 6

Notes: J statistics are significant at *10%, **5%, ***1% significance levels, based on the standard chi-squared critical value. ∅ denotes an empty *S*-set, and † denotes a nonempty *S*-set, which is not presented graphically to save space. MR is the return on the market portfolio (the proxy for the optimal portfolio). In each model there are 13 Euler equations ($G = 13$). See the notes to Table IV.
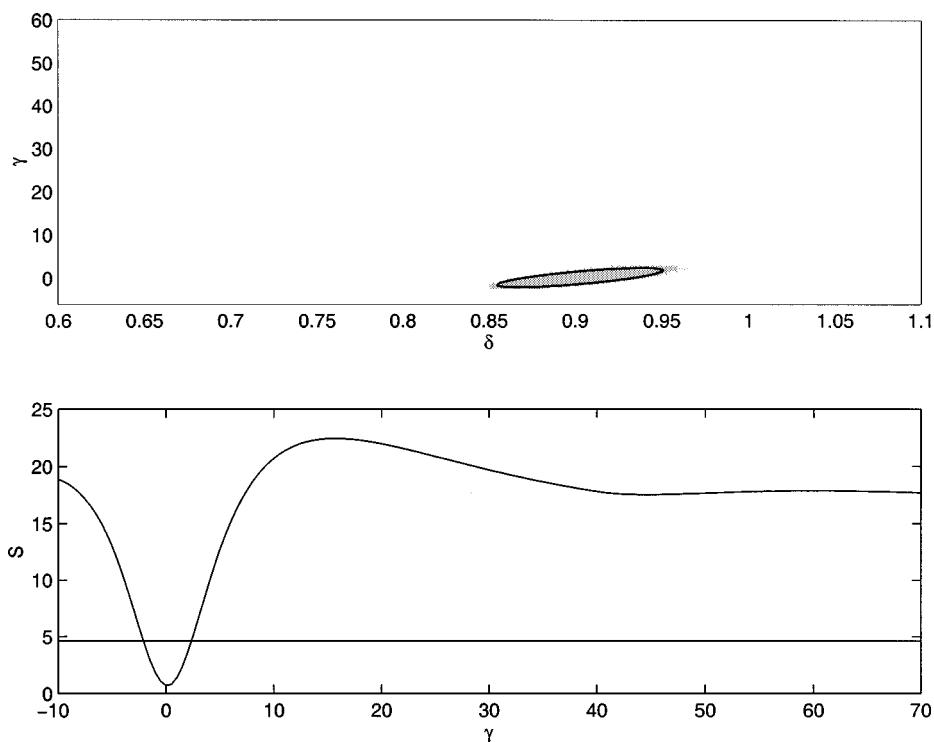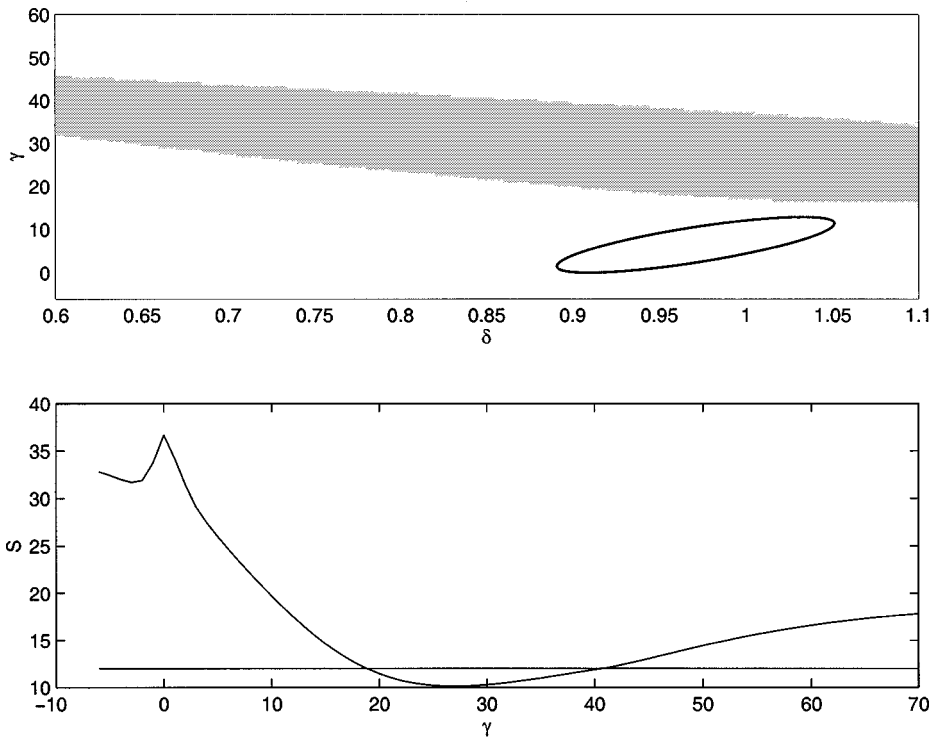
FIGURE 2.—Joint S-set and concentrated objective function: model CRRA-1. (a) Joint 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ) (upper panel); (b) objective function concentrated with respect to δ (lower panel).

using annual data (CRRA-1), inferences based on the S-sets and conventional GMM agree closely. Although the objective function in which δ is concentrated out has multiple minima (Figure 2b), the second minimum is well above the χ^2 critical value used to construct the concentrated S-set. Similarly, for models CRRA-2, CRRA-3, and CRRA-4, the S-sets are null, indicating that there are no parameter values consistent with the overidentifying conditions, the same conclusion as is reached using the conventional J statistic. However, when the second lags are used as instruments (Figure 3 for CRRA-5; CRRA-6 is similar), the conventional confidence ellipse and the S-sets have no points in common; the S-sets are much larger, and the degree of risk aversion is greater. Among the CRRA models with monthly data, the S-set and standard GMM inferences agree most closely when only stock returns and consumption growth are used as instruments (CRRA-7). When both stocks and bonds are priced (CRRA-8 and CRRA-9), the two-step J-statistic fails to reject but the S-sets are null. When second lags are used as instruments, the S-set differs sharply from the seemingly precise GMM confidence ellipse.

FIGURE 3.—Joint S-set and concentrated objective function: model CRRA-5. (a) Joint 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ) (upper panel); (b) objective function concentrated with respect to δ (lower panel).

The differences between conventional GMM inferences and those based on the S-sets are dramatic for the habit formation/durability models. Although the J statistic suggests that models H/D-1, H/D-2, and H/D-3 are rejected at the 10% level, the 90% S-sets are nonempty; in fact they contain a large set of parameters and are disjoint with the standard GMM confidence sets (see Figure 4 for H/D-1; H/D-2 is similar). Although H/D-4 is not rejected at the 10% level using the J-statistic, the S-sets and standard GMM ellipses for H/D-4 differ sharply. With only second lags as instruments the S-sets are somewhat larger. Comparing the S-sets for H/D-1 and H/D-5 (Figures 4 and 5), which use comparable sets of instruments, reveals that the annual data are consistent with habit formation while the monthly data are consistent with durability. This accords with the theoretical results in Heaton (1993) (although his functional form differs), but this is not revealed by the two-step GMM point estimates, which suggest durability in both the monthly and annual data.

Conventional and S-set inferences also disagree for the Epstein-Zin preferences. The only non-null S-sets obtain using second lags as instruments. For the long instrument list in EZ-4 (Figure 6), the S-sets are consistent with moderate
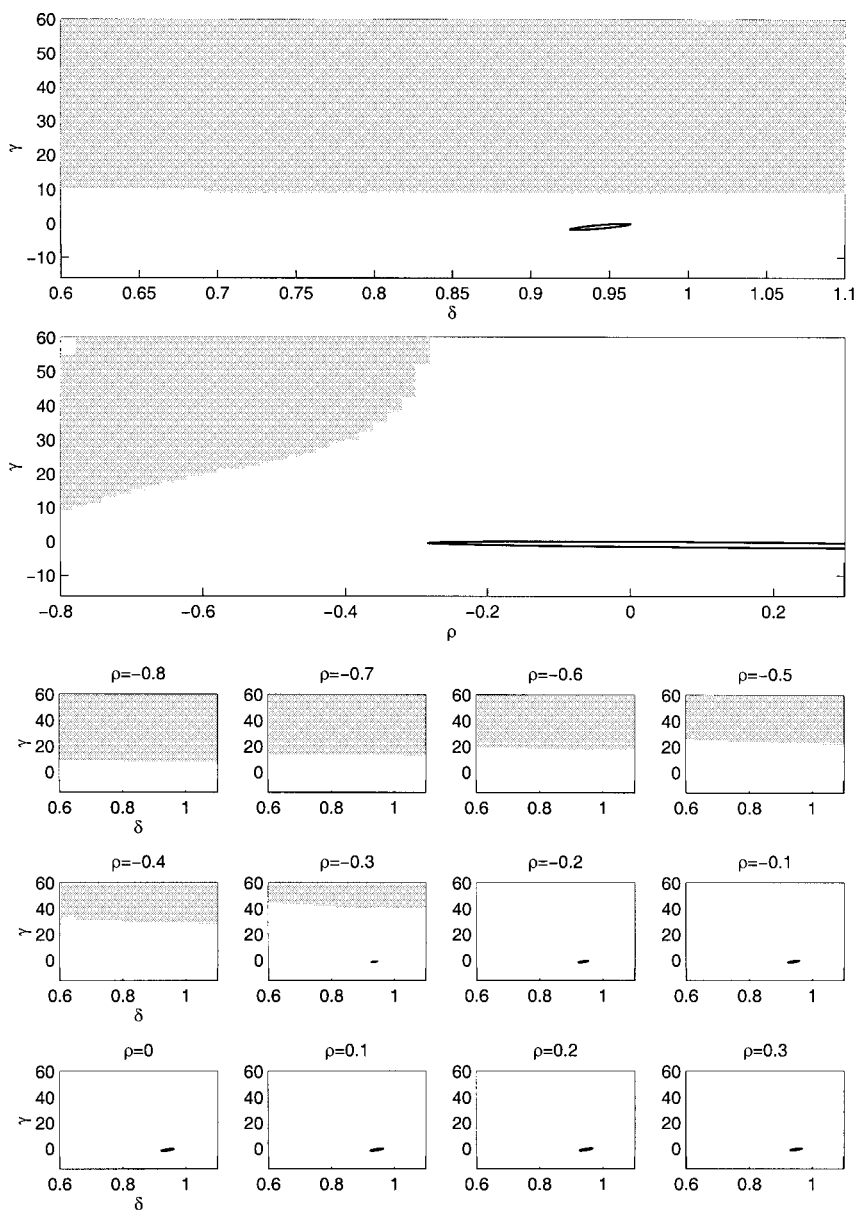
FIGURE 4.—Concentrated S-sets: model H/D-1. (a) 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ), ρ concentrated out (upper); (b) 90% S-Set (shaded) and 90% GMM confidence ellipse for (γ, ρ), δ concentrated out (lower); (c) joint S-set: model H/D-1. Joint 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ, ρ), sliced in ρ dimension.
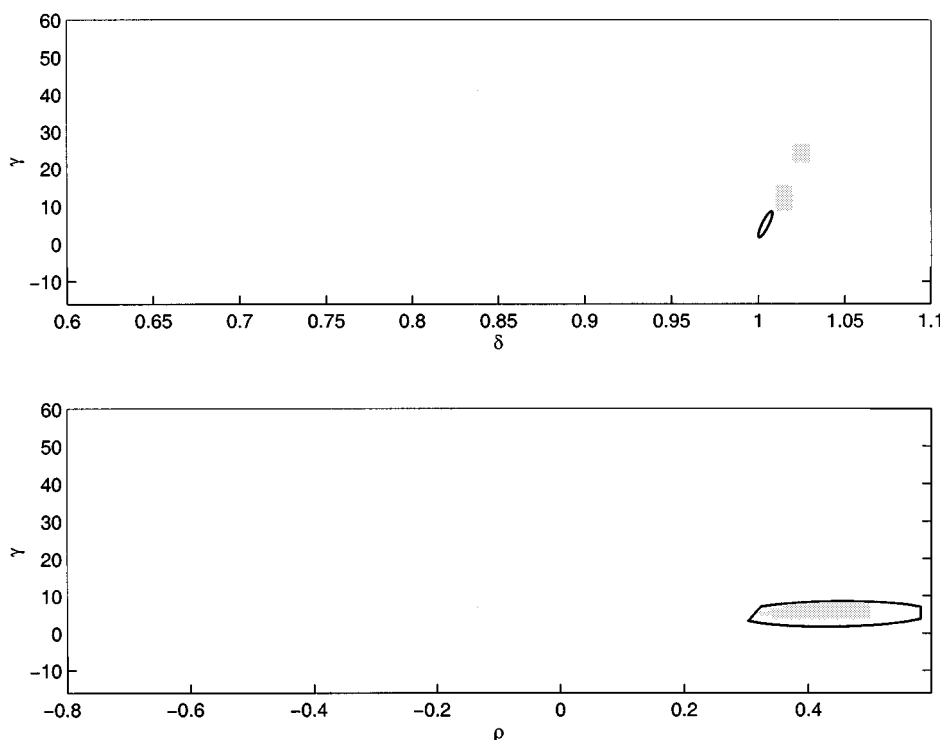
FIGURE 5(a, b).—Concentrated S-sets: model H/D-5. (a) 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, δ), ρ concentrated out (upper); (b) 90% S-set (shaded) and 90% GMM confidence ellipse for (γ, ρ), δ concentrated out (lower).

and high levels of risk aversion (the coefficient of relative risk aversion in the EZ model is $1 - \lambda + \lambda\gamma$). The S-sets for $\lambda = 1$ in models EZ-2 and EZ-4 are nonempty, indicating that these data are in fact consistent with the CRRA model, the opposite conclusion as reached using the standard GMM Wald statistic.

These results reveal several symptoms of weak identification. Conclusions based on the J statistic evaluated using the two step and continuous updating objective functions often differ (e.g. CRRA-5, CRRA-6). The two-step point estimates are sensitive to instrument choice even in models for which the two-step J statistic does not reject (e.g. CRRA-5, CRRA-6, or EZ-1 and EZ-2). The continuous updating estimates (not tabulated here to save space) often differ substantially from the two-step estimates, and in some cases tended towards arbitrarily large values. Importantly, the S-sets and conventional GMM confidence sets typically disagree, even when the J statistic does not reject. These observations lead us to conclude that the inferences based on conventional GMM methodology are unreliable.
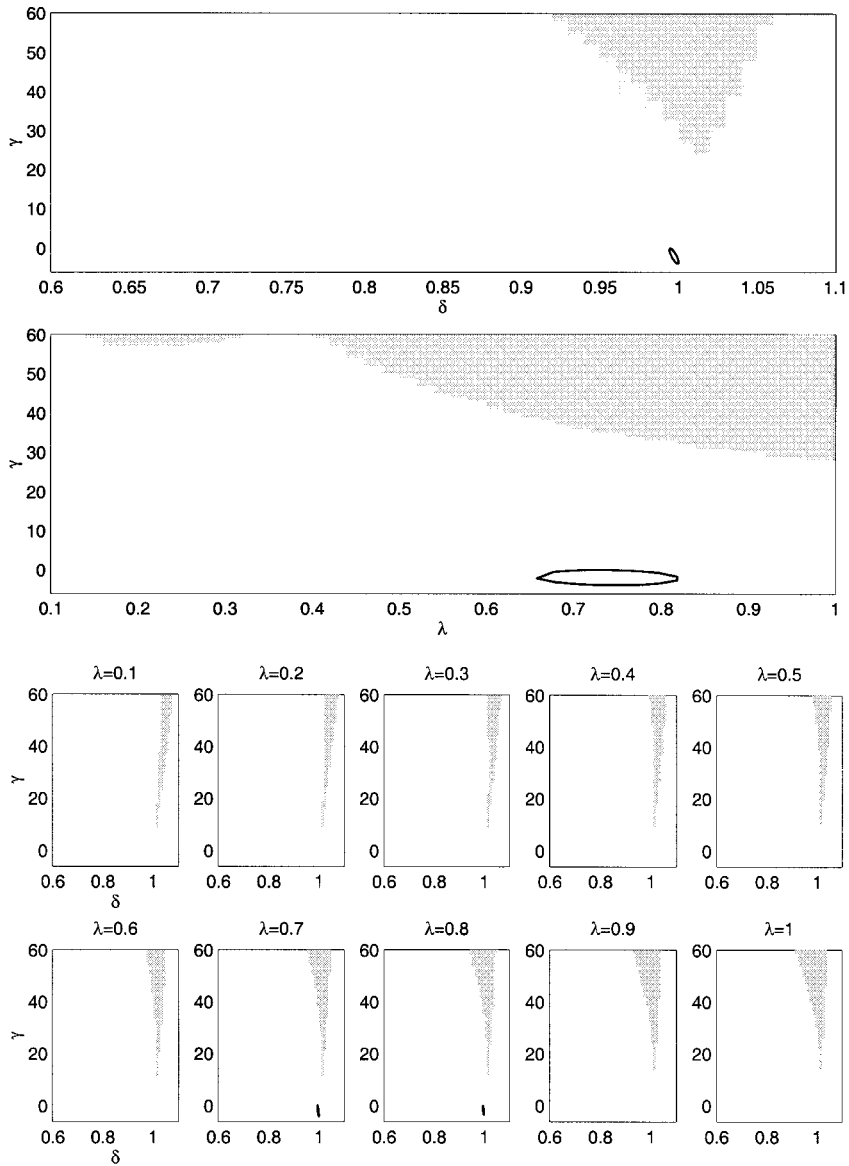
FIGURE 6.—Concentrated *S*-sets: model EZ-4. (a) 90% *S*-set (shaded) and 90% GMM confidence ellipse for (γ, δ), λ concentrated out (upper); (b) 90% *S*-set (shaded) and 90% GMM confidence ellipse for (γ, λ), δ concentrated out (lower); (c) joint *S*-set: model EZ-4. Joint 90% *S*-set (shaded) and 90% GMM confidence ellipse for $(\gamma, \delta, \lambda)$, sliced in λ dimension.

The S-sets lead to different substantive conclusions than the conventional GMM analysis. The S-sets generally indicate greater degrees of risk aversion than found using conventional GMM. Although such high risk aversion might seem counterintuitive, these results are consistent with the view, recapitulated in Campbell, Lo, and MacKinlay (1997, Ch. 8.2), that very high risk aversion is needed to resolve the equity premium puzzle. Among the specifications in which both stocks and bonds are priced and first lags are used as instruments, the only models for which the S-sets are nonempty are habit formation/durability models. In the annual data, these sets suggest habit formation, but in the monthly data, they suggest durability. For Epstein-Zin preferences, the only nonrejected specifications have second lags as instruments, and the associated S-sets provide little evidence against the CRRA specification in favor of Epstein-Zin preferences.

6. DISCUSSION AND CONCLUSIONS

From a methodological perspective, it is noteworthy that the empirical conclusions based on the S-sets and conventional GMM analysis differ. The puzzling Epstein-Zin conventional GMM point estimates in the nonrejected models are less puzzling when viewed in the context of the S-sets. Using conventional GMM, the habit formation/durability model is largely rejected, or the point estimates are nonsensical from an economic perspective, but using S-sets these models are often not rejected and the confidence sets are consistent with risk aversion. Generally speaking, the S-sets point to higher degrees of risk aversion than suggested by the conventional GMM analysis. These differences underscore the importance of using procedures that are robust to the problem of weak identification in Euler equation estimation.

The weak-identification asymptotic theory developed here might be extended in several ways. Although $\phi_t(\theta_0)$ is assumed to be a martingale difference sequence, if instead $\phi_t(\theta_0)$ is integrated of order zero and autocorrelated, then the efficient estimator would use a heteroskedasticity and autocorrelation consistent covariance matrix. The extension to the autocorrelated case is conceptually straightforward as long as Assumptions A–D are satisfied. Another extension is to develop approximations to the distributions of statistics testing q linear restrictions on θ when the instruments are weak. This is relevant for understanding distortions of sizes of tests and coverage rates of conventional confidence intervals. An explicit asymptotic representation of the likelihood ratio statistic for $q = K$ has been provided, and the specialization to $q < K$ is conceptually straightforward. The extension to Wald statistics and conventional standard errors appears to be more difficult. Although explicit limiting representations for Wald statistics can be obtained in some special cases (e.g. when $\phi_t(\theta)$ is a finite order polynomial in θ), in the general GMM problem with arbitrary nonlinearities it appears that additional assumptions are needed. This extension is left for future work.

Although some tools for inference robust to weak identification have been developed in this paper, here too work remains to be done. For example, as discussed in Section 2.4, the interpretation of the S-sets can be complicated because they jointly test $\theta = \theta_0$ and instrument validity. It would be useful to have a method for constructing confidence sets that are robust to weak identification, conditional on model (instrument) validity. Staiger and Stock (1997) and Wang and Zivot (1998) proposed such methods in the linear case, but these do not extend naturally to nonlinear models. It also would be useful for a researcher to have a statistical measure of whether she faces weak identification in a particular application. In general this depends on the global properties of $\mu(\alpha)'\mu(\alpha)$, an unobserved function. Outside of the linear case, where $\mu(\alpha)'\mu(\alpha)$ is quadratic, there are no extant methods for reliable inference about this function directly. The development of a simple and reliable statistic to detect weak identification remains an open challenge.

This analysis nevertheless points to several symptoms of weak identification that can be readily detected in empirical work. One such symptom is that the objective function is clearly nonquadratic and has plateaus or ridges that are not far (in terms of LR statistics) from its minimum value, as seen in Figures 3 and 4. A second, related symptom is that S-sets and conventional GMM confidence sets have substantial areas of disagreement. A third symptom is obtaining substantially different point estimates and inferences using GMM estimators that, under the conventional theory, are asymptotically equivalent. A fourth symptom is when a Monte Carlo study of a model calibrated to the empirical problem at hand yields economically significant biases in GMM point estimates and size distortions in LR and J statistics. If such symptoms are present, a diagnosis of weak identification is appropriate, and it is prudent to report S-sets in addition to, or instead of, conventional GMM statistics.

Kennedy School of Government, 79 J.F.K. St., Harvard University, Cambridge, MA 02138, U.S.A.

and

Board of Governors of the Federal Reserve System, Washington, D.C. 20551, U.S.A.

APPENDIX A

PROOFS OF THEOREMS

Before proving the theorems, it is shown that $\hat{\beta}$ is $\sqrt{T}$-consistent for β_0.

LEMMA A1: *Under the assumptions of Theorem 1, $T^{1/2}(\hat{\beta} - \beta_0) = O_p(1)$.*

PROOF: We first show that $\hat{\beta} \overset{p}{\to} \beta_0$. Let $m_T(\theta) = ET^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)$, so

$$S_T(\theta; \bar{\theta}_T(\theta)) = [\Psi_T(\theta) + m_T(\theta)]'W_T(\bar{\theta}_T(\theta))[\Psi_T(\theta) + m_T(\theta)].$$

By the various assumptions,

$$T^{-1}S_T(\theta; \bar{\theta}_T(\theta)) \xrightarrow{p} m_2(\beta)' W(\bar{\theta}(\theta)) m_2(\beta)$$

uniformly in θ. Because W is positive definite by Assumption D and $m_2(\beta) = 0$ iff $\beta = \beta_0$, by the continuity of the argmin operator, $\hat{\beta} \xrightarrow{p} \beta_0$.

To show $\sqrt{T}$-consistency, use the fact that $\hat{\theta}$ minimizes S_T and Assumption C to write,

$$S_T(\theta; \bar{\theta}_T(\hat{\theta})) - S_T(\theta_0; \bar{\theta}_T(\theta_0)) = [\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta}) + \sqrt{T} m_2(\hat{\beta})]'$$
$$\times W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta}) + \sqrt{T} m_2(\hat{\beta})]$$
$$- \Psi_T(\theta_0)' W_T(\bar{\theta}_T(\theta_0)) \Psi_T(\theta_0) \le 0$$

or, equivalently,

$$(A.1) \qquad T m_2(\hat{\beta})' W_T(\bar{\theta}_T(\hat{\theta})) m_2(\hat{\beta}) + 2\sqrt{T} m_2(\hat{\beta})' W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})] + d_{1T}(\hat{\theta}) \le 0,$$

where

$$d_{1T}(\theta) = [\Psi_T(\theta) + m_{1T}(\theta)]' W_T(\bar{\theta}_T(\theta))[\Psi_T(\theta) + m_{1T}(\theta)]$$
$$- \Psi_T(\theta_0)' W_T(\bar{\theta}_T(\theta_0)) \Psi_T(\theta_0).$$

Without loss of generality, let W_T be symmetric (if not, replace it by $1/2(W_T + W_T')$). Now $T m_2(\hat{\beta})' W_T(\bar{\theta}_T(\hat{\theta})) m_2(\hat{\beta}) \ge \|\sqrt{T} m_2(\hat{\beta})\|^2$ mineval$[W_T(\bar{\theta}_T(\hat{\theta}))]$, where mineval$(A)$ denotes the minimum eigenvalue of the matrix A. Also,

$$\sqrt{T} m_2(\hat{\beta})' W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})]$$
$$\ge -\|\sqrt{T} m_2(\hat{\beta})\| \|W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})]\|.$$

Using these inequalities and dividing (A.1) through by mineval$[W_T(\bar{\theta}_T(\hat{\theta}))]$ (which is positive with probability one by Assumption D),

$$(A.2) \qquad \|\sqrt{T} m_2(\hat{\beta})\|^2 - 2 d_{2T} \|\sqrt{T} m_2(\hat{\beta})\| + d_{3T} \le 0,$$

where

$$d_{2T} = \|W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\theta)]\| / \text{mineval}[W_T(\bar{\theta}_T(\hat{\theta}))] \quad \text{and}$$
$$d_{3T} = d_{1T}(\hat{\theta}) / \text{mineval}[W_T(\bar{\theta}_T(\hat{\theta}))].$$

Now take the roots of (A.2) and write $\sqrt{T} m_2(\hat{\beta}) = R(\tilde{\beta}) \sqrt{T}(\hat{\beta} - \beta_0)$, where $\tilde{\beta} \in (\beta_0, \hat{\beta})$ and $R(\beta) = \partial m_2(\beta)/\partial \beta'$. Thus for (A.2) to hold it must be that

$$(A.3) \qquad \|R(\tilde{\beta}) \sqrt{T}(\hat{\beta} - \beta_0)\| \le d_{2T} + (d_{2T}^2 - d_{3T})^{1/2}.$$

Because $\tilde{\beta} \xrightarrow{p} \beta_0$, $R(\tilde{\beta}) \xrightarrow{p} R(\beta_0)$ which has full column rank by Assumption C. The desired result, that $\sqrt{T}(\hat{\beta} - \beta_0) = O_p(1)$, follows if (i) $d_{2T} = O_p(1)$ and (ii) $d_{3T} = O_p(1)$. These are now shown.

(i) $\qquad d_{2T} = \|W_T(\bar{\theta}_T(\hat{\theta}))[\Psi_T(\hat{\theta}) + m_{1T}(\hat{\theta})]\| / \text{mineval}[W_T(\bar{\theta}_T(\hat{\theta}))]$

$$\le \sup_\theta \|W_T(\bar{\theta}_T(\theta))[\Psi_T(\theta) + m_{1T}(\theta)]\| / \inf_\theta \text{mineval}[W_T(\bar{\theta}_T(\theta))]$$

$$(A.4) \qquad \Rightarrow \sup_\theta \|W(\bar{\theta}(\theta))[\Psi(\theta) + m_1(\theta)]\| / \inf_\theta \text{mineval}[W(\bar{\theta}(\theta))]$$

by Assumptions B, C, and D. By these assumptions, the numerator in (A.4) is $O_p(1)$ and the denominator is a positive constant, so $d_{2T} = O_p(1)$.

(ii) $|d_{3T}| = |d_{1T}(\hat{\theta})|/\text{mineval}[W_T(\bar{\theta}_T(\hat{\theta}))]$

$\leq \sup_\theta |d_{1T}(\theta)|/\inf_\theta \text{mineval}[W_T(\bar{\theta}_T(\theta))]$

$\leq 2 \sup_\theta [\Psi_T(\theta) + m_{1T}(\theta)]' W_T(\bar{\theta}_T(\theta))[\Psi_T(\theta) + m_{1T}(\theta)]/$

$\inf_\theta \text{mineval}[W_T(\bar{\theta}_T(\theta))]$

(A.5) $\Rightarrow 2 \sup_\theta [\Psi(\theta) + m_1(\theta)]' W(\bar{\theta}(\theta))[\Psi(\theta) + m_1(\theta)]/\inf_\theta \text{mineval}[W(\bar{\theta}(\theta))]$

by Assumptions B, C, and D. The right-hand side of (A.5) is $O_p(1)$, so $d_{3T} = O_p(1)$.   Q.E.D.

PROOF OF THEOREM 1:
(i) By Lemma A1, it suffices to obtain a limiting representation for $S_T(\alpha, \beta_0 + b/T^{1/2})$ as an empirical process in $(\alpha', b') \in A \times \bar{B}$, where $\bar{B}$ is compact. Now,

$$T^{-1/2} \sum_{t=1}^{T} \phi_t(\alpha, \beta_0 + b/T^{1/2}) = \Psi_T(\alpha, \beta_0 + b/T^{1/2}) + m_{1T}(\alpha, \beta_0 + b/T^{1/2})$$

$$+ T^{1/2} m_2(\beta_0 + b/T^{1/2}).$$

By Assumption B, $\Psi_T(\alpha, \beta_0 + b/T^{1/2}) \Rightarrow \Psi(\alpha, \beta_0)$; by Assumption C(i), $m_{1T}(\alpha, \beta_0 + b/T^{1/2}) \to m_1(\alpha, \beta_0)$; by Assumption C(ii), $T^{1/2} m_2(\beta_0 + b/T^{1/2}) \to R(\beta_0)b$; and by Assumption D, $W_T(\bar{\theta}_T(\alpha, \beta_0)) \to W(\bar{\theta}(\alpha, \beta_0))$. These limits are all uniform in $(\alpha', b') \in A \times \bar{B}$, proving (i).

(ii) By Lemma 3.2.1 of van der Vaart and Wellner (1996, p. 286), it follows that $(\hat{\alpha}', T^{1/2}(\hat{\beta} - \beta_0)')' \Rightarrow (\alpha^{*'}, b^{*'})' = \text{argmin}_{(\alpha', b') \in A \times \bar{B}} \bar{S}(\alpha, b; \bar{\theta}(\alpha, \beta_0))$. To obtain the concentrated limiting objective function $S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$, fix α, differentiate $\bar{S}(\alpha, \beta; \bar{\theta}(\alpha, \beta_0))$, and rearrange the first order conditions to obtain

$$b^*(\alpha) = -[R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0)) R(\beta_0)]^{-1}$$

$$\times R(\beta_0)' W(\bar{\theta}(\alpha, \beta_0))[\Psi(\alpha, \beta_0) + m_1(\alpha, \beta_0)].$$

Setting $S^*(\alpha; \bar{\theta}(\alpha, \beta_0)) = \bar{S}(\alpha, b^*(\alpha); \bar{\theta}(\alpha, \beta_0))$ and rearranging yields the expression for $S^*(\alpha; \bar{\theta}(\alpha, \beta_0))$ in the theorem.

A consequence of the continuous mapping theorem and the envelope theorem is that $\hat{\alpha} \Rightarrow \alpha^* = \text{argmin}_{\alpha \in A} S^*(\alpha)$. Because $\hat{\beta} = \hat{\beta}(\hat{\alpha})$, $T^{1/2}(\hat{\beta} - \beta_0) \Rightarrow b^*(\alpha^*)$, which yields the expression in the theorem.   Q.E.D.

PROOF OF THEOREM 2: Because $E\phi_t(\theta_0) = 0$, $S_T(\theta_0) = \Psi_T(\theta_0)' W_T(\theta_0)\Psi_T(\theta_0)$ by Assumption A and the assumption $W_T(\theta_0) \overset{p}{\to} W(\theta_0) = \Omega(\theta_0, \theta_0)^{-1}$, $S_T(\theta_0, \theta_0) = \Psi_T(\theta_0)' W_T(\theta_0)\Psi_T(\theta_0) \overset{d}{\to} \Psi(\theta_0)'\Omega(\theta_0, \theta_0)^{-1}\Psi(\theta_0) \sim \chi^2_{GK}$.   Q.E.D.

PROOF OF THEOREM 3: Because $m_1(\alpha_0, \beta_0) = 0$ and by assumption $W(\theta_0) = \Omega(\theta_0, \theta_0)^{-1}$, from Theorem 1 we have

$$S_T(\alpha_0, \hat{\beta}; \alpha_0, \hat{\beta}) \Rightarrow [\Omega(\theta_0, \theta_0)^{-1/2}\Psi(\theta_0)]' \tilde{M}(\theta_0)[\Omega(\theta_0, \theta_0)^{-1/2}\Psi(\theta_0)],$$

where $\tilde{M}(\theta_0) = I - \tilde{R}_0[\tilde{R}_0'\tilde{R}_0]^{-1}\tilde{R}_0'$, where $\tilde{R}_0 = \Omega(\theta_0, \theta_0)^{-1/2}R(\beta_0)$. The result follows from noting that $\Omega(\theta_0, \theta_0)^{-1/2}\Psi(\theta_0)$ is a $GK \times 1$ standard normal random variable and $\tilde{M}(\theta_0)$ is idempotent with rank $GK - n_2$.   Q.E.D.

PROOF OF COROLLARY 4: For each of the estimators, the assumption $\bar{\theta}_T(\theta) \Rightarrow \bar{\theta}(\theta)$ in Theorem 1 must be verified. For the one step estimator, we can set $\bar{\theta}_T(\theta) = \bar{\theta}(\theta) = 0$ and the assumption is satisfied trivially, and parts (a) and (b) follow. For the two-step estimator, $\bar{\theta}_T(\theta) = \hat{\theta}_1$, and the

assumption is implied by part (b); parts (c) and (d) thus follow. For the continuous updating estimator, $\bar{\theta}_T(\theta) = \bar{\theta}(\theta) = \theta$ and the assumption is satisfied, so parts (e) and (f) follow. The remaining results are direct implications of parts (c)–(f). $Q.E.D.$

APPENDIX B

COMPUTATION OF WEAK ASYMPTOTIC APPROXIMATION FOR THE CRRA CCAPM

The weak asymptotic approximations to the GMM estimators and test statistics analyzed in the Monte Carlo analysis of Section 4 were computed in three steps: (i) for each design and set of instruments, compute the functions m_1, m_2, and Ω using a Taylor series approximation to (4.1), where the parameters were computed by simulation of the DGP of Section 4.2; (ii) given these functions, obtain 5000 Monte Carlo draws of the limiting objective functions of interest; and (iii) for each Monte Carlo realization of each objective function, compute the various statistics of interest (the optimizing estimator and the test statistics), which yields 5000 Monte Carlo draws of each statistic. The global Taylor series approximation is feasible because α_0 and δ_0 are known, along with the other parameters describing the DGP, and because the parameter space was bounded. The main numerical advantage of using a finite order Taylor series approximation is that the limiting random function Ψ can be represented as a function of a finite-dimensional normal random variable. The Taylor series approximation was only used to compute the weak-instrument asymptotic distributions.

Without loss of generality, let the first element of Z_t be a constant and let the remaining elements have sample mean zero. To order m, the Taylor series approximation is

$$
\text{(B.1)} \qquad h(Y_t, \theta) = \delta(C_{t+1}/C_t)^{-\gamma} R_{t+1} - \iota_G = \delta R_{t+1} e^{(\gamma_0 - \gamma)c_{t+1}}/e^{\gamma_0 c_{t+1}} - \iota_G
$$

$$
\cong \delta R_{t+1} e^{-\gamma_0 c_{t+1}}\left[1 + \sum_{i=1}^{m} c_{t+1}^i (\gamma_0 - \gamma)^i / i!\right] - \iota_G
$$

$$
= \delta \eta_{t+1} g(\gamma)' C_{t+1}^{(m)} - \iota_G
$$

$$
= \delta[\eta_{t+1} - \delta_0^{-1}\iota_G \, \eta_{t+1} \tilde{C}_{t+1}^{(m)'}]g(\gamma) + (\delta/\delta_0 - 1)\iota_G
$$

where

$$
\eta_{t+1} = R_{t+1} \exp(-\gamma_0 c_{t+1}),
$$

$$
g(\gamma) = [1 \quad (\gamma_0 - \gamma) \quad 1/2(\gamma_0 - \gamma)^2 \quad \cdots \quad (\gamma_0 - \gamma)^m / m!]', \quad \text{and}
$$

$$
C_{t+1}^{(m)} = [1 \quad \tilde{C}_{t+1}^{(m)\gamma}]',
$$

where $\tilde{C}_{t+1}^{(m)} = [c_{t+1} \quad c_{t+1}^2 \quad \cdots \quad c_{t+1}^m]'$. Thus,

$$
\text{(B.2)} \qquad T^{-1/2} \sum_{t=1}^{T} \phi_t(\theta) \cong \delta T^{-1/2} \sum_{t=1}^{T} [\zeta_t g(\gamma)] \otimes Z_t + T^{1/2}(\delta/\delta_0 - 1)\iota_G \otimes e_K
$$

$$
= \delta[(I_G \otimes g') \otimes I_K]T^{-1/2} \sum_{t=1}^{T} \text{vec}(\zeta_t') \otimes Z_t + T^{1/2}(\delta/\delta_0 - 1)\iota_G \otimes e_K
$$

where e_K is the $K \times 1$ vector $(1 \ 0 \ \cdots \ 0)'$ and $\zeta_t = [\eta_{t+1} - \delta_0^{-1}\iota_G \, \eta_{t+1} \tilde{C}_{t+1}^{(m)'}]$ (the first equality in (B.2) uses $\sum_{t=1}^{T} Z_t = Te_K$). It follows that $\Sigma_{hh}(\theta) = \delta^2(I_G \otimes g(\gamma))'\Sigma_{\zeta\zeta}(I_G \otimes g(\gamma))$, where

$$
\Sigma_{\zeta\zeta} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E[\text{vec}(\zeta_t') - E\text{vec}(\zeta_t')][\text{vec}(\zeta_t') - E\text{vec}(\zeta_t')]'.
$$

Also, under conventional moment assumptions, $T^{-1/2}\sum_{t=1}^{T} \text{vec}(\zeta_t') \otimes Z_t - E[\text{vec}(\zeta_t') \otimes Z_t] \Rightarrow \upsilon$, where $\upsilon \sim N(0, \omega)$, where ω is the average covariance matrix of $\text{vec}(\zeta_t') \otimes Z_t$.

With this notation, Assumption C is satisfied by assuming

(B.3) $ET^{-1/2} \sum\limits_{t=1}^{T} \text{vec}(\zeta_t') \otimes Z_t \rightarrow M$

uniformly in θ. Thus

(B.4a) $m_1(\theta) = \delta[(I_G \otimes g(\gamma)') \otimes I_K]M,$

(B.4b) $m_2(\delta) = (\delta/\delta_0 - 1)\iota_G \otimes e_K,$

(B.4c) $R(\delta_0) = \delta_0^{-1}\iota_G \otimes e_K,$

(B.4d) $\Omega[(\gamma_1, \delta_1), (\gamma_2, \delta_2)] = \delta_1 \delta_2[(I_G \otimes g(\gamma_1)') \otimes I_K]\omega[(I_G \otimes g(\gamma_2)) \otimes I_K].$

Computation of the asymptotic distributions proceeds as follows. Suppose that M, ω, and Q_{ZZ} are known. Given θ, then m_1, m_2, R, and Ω are computed using (B.4). A realization of the random variable υ is obtained as a pseudorandom draw from a $N(0, \omega)$ distribution. Then

$$\mu(\gamma) = \Omega_{\gamma, \delta_0}^{-1/2\prime} \delta_0[(I_G \otimes g(\gamma)') \otimes I_K]M \quad \text{and}$$

$$z(\gamma) = \Omega_{\gamma, \delta_0}^{-1/2\prime} \delta_0[(I_G \otimes g(\gamma)') \otimes I_K]\upsilon.$$

These expressions are then used to compute a realization of the objective functions and their minimizers in Corollary 4 or their nonrobust counterparts such as (2.12). Repeating this for multiple draws of υ gives multiple draws from the limiting distributions of these statistics.

The only information about the DGP required for computing these asymptotic distributions by this Monte Carlo method are the values of M, ω, and Q_{ZZ}. These moments are not readily obtained analytically and instead were computed by averaging their sample counterparts over 5000 Monte Carlo replications generated according to the design in Section 4.2. Given these moments, the asymptotic distributions of the various statistics were computed by numerical minimization of the limiting stochastic process for the objective function. The Taylor series expansion (B.1) was implemented with $m = 6$. To check the sensitivity of the results to the choice of m, the approximation for model M1b was recomputed using $m = 4$, 8, and 10, with negligible change in the results.

REFERENCES

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

ANDERSON, T. W. (1977): "Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Parameter Sequences," *Econometrica*, 45, 509–518.

ANDERSON, T. W., AND T. SAWA (1979): "Evaluation of the Distribution Function of the Two Stage Least Squares Estimate," *Econometrica*, 47, 163–182.

ANDREWS, D. W. K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle and D. McFadden. Amsterdam: North Holland, 2247–2294.

BOBKOSKI, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," unpublished Ph.D. Thesis, Department of Statistics, University of Wisconsin.

CAMPBELL, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.

CAMPBELL, J. H., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

CAMPBELL, J. Y., AND R. J. SHILLER (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

CAVANAGH, C. (1985): "Roots Local to Unity," Manuscript, Department of Economics, Harvard University.

CHAMBERLAIN, G., AND G. IMBENS (1996): "Hierarchical Bayes Models with Many Instrumental Variables," Manuscript, Harvard University.

CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference For Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

CHOI, I., AND P. C. B. PHILLIPS (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113–150.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

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

EICHENBAUM, M., AND L. P. HANSEN (1990): "Estimating Models with Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business and Economic Statistics*, 8, 53–69.

EPSTEIN, L. G., AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

——— (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.

FERSON, W., AND G. CONSTANTINIDES (1991): "Habit Formation and Durability in Aggregate Consumption, Empirical Tests," *Journal of Financial Economics*, 29, 199–240.

FISHER, F. M. (1966): *The Identification Problem in Econometrics*. New York: McGraw Hill.

FUHRER, J., G. MOORE, AND S. SCHUH (1995): "Estimating the Linear-Quadratic Inventory Model: Maximum Likelihood Versus Generalized Method of Moments," *Journal of Monetary Economics*, 35, 115–158.

HALL, A., G. RUDEBUSCH, AND D. WILCOX (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–298.

HANSEN, B. E. (1996): "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12, 347–359.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

HANSEN, L. P., AND K. SINGLETON (1982): "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

HEATON, J. (1993): "The Interaction Between Time-Nonseparable Preferences and Time Aggregation," *Econometrica*, 61, 353–385.

KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

KOCHERLAKOTA, N. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

MADDALA, G. S., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

NEELEY, J. (1994): "A Reconsideration of the Properties of the Generalized Method of Moments Asset Pricing Models," Manuscript, Federal Reserve Bank of St. Louis.

NELSON, C. R., AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

NELSON, C. R., R. STARTZ, AND E. ZIVOT (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed by. R. Engle and D. McFadden. Amsterdam: North Holland, 2113–2247.

PAGAN, A., AND J. ROBERTSON (1997): "GMM and its Problems," Manuscript, Australia National University.

——— (1998): "Structural Models of the Liquidity Effect," *Review of Economics and Statistics*, 80, 202–217.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, *Vol. I*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, 449–516.

——— (1987): "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer Verlag.

SARGAN, J. D. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

SHEA, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, 79, 348–352.

SHILLER, R. J. (1982): "Consumption, Asset Markets and Macroeconomic Fluctuations," *Carnegie-Rochester Series on Public Policy*, 17, 203–238.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

TAUCHEN, G. (1986): "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397–425.

TAUCHEN, G., AND R. HUSSEY (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59, 371–396.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): Weak Convergence and Empirical Processes. New York: Springer.

WANG, J., AND E. ZIVOT (1998): "Inference on a Structural Parameter in Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 66, 1389–1404.

WEST, K. D., AND D. W. WILCOX (1994): "Some Evidence on Finite Sample Distributions of an Instrumental Variables Estimator of the Linear Quadratic Inventory Model," in *Inventory Cycles and Monetary Policy*, ed. by R. Fiorito. Berlin: Springer-Verlag, 253–282.