Near Common Factors and Confidence Regions for Present Value Models

by Stephen R. Blough

Federal Reserve Bank of Boston

Working Paper Series

No. 94-3 August 1994

Near Common Factors and Confidence Regions for Present Value Models

by Stephen R. Blough

August 1994
Working Paper No. 94-3

Federal Reserve Bank of Boston

Near Common Factors and Confidence Regions for Present Value Models

Stephen R. Blough*

Revised August 1994

Abstract

The evidence for excess smoothness of aggregate consumption and excess volatility of stock prices is reexamined, using a method that nests parsimonious trend- and difference-stationary specifications of the forcing processes. The confidence interval for the present value of an innovation to labor income is found to be very wide, so that aggregate consumption may be anything from much too smooth to much too volatile. The confidence interval for the present value of an innovation to dividends is narrower and weakly supports excess volatility in the stock market.

---

*Research Department, Federal Reserve Bank of Boston, Boston, MA 02106-2076. I am indebted to Ken West for comments, and for sharing his RATS programs. I would also like to thank Pedro DeLima and Lou Maccini for comments, John Campbell for providing the Blinder-Deaton data, and Robert Shiller for providing his stock market data. Errors are my own.

Inference in present value models is notoriously sensitive to the time series specification of the forcing process. Consumption, equity prices, and long-term bond yields have all been found to be excessively volatile relative to the predictions of simple present value models when income, dividends, and short rates, respectively, are modelled as trend-stationary processes; these results are reversed or greatly diminished when parsimonious difference-stationary specifications are used instead.[1] Unit root tests and related methods are commonly used to choose between the *non-nested* specifications and the results derived therefrom. However, such procedures carry their own sensitivities to specification as well as problems of finite sample level and power.[2]

This paper proposes an alternative method for inference. Confidence intervals for the present value of an innovation to the forcing process are computed for a model that *nests* the competing parsimonious specifications. The model uncertainty to which inference is sensitive is subsumed in the confidence intervals calculated. The nested model also allows plausible time series behavior which parsimonious ARIMA specifications do not capture. The method avoids entirely issues of the statistical properties of unit root tests.

The proposed method uses difference specifications that permit near common factors and moving average roots near or equal to unity. Inclusion of near common factors is directly contrary to the prescriptions of Box and Jenkins (1976), whose recommendation is based on the poor statistical properties of parameter estimates when such factors are included. Estimation of models with (near) unit moving average roots is also problematic, as documented by Plosser and Schwert (1977) and others. This paper shows how to construct confidence intervals that do not inherit the poor distributional properties of the point estimates and permit valid inference.

This paper applies the proposed near common factor method to the questions of excess smoothness of consumption and excess volatility of stock prices, using postwar quarterly data in the first case and annual data since 1879 in the latter. In both cases the computed confidence regions include the results of both parsimonious trend and difference specifications. Very wide confidence intervals are found for the present value of an income innovation, so that consumption

---

[1]Examples of such results include Deaton (1987) for consumption, DeJong and Whiteman (1991) for equity prices, and Cushing and Ackert (1994) for bond yields.

[2]Stock (1993) provides a recent survey.

might be anywhere from much too smooth to much too volatile. Confidence intervals for the present value of a dividend innovation are narrower and weakly support excess volatility of stock prices. The two examples illustrate the importance of the discount rate per observation for inference in present value models: a higher discount rate reduces the importance of long horizons of the impulse response function, about which the data contain little information.

The next section will discuss the use of near common factors to flexibly represent persistence, with particular attention to their implications for present value estimates. Section 2 will review the undesirable statistical properties of standard methods when applied to processes with near common factors, and propose an alternative method with more attractive behavior. It will also compare the proposed method to fractionally integrated specifications. Section 3 will present the empirical results, and Section 4 concludes.

1. Near-Common Factors and Persistence

A number of authors have made the point that the standard Box-Jenkins "principle of parsimony" for empirical time series specification is not appropriate for investigation of longer-horizon properties. This work predominantly concerns persistence in U.S. real GNP. Cochrane (1988) and Christiano and Eichenbaum (1990) show how parsimonious ARIMA models constrain the relationship of the estimated short- and long-run dynamics of a process, with the observed short-run dynamics predominant in determining the estimates. Clark (1988) shows how the slow, partial mean reversion estimated by an unobserved components model of GNP (see also Watson 1986 and Cochrane 1988) implies near common factors in an ARIMA representation, and gives simulation evidence of the poor distributional properties of near common factor estimates.

The importance of near-common factors in modelling persistence may be illustrated with the ARIMA(1,1,1) model

(1) $$(1-\rho L)\Delta y_t = (1+\theta L)\varepsilon_t$$

where L is the lag operator, Δ is the difference operator ($\Delta \equiv 1-L$), ε_t is a white noise innovation, and ρ and θ are respectively autoregressive and moving average parameters. In this paper, ρ is assumed to lie in the open interval (-1,1), and θ is assumed to lie in the half-open interval [-1,1).

A unit moving average root (θ=-1) cancels the difference operator and implies a stationary AR(1) process; otherwise ($\theta \in (-1,1)$) the processes generated by (1) have a single unit root.

The impulse response function for the levels of $\{y_t\}$ as defined in (1) is given by

$$\mu(s) = 1 + (\rho + \theta) \cdot \sum_{i=0}^{s-1} \rho^i \tag{2}$$

for horizon $s = 1,\ldots,\infty$. Thus, the long-run effect of an innovation is given by

$$\mu(\infty) = \frac{1+\theta}{1-\rho} \tag{3}$$

while the rate at which the impulse response function approaches this limit is controlled by the autoregressive parameter ρ.

If θ lies in the interval (-1, -ρ), (3) indicates that a unit innovation will have a long-run effect between zero and one, which property is denoted "partial mean reversion." If, in addition, ρ is near one, this limit will be approached slowly. For example, $\rho = 0.8$, $\theta = -0.9$ implies a long-run effect of 0.5, and the impulse response function has declined halfway to this long-run level three periods after the shock. $\rho = 0.9$, $\theta = -0.95$ give the same long-run effect, but the halfway point is not reached until six periods have passed.

Since slow partial mean reversion requires ρ to be near one and -θ to be between ρ and one, it requires -θ to be near ρ and also near one. In such a case, the process (1) will be difficult to distinguish from a random walk (the moving average factor nearly cancels the autoregressive factor) and from a stationary AR(1) (the moving average factor nearly cancels the difference operator). The principle of parsimony will prefer to model $\{y_t\}$ as an AR(1) with root near or equal to unity, and thus as a process whose impulse response either decays to zero at rate ρ or does not decay at all.[3] The slow partial mean reversion of the true process will not appear in the parsimonious specifications.

[3]A Monte Carlo experiment reported in Blough (1994) indicates that, with $\rho = 0.8$, $\theta = -0.9$, T=200, the Schwartz criterion chooses the true model for only three of 1000 replications.

The representation (1) nests the random walk and stationary AR(1) models as special cases, while allowing slow partial mean reversion. Model (1) also allows mean perversion: when $\rho > -\theta$, the long-run effect of a shock is greater than one.[4] Campbell and Mankiw (1987) report results for GNP with this property. If this condition holds but $-\theta$ is near to ρ, the principle of parsimony will eliminate the resulting near common factors, suggesting that standard methods will understate the extent to which the data admit mean perverting representations. Blough (1994) uses the method proposed in this paper to compute confidence intervals for the univariate GNP impulse response function; the upper 95 percent confidence bound lies substantially *above* that reported by Campbell and Mankiw.

The present value of an innovation in a general ARIMA(p,1,q) model is given by

$$\Psi(\rho(L),\theta(L),\beta) = \frac{1}{1-\beta}\cdot\frac{1+\theta(\beta)}{1-\rho(\beta)} \tag{4}$$

where β is the (constant) discount factor and $\theta(L)$ and $\rho(L)$ are, respectively, the autoregressive and moving average polynomials. The annuity value of an innovation is

$$\psi(\rho(L),\theta(L),\beta) \equiv (1-\beta)\cdot\Psi = \frac{1+\theta(\beta)}{1-\rho(\beta)} \quad . \tag{5}$$

In the special case of the ARIMA(1,1,1) given in (1), this annuity value is

$$\psi(\rho,\theta,\beta) \doteq \frac{1+\theta\beta}{1-\rho\beta} \quad . \tag{6}$$

Figure 1 shows the importance for estimation of present value models of the possible existence of slow, mean reverting and mean perverting processes. The graph show level contours of $\psi(.)$ defined in (6), plotted against ρ and θ for discount rate β=0.995. Note that the contours are logarithmically scaled, and that each contour represents a doubling of the function.

[4]Unobserved components models with orthogonal innovations are unable to represent such behavior, as noted by Watson (1986).

The contours fan out from the point (off the graph) $\rho=1/\beta$, $\theta=-1/\beta$. They lie very close together in the upper-left-hand corner of the graph, where ρ is near one and θ is near minus one. This is exactly the region where slow mean reverting and mean perverting processes are found. Thus, if values of the parameters in this region are admissible, very precise estimation of the parameters is necessary for precise estimation of the present value function. Further, accurate determination of confidence intervals for the present value function requires careful determination of the joint confidence region for ρ and θ.

2. Confidence Regions for Near Common Factors

Standard Box-Jenkins analysis chooses the lowest order representation of a time series consistent with the data, and hence eliminates from consideration specifications with near common factors. This constraint can improve short-term forecasts even when the true process does have near common factors, but can introduce severe specification biases in inference about long horizon properties.

Estimation and inference are problematic when near common factors are permitted. This section will review the inadequacy of standard methods, and then propose an alternative. The results deal formally with an ARIMA(1,1,1) model. An approach to more general models will be suggested, which will be used for the empirical work in Section 3.

2.1 Problems with Standard Methods

The ARIMA(1,1,1) model is nonlinear in the data. Standard methods (non-linear least squares and maximum likelihood, denoted NLLS and ML respectively) arrive at estimates and form variance / covariance matrices using an assumption that the objective function (the sum of squared errors or minus the log-likelihood) is approximately quadratic in the neighborhood of the minimum. In particular, the estimated variance / covariance matrix $\hat{V}$ is formed using the Hessian of the objective function at the minimum, or an approximation thereof. This estimate implies a confidence ellipse for the parameters $\delta \equiv (\rho\,,\,\theta)'$ defined by

$$(\delta-\hat{\delta})'\hat{V}^{-1}(\delta-\hat{\delta})=\alpha \qquad (7)$$

where α is an appropriate critical value. If the quadratic approximation is accurate, the ellipse (7) is a good approximation of a level set of the log-likelihood function. Hence, tests based on (7) are approximately likelihood ratio tests.[5]

This procedure, which we will call the VCV method (in either its NLLS or ML flavors), is asymptotically valid as long as (i) ρ and θ are each in the interior of the interval (-1,1), and (ii) $\rho \neq -\theta$. These conditions require the true process to be stationary, invertible, and without exact common factors. However, the VCV method has poor finite sample properties when the true process has near common factors or moving average parameter near -1.

All points on the line $\rho = \theta$ are observationally equivalent and therefore have the same likelihood. Level sets of the likelihood function cannot be elliptical in this region, implying that the likelihood is poorly approximated by a quadratic. Parameter estimates have poor distributional properties, as documented by Dent and Min (1978), Ansley and Newbold (1980), and Clark (1988), among others. Confidence regions formed using (7) are badly behaved for two reasons. First, because they are centered on the parameter estimates, they inherit the unpleasant distributional properties of the latter. Second, because they are by construction elliptical, they are poor approximations of level sets of the likelihood function. Ansley and Newbold find severe size distortion in *t*-tests of individual parameter values.

Table 1 shows how the Ansley and Newbold results extend to the properties of joint confidence ellipses for ρ and θ. The columns labelled "VCV Method" give the empirical probabilities that nominal 95 percent confidence ellipses fail to cover the true parameter values (so that ideally the entries should all be 0.05). Results are for 124 observations (for use in the applications in Section 4); estimation was performed using a standard NLLS algorithm.[6] The

[5]The ellipse (7) is based on a Wald statistic, and is known at least in the linear model to give a smaller confidence region in finite samples than the likelihood ratio statistic (Berndt and Savin 1977). This effect is not the source of the problems to be discussed, however. In the experiments below, Wald and likelihood ratio confidence regions differ trivially for models without near common factors.

[6]Specifically, the BOXJENK command in RATS version 4.0. No attempt was made to search over different starting values; replications that did not converge in 50 iterations, or converged to points outside the parameter space, were discarded.

Table 1

Monte Carlo Results

Empirical probabilities of estimated 95 percent confidence regions failing to cover true parameter values.
T=124; 1000 replications; standard errors in parentheses.

		VCV Method		LRCR Method	
ρ	θ	$\chi^2(2)$	F(2,122)	$\chi^2(2)$	F(2,122)
0.60	0.00	0.077 (0.008)	0.070 (0.008)	0.049 (0.007)	0.046 (0.007)
0.00	0.00	0.364 (0.015)	0.363 (0.015)	0.075 (0.008)	0.070 (0.008)
0.10	-0.20	0.363 (0.015)	0.360 (0.015)	0.072 (0.008)	0.068 (0.008)
0.85	-0.90	0.437 (0.016)	0.433 (0.016)	0.064 (0.008)	0.060 (0.008)
0.95	-0.90	0.325 (0.015)	0.321 (0.015)	0.084 (0.009)	0.082 (0.009)
0.99	-0.90	0.138 (0.011)	0.135 (0.011)	0.059 (0.007)	0.054 (0.007)
0.80	-0.95	0.434 (0.016)	0.429 (0.016)	0.042 (0.006)	0.040 (0.006)
0.85	-0.95	0.460 (0.016)	0.451 (0.016)	0.045 (0.007)	0.043 (0.006)
0.90	-0.95	0.416 (0.016)	0.414 (0.016)	0.072 (0.008)	0.068 (0.008)
0.99	-0.95	0.326 (0.015)	0.321 (0.015)	0.090 (0.009)	0.083 (0.009)
0.80	-0.99	-	-	0.025 (0.005)	0.025 (0.005)
0.85	-0.99	-	-	0.029 (0.005)	0.026 (0.005)

continued

Table 1 (cont.)

		VCV Method		LRCR Method	
ρ	θ	$\chi^2(2)$	F(2,122)	$\chi^2(2)$	F(2,122)
0.80	-1.00	0.493 (0.016)	0.486 (0.016)	0.028 (0.005)	0.024 (0.005)
0.85	-1.00	0.496 (0.016)	0.489 (0.016)	0.026 (0.005)	0.025 (0.005)
0.90	-1.00	0.450 (0.016)	0.445 (0.016)	0.046 (0.007)	0.044 (0.006)
0.95	-1.00	0.467 (0.016)	0.465 (0.016)	0.051 (0.007)	0.046 (0.007)
0.99	-1.00	0.463 (0.016)	0.455 (0.016)	0.078 (0.008)	0.071 (0.008)

column headed "$\chi^2(2)$" takes the critical value α in (7) from the $\chi^2(2)$ distribution. The column headed "F(2,122)" implements a finite sample correction by setting α equal to twice the 5 percent critical value of the F(2,122) distribution (under a normal approximation, one-half the LHS of (7) has this distribution).

For $\rho = 0.6$, $\theta = 0$, the 95 percent confidence ellipses fail to cover the true parameter value with about 7 percent probability. The remaining parameter pairs have either near common factors, moving average roots near -1, or both, and the results are very bad indeed. With the exception of the strongly mean-perverting process $\rho = 0.99$, $\theta = -0.90$ (for which the rejection probability is 13.5 percent), the confidence ellipses fail to cover the true parameter values with probabilities ranging from 30 percent to 50 percent.[7] The VCV method is clearly useless for inference when near common factors and/or slow mean reversion or perversion are possibilities.

[7] The entry for $\rho = \theta = 0$ gives the probability that the ellipse fails to cover that point. The probability of the ellipse failing to cover observationally equivalent points along the common factor line - the probability of falsely rejecting white noise - is much lower.

These results are bad enough, but another complication arises in translating the confidence regions for the parameter vector δ into confidence intervals for the present value function ψ. The function ψ is nonlinear; the VCV method forms confidence intervals using

$$\text{var}(\psi) \cong \psi_\delta(\hat{\delta})' \hat{V} \cdot \psi_\delta(\hat{\delta}) \tag{8}$$

where $\psi_\delta(\hat{\delta})$ is the vector of first derivatives of ψ with respect to ρ and θ evaluated at the parameter estimates. The approximation (8) is based on a linearization. Accuracy of confidence intervals constructed using (8) depends on the linearization being adequate over the confidence ellipse for ρ and θ. Figure 1 suggests that this linearization will be inaccurate when the confidence region for δ is elongated along the $\rho=-\theta$ diagonal: a given distance gives a much smaller change in the function in the middle of the graph than in the upper left corner. The consequences of this approximation in practice will be apparent in Section 3.[8]

2.2. Likelihood Ratio Confidence Regions

The discussion in the previous section emphasized the failure of the quadratic approximation to the likelihood function as a cause of the poor finite sample properties of the VCV method for near common factor models. This failure means that a VCV confidence ellipse can be a poor approximation to a level set of the likelihood function.

The alternative procedure proposed here calculates a level set of the likelihood function directly. The procedure defines a joint confidence region for ρ and θ as the set of points in the parameter space that are not rejected by a likelihood ratio test.

Definition. A Likelihood Ratio Confidence Region (LRCR) for the ARIMA(1,1,1) model (1) is defined by:

---

[8]Were the confidence regions for (ρ,θ) accurate, this problem could be at least mitigated by using corrections derived from an Edgeworth expansion (Phillips and Park 1988). That approach is not pursued in this paper, because a more direct method for computing confidence intervals for ψ will be available.

$$\left\{(\rho,\theta):2\cdot\left(\log L\left(\hat{\rho},\hat{\theta}\right)-\log L(\rho,\theta)\right)<\alpha\right\}$$

where $(\hat{\rho}\ ,\hat{\theta}\ )$ are the maximum likelihood estimates and α is an appropriate critical value (discussed below).

An LRCR need not be an ellipse. In fact, since the likelihood function takes the same value for every point on the line $\rho{=}{-}\theta$, the entire line must fall in the LRCR if any point does so.[9] Since the likelihood function is continuous, inclusion of the common factor line in the LRCR implies inclusion of near common factor points near that line along its entire length, including points with slow partial mean reversion and mean perversion. Further, unlike a standard confidence ellipse, an LRCR need not be centered on the parameter estimates. Thus, while the maximum of the likelihood function is found and therefore maximum likelihood estimates are implied, the poor distributional properties of the maximum likelihood estimates are not inherited by the LRCR.

Of course, in practice the likelihood function can only be evaluated at a finite number of points. In what follows, it is evaluated at all points on the grid

(9)
$$\begin{gathered}\rho\in\{-0.99,-0.95,-0.90,\ldots,0.90,0.95,0.99\}\\ \in\{-1.0,-0.99,-0.95,-0.90,\ldots,0.90,0.95,0.99,1.0\}\end{gathered}$$

Two sources of approximation error are introduced by use of a grid. Unless the true maximum is at a point on the grid, computed likelihood ratio statistics will understate the true likelihood ratio, which may lead to inclusion of points that should be rejected. On the other hand, the level set of the likelihood function will not pass exactly through points on the grid, so points off the grid that should be included may be ignored. These errors have opposite signs, and the results suggest that neither is large.

---

[9]This is true only if the exact likelihood is computed. Approximations that assume presample errors to be zero yield different values of the objective function for observationally equivalent models along the common factor line.

When the critical value α is taken from the $\chi^2(2)$ distribution, the LRCR is asymptotically justified for all elements of the set $\{\rho \in (-1,1),\ \theta \in (-1,1),\ \rho \neq -\theta\}$ (in fact, the LRCR is asymptotically equivalent to the VCV confidence ellipses for these points). The asymptotic theory does not apply for points on the common factor line, nor for points on the line $\theta = -1$. Hansen (1991) presents asymptotic theory for the likelihood ratio statistic for $\rho = -\theta$. He argues that a χ^2 critical value performs poorly in finite samples in this case. However, Hansen uses the critical values for the $\chi^2(1)$ rather than the $\chi^2(2)$.[10]

Monte Carlo results supporting the use of the LRCR are presented in Table 1, above.[11] The columns of the table headed "LRCR Method" give empirical probabilities that computed 95 percent LRCRs fail to cover true parameter values; as with the VCV method, results are for T=124 and 1000 replications, and are given using both $\chi^2(2)$ and F(2,122) critical values.

The results are very encouraging. The LRCR improves on the VCV confidence ellipse even for the parameter pair ($\rho = 0.6$, $\theta = 0$) for which the latter performed well, having a 5 percent rather than a 7 percent rejection probability. More importantly, the LRCR performs quite well for the processes for which the VCV ellipse failed miserably. When $\rho = -\theta$ (and the likelihood ratio test is not asymptotically χ^2), the 95 percent LRCR fails to include the common factor line 7 percent of the time. None of the estimated probabilities exceeds 9 percent, in comparison to the typical 30 percent to 40 percent probabilities found for the VCV confidence ellipses.

The most noticeable distortion of the LRCR occurs for processes with $\theta = -1$. As ρ declines away from the common factor line, the LRCR includes the true parameter value too often. The probability of rejecting the true parameter value falls to 2.5 percent when ρ is 0.85 or less. Some distortion also occurs for $\theta = -0.99$, but it has largely disappeared for $\theta = -0.95$.

---

[10]The condition $\rho = -\theta$ appears to impose only a single restriction, hence the single degree of freedom in Hansen's paper. On the other hand, the unrestricted model has two degrees of freedom while the restricted model has none, motivating the presumption of two restrictions in this paper. The proof is in the pudding, of course; the results presented below show that the $\chi^2(2)$ critical values perform quite well.

[11]For each replication, the log-likelihoods were evaluated over the grid (9) using the Kalman Filter (see Harvey 1981, ch. 5). Computations were performed in GAUSS.

Table 2

Sample Size Effects: LRCR Method

Empirical probabilities of estimated 95 percent confidence regions failing to cover true parameter values. 1000 replications, standard errors in parentheses.

		T=50		T=124		T=500	
ρ	θ	$\chi^2(2)$	F(2,48)	$\chi^2(2)$	F(2,122)	$\chi^2(2)$	F(2,498)
0.00	0.00	0.058 (0.007)	0.049 (0.007)	0.049 (0.007)	0.046 (0.007)	0.089 (0.009)	0.087 (0.009)
0.90	-0.95	0.037 (0.006)	0.031 (0.005)	0.072 (0.008)	0.068 (0.008)	0.052 (0.007)	0.050 (0.007)
0.95	-1.00	0.057 (0.007)	0.041 (0.006)	0.051 (0.007)	0.046 (0.007)	0.012 (0.003)	0.012 (0.003)

In the other direction, there is some tendency for the LRCR to miss the true parameter values too often for slow, mean perverting processes with near common factors, e.g. $\rho = 0.99$, $\theta = -0.95$.

Table 2 presents evidence on the effect of sample size on the performance of the LRCR. For three parameter pairs, empirical rejection probabilities are given for 50, 124, and 500 observations. The results are consistent with asymptotic theory. Performance improves with sample size for the point $\rho = 0.9$, $\theta = -0.95$, where the LRCR is asymptotically justified. Performance deteriorates with sample size for the points $\rho = \theta = 0$ and $\rho = 0.95$, $\theta = -1$ for which the asymptotic theory does not apply. Overall, performance is quite acceptable for T=50 and T=124, but the results for T=500 suggest that the LRCR should be used with caution for large samples.[12]

[12]Hansen (1991) provides asymptotic theory for the likelihood ratio statistic in the presence of common factors. Asymptotic results for processes with ρ converging to one and θ converging to minus one are given by Nabeya and Perron (1991). These results may be useful in improving the performance of the LRCR for large samples.

Once the LRCR for ρ and θ is determined on the grid, computation of confidence sets for functions of the parameters is straightforward. The function is evaluated at each of the points in the LRCR; a confidence region is formed by a suitable closure of these points. The confidence interval for the present value function is simply the minimum and maximum of $\psi(\rho,\theta)$ over the LRCR. A simple graphical representation results from plotting the LRCR on top of a contour map of the function, as will be done below. Such graphs may provide insight into the relationship between the confidence region for (ρ,θ) and the confidence intervals for less well behaved functions.

Use of the LRCR is not computationally burdensome for the two-parameter ARIMA(1,1,1) model, but full computation of LRCRs for higher-order processes is not practical. In this case, inference can proceed by fitting a parsimonious model using standard methods, e.g.

(10) $$\left(1-a(L)\right)\Delta y_t = \left(1+b(L)\right)u_t \cdot$$

An LRCR can then be computed for the parameters of an ARMA(1,1) model fitted to the residuals

(11) $$(1-\rho L)u_t = (1+\theta L)\varepsilon_t \cdot$$

This implies the near common factor model

(12) $$(1-\rho L)\left(1-a(L)\right)\Delta y_t = (1+\theta L)\left(1+b(L)\right)\varepsilon_t$$

for the original series. The LRCR for (11) will imply a confidence interval for

(13) $$\tilde{\psi} \equiv \frac{1+\theta\beta}{1-\rho\beta}$$

A confidence interval for ψ can be formed from

(14) $$\psi \equiv \frac{1+b(\beta)}{1-a(\beta)} \cdot \tilde{\psi}$$

treating the estimates $a(L)$ and $b(L)$ as fixed. This will understate uncertainty about ψ ; however, in the applications below, uncertainty about $\tilde{\psi}$ swamps the omitted variance.

2.3 Comparison to Fractionally Integrated Models

Fractionally integrated, or long-memory, models have recently been used for flexible representation of the low-frequency properties of time series. An ARFIMA (autoregressive, fractionally integrated, moving average) model is given by

(15) $$a(L)(1-L)^d y_t = b(L)\varepsilon_t$$

where the roots of a(L) and b(L) lie outside the unit circle and the fractional integration parameter d may take values on the real line. Such models are stationary and have invertible moving average representation if $d \in$ (-1/2,1/2), and by extension are difference stationary if $d \in$ (1/2,3/2) (see Sowell 1992 and the references therein).

Standard stationary and difference stationary (ARMA and ARIMA) representations correspond to the special cases d=0 and d=1, respectively. Allowing d to take non-integer values allows the ARFIMA model to represent long-run dynamics in ways that are unavailable in parsimonious ARMA and ARIMA specifications. The impulse response function of an ARFIMA model displays partial mean reversion for $d \in$ (1/2,1). In comparison to the near common factor model (1), the impulse response function converges to its long-run value more slowly; decay is hypergeometric rather than geometric.

Suppose we wish to estimate a model that nests random walk and stationary AR(1) processes, while allowing partial mean reversion. The near common factor model (1) does so by considering $\rho < 1$ and letting θ vary from -1 (implying a stationary AR(1) process) to $-\rho$ (implying a random walk). The same nesting can be accomplished with the ARFIMA model

(16) $$(1-\rho L)(1-L)^d y_t = \varepsilon_t$$

with d varying from zero to one. However, the statistical properties of estimates of (16) must suffer from difficulties conceptually similar to those produced by near common factor models.

The points (ρ=0,d=1) and (ρ=1,d=0) are observationally equivalent and therefore approximation of the likelihood function by a quadratic in the vicinity of these points will be inaccurate.

Existing results on the properties of ARFIMA models (e.g. Cheung and Diebold 1990, Sowell 1992) do not consider joint estimation of ρ and d in (16) when ρ may be near one. In practice, the fractional differencing parameter d is a substitute for autoregressive roots near unity. Diebold and Rudebusch (1989) examine consumption smoothness in an ARFIMA specification. Their results, like those reported below, show a wide confidence interval for the present value of an innovation to income. They are not as wide as those obtained below with a near common factor model, and in particular do not include the results of parsimonious trend stationary specifications. This results from the restrictions imposed by their specification procedure, and not from any additional information that their procedure employs.

3. Application to Present Value Models

As discussed in the introduction, inference in present value models is commonly found to be fragile with respect to the level of integration assumed for the stochastic process of the fundamentals. The confidence regions proposed in Section 2 may be used to provide easily interpreted results that are not subject to this fragility. Since the confidence regions are constructed using the residuals of a parsimonious difference specification, the estimates of such a specification necessarily lie in the confidence region. Such a specification implies ρ=$-\theta$; by continuity, the confidence region will also include points where ρ is near -θ, and in particular points where ρ is less than but near to one and θ=-1. These latter points are trend stationary processes; thus, trend stationary processes with roots near unity are necessarily included in the constructed confidence intervals. In this way, results using the near common factor model can determine whether the apparently conflicting results of level and difference specifications are in fact elements of the confidence region of a nested model.

This section provides examples of such analyses for two prominent present value models: the permanent income hypothesis of consumption and the present value model of stock prices. The results extend work by West (1988a,b) by explicitly allowing near common factors in the ARIMA representations of labor income and dividends.

The data samples contain approximately the same number of observations in the two examples (T=124 for labor income, T=120 for dividends). In both cases a constant discount factor is used.[13] However, the appropriate per-period discount factor for the annual dividend data is substantially lower than that for the quarterly income data. Because a lower discount factor implies smaller weights on long horizons of the impulse response function, this difference will have an important effect on the results.

3.1 Excess smoothness of consumption?

Deaton (1987) indicated that when labor income is modeled as a trend stationary process, consumption appears to be too volatile relative to permanent income, while if labor income is modeled as a difference stationary process, consumption appears to be too smooth. West (1988a) showed that the latter result is robust to a variety of ARIMA specifications, and furthermore continues to hold when account is taken of the information advantage of agents over the econometrician.[14] He concludes that "If . . . income does have a unit root . . ., a stylized fact is that consumption is insensitive to news about income."

West's conclusion is based on analysis of a number of ARIMA models of the Blinder - Deaton (1985) postwar quarterly data on labor income. West does not restrict the analysis to parsimonious parameterizations: while Box-Jenkins criteria select an ARIMA(1,1,0) specification, West analyzes parameterizations up to ARIMA(2,1,2).

In the absence of private information of agents, the PIH implies that the variance of consumption changes should be $\psi^2 \cdot \sigma^2$, where ψ^2 is the annuity value of an innovation defined

[13]Much of the recent literature on both hypotheses permits the discount rate to vary. The results reported here concern the extent to which the constant discount rate hypothesis is consistent with the data, and thus the extent to which time-varying discount rates are empirically required.

[14]Thus West's results are not subject to the criticism of Quah (1990) regarding the permanent/ transitory decomposition observed by agents. West's results are however subject to the caveat raised by Flavin (1988): the correction for superior information of agents is valid only if the permanent income null is true. Thus, in Flavin's view, West provides valid evidence against the permanent income null, but not necessarily in favor of an excess smoothness alternative. These points are moot unless we are confident that consumption is too smooth relative to the univariate labor income process.

in (5) and σ^2 is the innovation variance of the labor income process. For each estimated process, West computes an estimate of ψ from the NLLS estimates of the ARIMA parameters, and computes standard errors using the VCV method described in Section 2. West found statistically significant evidence of excess smoothness for all specifications used, apparently providing robust evidence for the conclusion. However, VCV confidence intervals for ψ were shown above to be unreliable.

Construction of an LRCR for ψ begins with a parsimonious representation. In this case, such is given by the ARIMA(1,1,0)

$$(1-0.44L)\Delta y_t = u_t \tag{17}$$

where the sample runs from 1954:I to 1984:IV (124 observations). The next step constructs an LRCR for the parameters ρ and θ of an ARMA(1,1) fitted to the residuals u_t. This implies an ARIMA(2,1,1) process for labor income, with one autoregressive root (given in (17)) treated as fixed.

The resulting LRCR for ρ and θ is graphed in Figure 2. The $\chi^2(2)$ critical values are used; the results in Table 1 suggest that these will yield a confidence region that is somewhat too small. Also plotted are contour lines for

$$\psi = \frac{1}{1-0.44\beta} \cdot \frac{1+\theta\beta}{1-\rho\beta} \tag{18}$$

using β=0.995, as in West. Table 3(a) gives the resulting confidence interval for ψ, in comparison to those found by West for ARIMA(1,1,0) and ARIMA(2,1,1) specifications, and also shows the consequences for the excess smoothness inferences (results for West are taken from his corrected Tables 1 and 2, 1992).[15]

The effect of using the LRCR is dramatic. The 95 percent confidence interval for ψ is almost six times as wide as that found using the VCV method on an ARIMA(2,1,1) specification.

---

[15]The confidence intervals are not exactly comparable. West's computed standard error allows for uncertainty about all three parameters. Also, West's standard errors are corrected for possible heteroskedasticity.

Table 3

Confidence Interval Comparison

(a) Consumption ($\beta = 0.995$)

Model	Method	95% CI for ψ	95% CI for σ_H^2	σ_c^2
ARIMA(1,1,0)	VCV	(1.39 , 2.19)	(687.8 , 3369)	246.1
ARIMA(2,1,1)	VCV	(1.33 , 2.49)	(726.2 , 3527)	236.9
ARIMA(2,1,1)	LRCR	(0.085 , 6.51)	(4.67 , 27386)	236.9

Note: First two lines are derived from Tables 1 and 2 of West (1992). Last line is computed as described in text; confidence intervals for that line treat as fixed the innovation variance of income, and well as one autoregressive root taking the value 0.44.

(b) Equity Prices ($\beta = 0.94$)

Model	Method	95% CI for Ψ	95% CI for $\Psi^2\sigma_H^2$	σ_P^2
ARIMA(4,0,0)	VCV	(4.56 , 17.0)	(3.12 , 43.1)	252.8
ARIMA(4,1,0)	VCV	(7.17 , 29.7)	(7.91 , 136)	244.4
ARIMA(1,1,2)	LRCR	(5.35 , 23.3)	(4.05 , 76.9)	254.6

Note: First two lines are derived from West (1988b). Last line is computed as described in text. All lines treat as fixed the discount factor and the innovation variance of dividends. Last line treats as fixed one moving average root taking the value 0.35.

The LRCR confidence interval for the variance in permanent income is (4.67, 27386) if the innovation variance σ^2 is treated as fixed. This confidence interval easily includes the variance of consumption changes estimated as 236.9. Thus, the apparent robustness of the excess smoothness evidence is an artifact of the use of the VCV method in the presence of near common factors.

The low end of the LRCR confidence interval comes from the trend stationary specification $\rho = 0.90$, $\theta = -1$.[16] This result coincides with Deaton's observation that consumption

[16]There is no reason to believe that this point is spuriously included in the confidence set: the empirical rejection probability in Table 1 is insignificantly different from 5 percent. In fact, the likelihood ratio statistic for this point is 4.82, which has an empirical p-value of 9 percent. If the grid were arbitrarily fine, the borderline processes would all have likelihood ratio statistics

appears too volatile if income is estimated as a stationary process. The LRCR confidence interval shows how this result fits into a nested analysis.

Further insight into the breakdown of the VCV confidence interval is obtained by factoring West's ARIMA(2,1,1) model as

$$(1-\rho L)(1-0.55L)\Delta y_t = (1+\theta L)\varepsilon_t \tag{19}$$

and estimating the parameters of this equation by NLLS. The ellipse resulting from the estimated covariance matrix for (ρ,θ) is plotted in Figure 3. The ellipse covers even more strongly trend-reverting processes than the LRCR, but these are not reflected in the VCV confidence interval for ψ. The culprit is the linearization of ψ implicit in (8). The estimated variance assumes that ψ is approximately linear in ρ and θ, implying parallel contour lines. The contours are in fact much farther apart in the neighborhood of the estimated parameters (around which the function is linearized) than in the upper left corner of the parameter space, hence the dramatic understatement of the confidence interval when the linearization is used. The LRCR method both produces a more accurate confidence region for the estimated parameters and provides a more accurate translation to a confidence interval for the function.

4.1 Excess volatility of stock prices?

The analysis of the permanent income hypothesis used a relatively high discount factor (β = 0.995), implying high weights on long horizons of the impulse response function. Application to stock market volatility illustrates the sharper inference provided by more rapid discounting. The Shiller (1981) data, updated by Shiller through 1985 and by the author through 1991, provide 123 annual observations, almost equal in size to the consumption sample. Because the data are annual and because of the well-known "equity premium" in the rate of return on stocks, a significantly lower discount factor per period of observation is appropriate. The value used here is $\beta = 0.94$, which is consistent with the estimates in West (1988b). To grasp the

---

of 5.99; the procedure used here therefore underestimates the LRCR. The adjacent point on the grid ($\rho = 0.85$, $\theta = -1$, $\psi = 0.032$) has a likelihood ratio statistic of 15.8 and is therefore strongly rejected.

importance of the difference, observe that the coefficient of the 100-period impulse response in the present value function is 0.61 when β=0.995, and it is 0.002 when β=0.94. Uncertainty about the long-horizon response receives dramatically lower weight with the lower discount rate.

A parsimonious model of differenced dividends is the ARIMA(0,1,1)

$$\Delta DIV_t = 0.02 + u_t + 0.35u_{t-1} \quad (20)$$

Proceeding as before, an LRCR is computed for the parameters of an ARMA(1,1) fitted to the residuals u_t, implying an ARIMA(1,1,2) model for dividends. This LRCR is graphed in Figure 4. Note that, in comparison to the LRCR for labor income, few mean perverting processes are included in the LRCR for dividends. The evidence presented suggests that the process (20) (represented by the common factor line in Figure 4) displays about the maximum amount of persistence consistent with the data.

The implications for stock price volatility are presented in Table 3(b) in comparison to the results of West (1988b). The present value model implies that

$$\Psi^2 \cdot \sigma_v^2 \geq \sigma_p^2 \quad (21)$$

where Ψ is the present value of an innovation in dividends as defined in (4), σ_v^2 is the innovation variance of dividends, and σ_p^2 is the variance of excess returns

$$\sigma_p^2 \equiv E[\beta(p_{t+1} + d_{t+1}) - p_t]^2 \quad (22)$$

The first two lines of Table 3(b) give West's results using level and difference AR(4) specifications. Since no moving average terms are included in these models, near common factors are not a problem; however, the two specifications are not nested. The third line gives LRCR results for the ARIMA(1,1,2) specification described above.

Because West's confidence intervals reflect uncertainty about the discount factor, which is treated as fixed in the LRCR results, comparability is not as close as for the consumption results. The LRCR confidence interval for Ψ lies within the union of West's level and difference confidence intervals. All reported confidence intervals for the left hand side of (21) lie well below the point estimate for σ_p^2, providing some evidence of excess volatility. However, the estimates of σ_p^2 have high variance (see West), and so the LRCR evidence is not definitive without more careful accounting for uncertainty.

The results do indicate that use of the near common factor model does not necessarily lead to a finding that the data are uninformative in present value models. Rather, it reflects the importance of uncertainty about long horizons more accurately than parsimonious methods. Confidence intervals depend on the discount rate relative to the number of observations; with a low discount rate, long-horizon uncertainty is relatively unimportant.

5. Conclusion.

This paper introduces a well-behaved method for inference for univariate time series models with near common factors. Applications to present value models show how valid inference can be made using models that allow for flexible long-run behavior, and in particular models that nest the results of parsimonious level and difference specifications. The informativeness of the data about the present value of innovations is found to depend sensitively on the discount rate. The analysis finds that postwar quarterly data are simply not informative about the present value of an innovation to labor income, and "excess smoothness" of consumption with respect to income is not a robust result. The evidence is somewhat stronger for excess volatility of stock prices, assuming a stable process for dividends over the much longer sample.

References

1. Ansley, Craig F. and Paul Newbold, 1980, Finite sample properties of estimators for autoregressive moving average models, Journal of Econometrics 13, 159-183.
2. Berndt, E.R. and N.E. Savin, 1977, Conflict among criteria for testing hypotheses in the multivariate linear regression model, Econometrica 45, 1263-1277.
3. Blinder, Alan S. and Angus Deaton, 1985, The time series consumption function revisited, Brookings Papers on Economic Activity 1985:2, 465-511.
4. Blough, Stephen R., 1994, Near observational equivalence of unit root and stationary processes: theory and evidence, manuscript, Federal Reserve Bank of Boston.
5. Box, G.E.P. and G.M. Jenkins, 1976, Time Series Analysis, Forecast, and Control, revised edition. San Francisco, Holden Day.
6. Campbell, John Y. and N. Gregory Mankiw, 1987, Are output fluctuations transitory? Quarterly Journal of Economics 102, 857-880.
7. Cheung, Y. and F.X. Diebold, 1990, On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean, manuscript, University of Pennsylvania.
8. Christiano, Lawrence J. and Martin Eichenbaum, 1990, Unit roots in GNP: do we know and do we care? Carnegie-Rochester Conference Series on Public Policy 32 , 7-62.
9. Clark, Peter K., 1988, Nearly redundant parameters and measures of persistence in economic time series, Journal of Economic Dynamics and Control 12, 447-461.
10. Cochrane, John H., 1988, How big is the random walk in GNP? Journal of Political Economy 96, 893-920.
11. Cushing, Matthew J. and Lucy F. Ackert, 1994, Interest Rate Innovations and the Volatility of Long-term Bond Yields, Journal of Money, Credit and Banking 26, 171-344.
12. Deaton, Angus, 1987, Life cycle models of consumption: is the evidence consistent with the theory? in T.F. Bewley, Advances in Econometrics, Fifth World Congress, Volume II. Cambridge, Cambridge University Press.
13. DeJong, D. and C.H. Whiteman, 1991, The temporal stability of dividends and stock prices: evidence from the likelihood function, American Economic Review 81, 600-617.

14. Dent, Warren and An-Sik Min, 1978, A Monte-Carlo study of autoregressive integrated moving average processes, Journal of Econometrics 7, 23-55.

15. Diebold, Francis X. and Glenn Rudebusch, 1989, Long memory and persistence in aggregate output, Journal of Monetary Economics 24, 189-209.

16. _______________, 1991, Is consumption too smooth? Long memory and the Deaton paradox, Review of Economics and Statistics 73, 1-9.

17. Flavin, Marjorie, 1988, The excess smoothness of consumption: identification and interpretation, NBER Working Paper #2807.

18. Hansen, Bruce E., 1991, Inference when a nuisance parameter is not identified under the null hypothesis, University of Rochester working paper.

19. Harvey, Andrew C., 1981, Time Series Models, Phillip Allan and Humanities Press.

20. Nabeya, S. and P. Perron, 1991, Local asymptotic distributions related to the AR(1) model with dependent errors, manuscript, Princeton University.

21. Phillips, P.C.B. and J. Park, 1988, On the formulation of Wald tests of nonlinear restrictions, Econometrica 56, 1065-1084.

22. Plosser, Charles I. and G. William Schwert, 1977, Estimation of A Non-Invertible Moving Average Process, Journal of Econometrics 6, 199-224.

23. Quah, Danny, 1990, Permanent and transitory movements in labor income: an explanation for 'excess smoothness' in consumption, Journal of Political Economy 98, 449-475.

24. Shiller, Robert J., 1981, Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? American Economic Review 71, 421-436.

25. Sowell, Fallaw, 1992, Modeling long-run behavior with the fraction ARIMA model, Journal of Monetary Economics 29, 277-302.

26. Stock, James H., 1993, Unit Roots, Structural Breaks and Trends, manuscript, Harvard University.

27. Watson, Mark W., 1986, Univariate Detrending Methods with Stochastic Trends, Journal of Monetary Economics 18, 49-75.

28. West, Kenneth D., 1988a, The insensitivity of consumption to news about income, Journal of Monetary Economics 21, 17-33.

29. ___________, 1988b, Dividend innovations and stock price volatility, Econometrica 56, 37-61.

30. ___________, 1992, Errata, Journal of Monetary Economics 29, 337.

FIGURE 1

Contours of $\psi(0.995)$

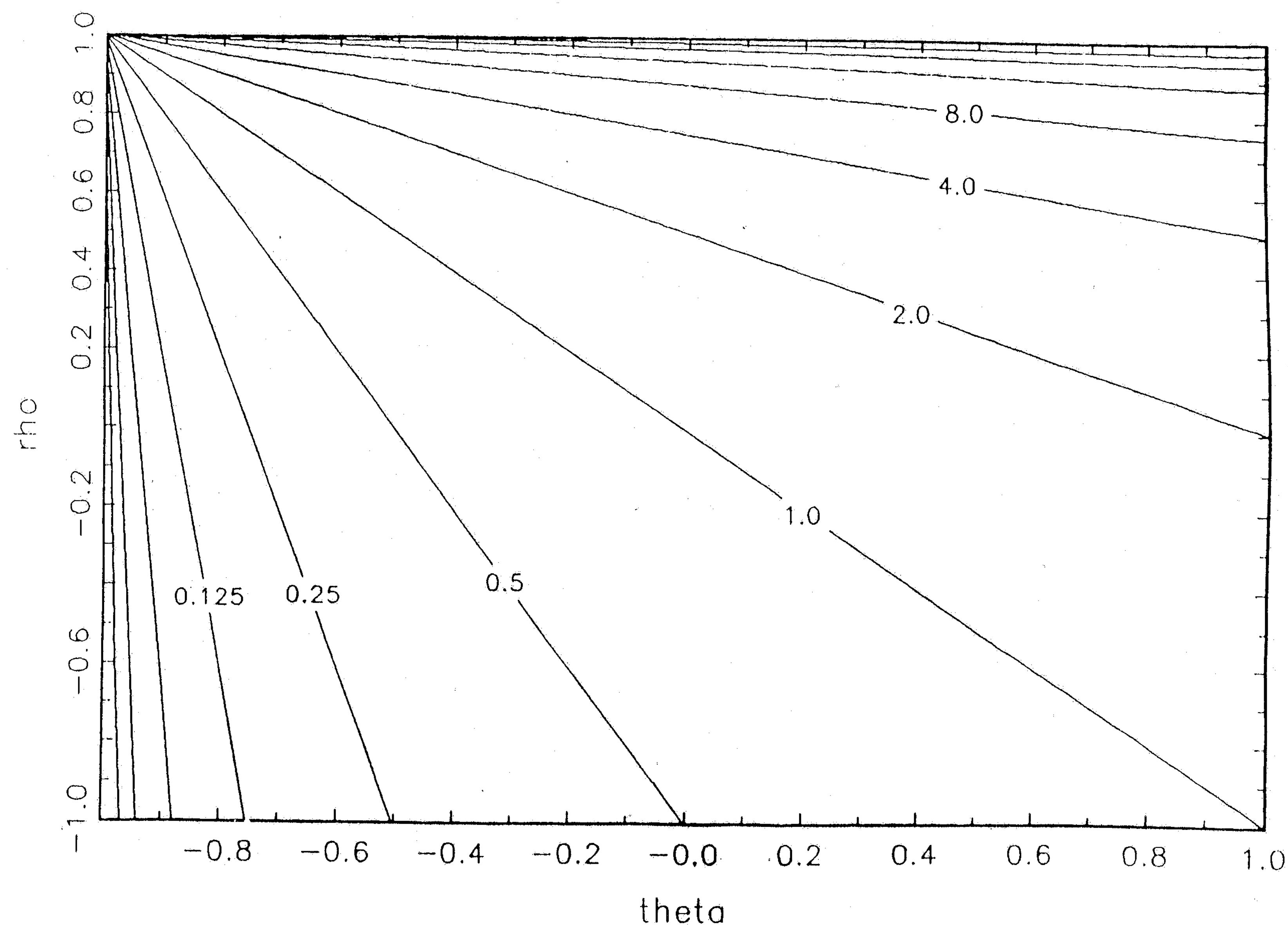


FIGURE 2

95% LRCR for $\Delta y^{l} - 0.44 \Delta y^{l}_{-1}$

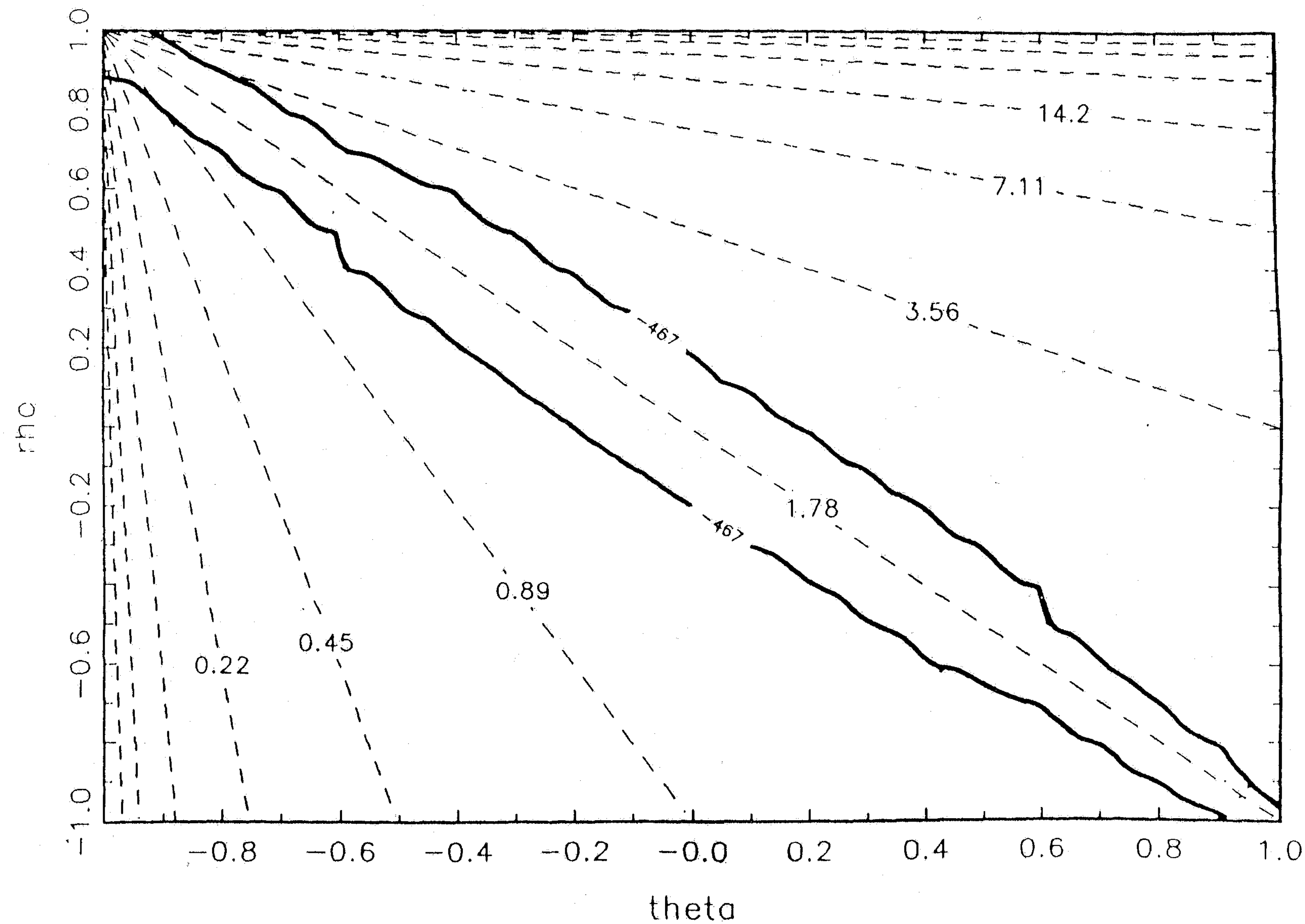


FIGURE 3

95% VCV Ellipse for $\Delta y^{I} - 0.55\Delta y^{I}_{-1}$

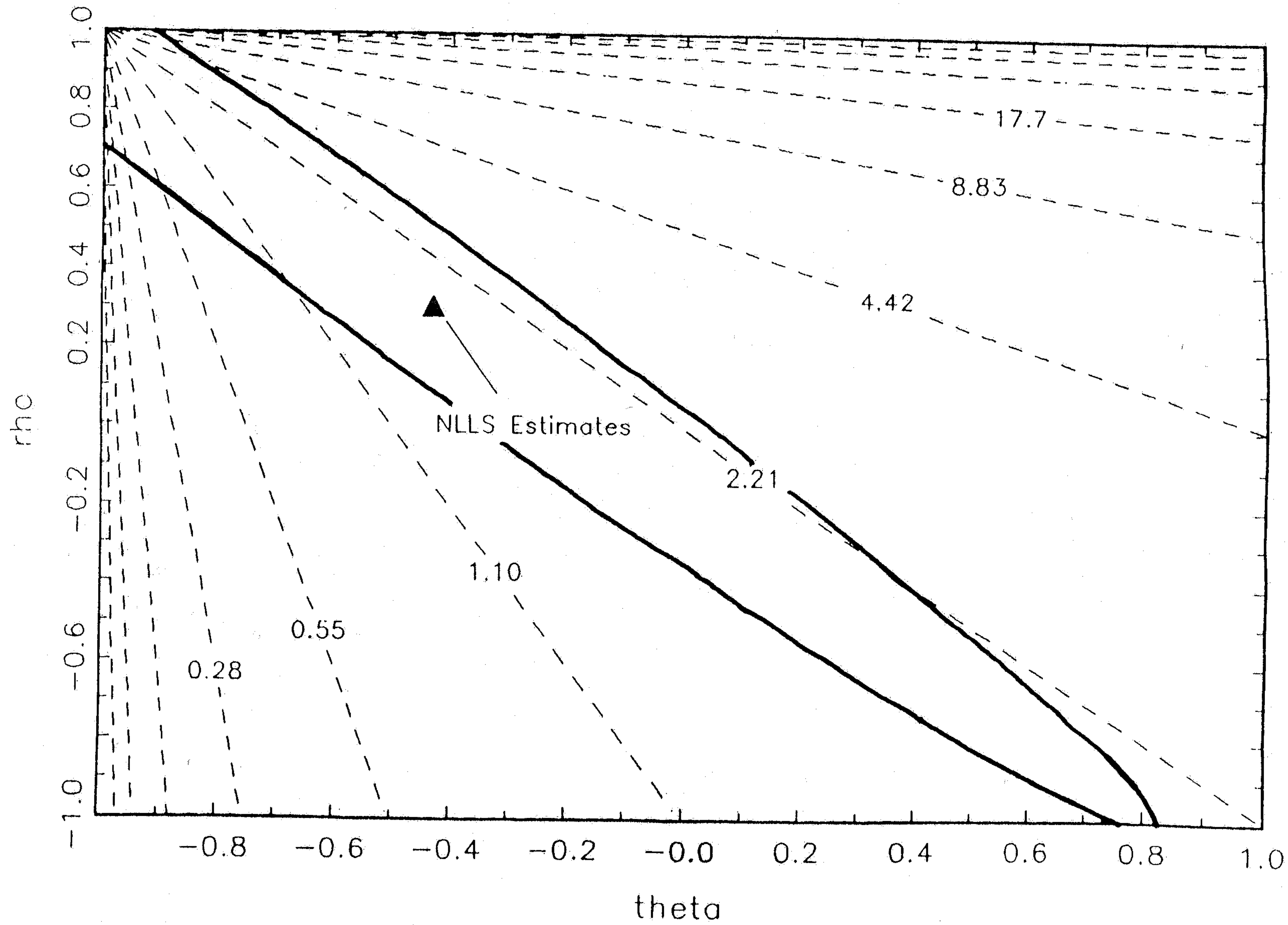
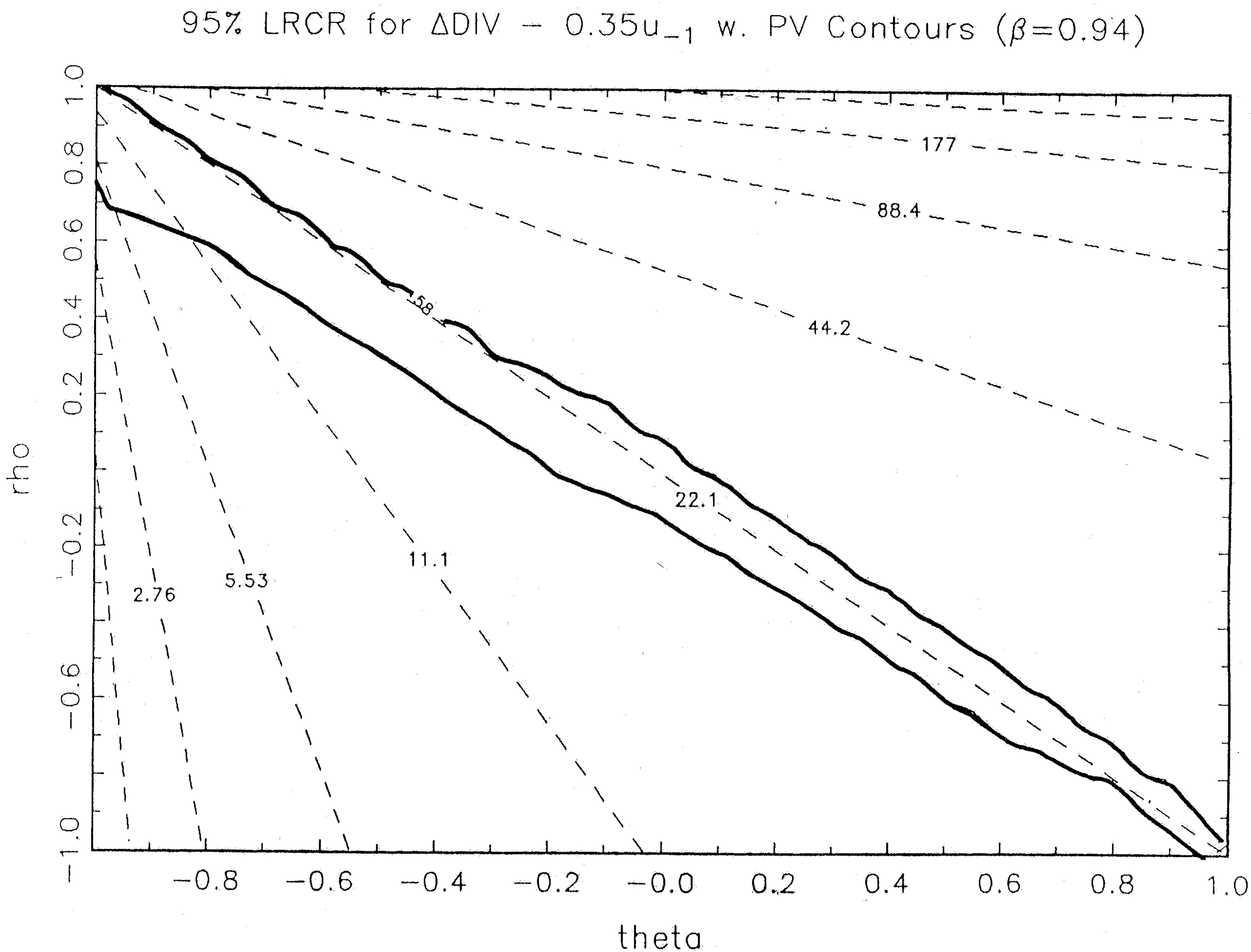


FIGURE 4


95% LRCR for ΔDIV – 0.35u₋₁ w. PV Contours (β=0.94)
177
88.4
44.2
1.58
22.1
11.1
5.53
2.76
rho
1.0
0.8
0.6
0.4
0.2
-0.2
-0.6
-1.0
theta
-0.8
-0.6
-0.4
-0.2
-0.0
0.2
0.4
0.6
0.8
1.0