UvA ECONOMETRICS

Discussion Paper: 2009/12

Conditional bimodality in a structural equations model

Cesar Ariza and Kees Jan van Garderen

(revised October 2010)

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Conditional Bimodality in a Structural Equations Model

Cesar Ariza
University of Amsterdam
Universidad del Rosario, Colombia

Kees Jan van Garderen
University of Amsterdam
UvA-Econometrics
Tinbergen Institute

7 October 2010

Abstract

We investigate the conditional density of the MLE in a simple structural Keynesian model as in Phillips (2006). The marginal distribution is known to be bimodal and various contending explanations have been offered. We give a clear geometric explanation for the bimodality of the MLE. We further show that the degree of bimodality depends heavily on the value of an appropriate ancillary statistic, as well as on the relevance and strength of the instrument in the observed sample. The relevant conditional distribution is still bimodal and we show that the saddlepoint approximation captures this conditional distribution and its bimodality extremely well.

JEL: C01, C10, C30

Keywords: Saddlepoint Approximation, Weak Instruments, Conditional Inference.

1 Introduction

Bimodality of the Maximum Likelihood Estimator (MLE) in the structural equation model has attracted renewed interest recently because of the relations to the weak instrument literature. Originally, Bergstrom (1962) derived the exact distribution of the MLE, and of the least squares estimator, in a simple Keynesian structural model consisting of one behavioural equation and one identity. Phillips and Wickens (1978) showed, and drew attention to the fact that this distribution is bimodal for all parameter values and sample sizes. Nelson and Startz (1990) also find a bimodal distribution for the Two Stage Least Squares (TSLS) estimator by specializing previous distributional results for TSLS to the case where there is one instrument that is only weakly correlated with a single right hand side endogenous variable. They argue that the bimodality is caused by the weakness of the instrument. Maddala and Jeong (1992) challenge this view and argue that bimodality is caused by a high degree of endogeneity. Woglom (2001) unifies both views using results from Marsaglia (1965) to show that the degree of bimodality depends on both the instrument strength and degree of endogeneity. Forchini (2006) extends these results to more than one instrument. Using the exact density he gives necessary conditions and explains bimodality through the interaction between the leading term in the expansion of the density and the noncentrality term. He shows that not only instrument strength and endogeneity play a role, but also the degree of overidentification. Phillips (2006) shows that the weaker the instrument, the stronger the bimodality in a model close to that of Bergstrom. Hillier (2006) considers both TSLS and Limited Information Maximum Likelihood (LIML) estimators and shows that possible bimodality is an artifact of the parameterization. He shows that the degree of overidentification affects OLS and TSLS but that LIML remains unaffected.

In this paper we give an alternative explanation of the bimodality based on the geometric properties of the model. We show that the sample space consists of two connected regions where the estimator maps the data to estimates distinctly to the left and distinctly to the right of the parameter value for which the model degenerates. This does not depend on the parameterization of the model, the weakness of the instrument, or the degree of endogeneity.

The second contribution of the paper is showing that the degree of bimodality depends on the value of an ancillary statistic. This depends on the sample actually being observed rather than on the parameters of the model. For the various reasons sanctioning conditional inference, including information recovery, relevance of inference for the sample actually observed, and the conditionality principle we refer to literature, e.g. Barndorff-Nielsen and Cox (1994, pp.32-35). Previous results by for instance Bergstrom (1962), Phillips and Wickens 1978) etc, related to the *marginal* distributions of the MLE, but we will show that the value of the ancillary has a significant effect on the distribution of the estimator. We also give a graphical explanation why particular samples are inherently less informative about the unknown parameters than others. The ancillary statistics are instructive in identifying the type of samples that are less informative and bimodality poses a bigger problem, and this is an important reason for reporting the conditional distribution given the value of the ancillary observed.

A third and major contribution is to derive the saddlepoint approximation for the conditional density. We show that it captures the bimodality of the distribution perfectly and performs very well even in very small samples. The fact that the saddlepoint approximation can capture the bimodality of distributions of estimators is known, also in econometrics, see e.g. Sowell (2007), but has not received much attention.

2 The Model

We use the model as in Phillips (2006) which is a simple Keynesian model, with one structural behavioural equation and an identity:

$$y_t = \beta x_t + u_t, \tag{1}$$

$$x_t = y_t + \gamma z_t. \tag{2}$$

The model has two endogenous variables, x_t and y_t, an exogenous variable z_t that can be used as an instrument, and a stochastic disturbance u_t assumed to be $IIN(0, \sigma^2)$. The constant γ controls the relevance of the instrument and is assumed know, as in Phillips (2006). In order to guarantee the existence of an equilibrium solution, we assume that the parameter β satisifies $\beta \neq 1$ and the parameter σ^2 needs to be finite. The behavioural equation (1) only makes economic sense if the variance of u_t is finite.We will assume that σ^2 is known and without loss of generality set it to 1.

The unrestricted reduced-form:

$$y_t = \pi_y z_t + v_t, \tag{3}$$

$$x_t = \pi_x z_t + w_t, \tag{4}$$

has 5 parameters: two mean related parameters π_y and π_x and three variance-related parameters, that satisfy a number of restrictions. The mean parameters satisfy: $\pi_y = \gamma\beta/(1-\beta)$, $\pi_x = \gamma/(1-\beta)$ with the parameter of interest $\beta = \pi_y/\pi_x$. The disturbances satisfy: $v_t = w_t = u_t/(1-\beta)$ and the contemporaneous covariance matrix is singular: all four entries equal $1/(1-\beta)^2$, times the variance of u, but σ^2 was set equal to 1.

The restricted reduced-form equations become:

$$y_t = \frac{\beta\gamma}{1-\beta}z_t + \frac{1}{1-\beta}u_t, \tag{5}$$

$$x_t = \frac{\gamma}{1-\beta}z_t + \frac{1}{1-\beta}u_t, \tag{6}$$

An interesting feature of the model is that the unknown parameter, β, enters both the mean and the variance part of the model. It shares this feature with e.g. the distribution of the AR(1) parameter estimate, continuous updating GMM, empirical saddlepoint approximations etc.

It is easy to see, by adding the two equations (5) and (6) that the log-likelihood function equals:

$$\begin{aligned}\mathcal{L}(\beta \mid t_1, t_2) &= t_1\tfrac{1}{4}\gamma(1-\beta^2) - t_2\tfrac{1}{8}(1-\beta)^2 + \\ &\quad -n\tfrac{1}{8}\gamma^2(1+\beta)^2 s_{zz} + n\frac{1}{2}\log((1-\beta)^2) - n\tfrac{1}{2}\log(8\pi),\end{aligned} \tag{7}$$

where $s_{zz} = \frac{1}{n}\sum_{t=1}^{n} z_t^2$ and the statistics t_1 and t_2 equal:

$$t_1 = \sum_{t=1}^{n}(x_t + y_t)z_t \ ; \quad t_2 = \sum_{t=1}^{n}(x_t + y_t)^2. \tag{8}$$

The minimal sufficient statistic $t = (t_1, t_2)'$ is of dimension two and since there is only one unknown parameter the model is a Curved Exponential Model (CEM) of dimensions (2,1) (see e.g. Van Garderen 1997). Such models are embedded in Full Exponential Models (FEMs) where the canonical parameter is a smooth vector function of a lower dimensional parameter. The model has canonical representation

$$p(x;\beta) = \exp\{t'\eta(\beta) - \kappa(\beta) - h(t)\}. \tag{9}$$

where the two dimensional canonical parameter $\eta = (\eta_1, \eta_2)'$ is a smooth vector function of β and equals

$$\eta(\beta) = \frac{1}{8}\begin{pmatrix} 2\gamma(1-\beta^2) \\ -(1-\beta)^2 \end{pmatrix}. \tag{10}$$

The cumulant function $\kappa(\cdot)$ in terms of β equals

$$\kappa(\beta) = \tfrac{1}{8}n\gamma^2(1+\beta)^2 s_{zz} + \tfrac{1}{2}n\log\left((1-\beta)^2\right). \tag{11}$$

and $h(t)$ ensures that the density integrates to 1 but is otherwise irrelevant and could be absorbed in the dominating measure. It is the fact that $\eta(\beta)$ is curved that renders the model a CEM. Obviously, if $\gamma = 0$ the result is a degenerate vertical line along the axis η_2 and the model is actually a one dimensional FEM with complete sufficient statistic t_2.

The cumulant function in terms of β does not enable us to derive all the moments and cumulants we need. We need the function in terms of the canonical parameter η. Using the reduction technique by Van Garderen (1999), however, we can derive the cumulant function of the embedding, 2-dimensional FEM as a function of η to obtain:

$$\kappa(\eta) = -\tfrac{1}{4}n\frac{\eta_1^2}{\eta_2}s_{zz} - \tfrac{1}{2}n\log(-2\eta_2). \tag{12}$$

For values η on the canonical manifold (10) we have $\kappa(\eta(\beta)) = \kappa(\beta)$, but the cumulant function of the FEM is useful for obtaining various terms, such as moments of t, by simple differentiation with respect to the canonical parameter, and calculating the statistical curvature of the model. The expectation, covariance matrix, and 3^{rd}-cumulants are

$$E[t] = \left.\frac{\partial\kappa(\eta)}{\partial\eta}\right|_{\eta=\eta(\beta)} = \begin{pmatrix} \gamma n s_{zz}\frac{(1+\beta)}{(1-\beta)} \\ n\frac{1}{(1-\beta)^2}(4+s_{zz}\gamma^2(1+\beta)^2) \end{pmatrix} \equiv \tau(\beta)\ ,\ \ say$$

$$Var[t] = \left.\frac{\partial^2\kappa(\eta)}{\partial\eta\partial\eta'}\right|_{\eta=\eta(\beta)} = ns_{zz}\frac{4}{(1-\beta)^4}\begin{pmatrix} (1-\beta)^2 & 2\gamma(1-\beta^2) \\ 2\gamma(1-\beta^2) & 4(2/s_{zz}+(1+\beta)^2\gamma^2) \end{pmatrix},$$

$$Cum[t_i,t_j,t_k] = \left.\frac{\partial^3\kappa(\eta)}{\partial\eta_i\partial\eta_j\partial\eta_k}\right|_{\eta=\eta(\beta)} = \begin{cases} & :\{i,j,k\} \\ 0 & :\{1,1,1\}, \\ ns_{zz}\frac{32}{(1-\beta)^4} & :\{1,1,2\},\{1,2,1\},\{2,1,1\} \\ n\gamma s_{zz}\frac{128(\beta+1)}{(1-\beta)^5} & :\{1,2,2\},\{2,1,2\},\{2,2,1\} \\ n\gamma\frac{128(4+3s_{zz}(\beta+1)^2\gamma^2}{(1-\beta)^6} & :\{2,2,2\} \end{cases}$$

etc.

The curved exponential nature of the model leads to a number of interesting insights. First, since the dimension of the minimal sufficient statistic t is larger than the dimension of the parameter, any mapping from t to an estimate $\hat{\beta}$ must lead to loss of information. This information can be recovered by conditioning on an ancillary statistic. Although no exact ancillary is known to exist in our model, approximate ancillaries are available to recover the information approximately. In order to do so we need to append the MLE $\hat{\beta}$, or another estimator, with an ancillary statistic a such that the mapping $t \to (\hat{\beta}, a)$ is one-to-one, and derive the conditional distribution of $\hat{\beta}$ given a. The value of a contains no information on β by itself, since its distribution does not depend on β, but it can contain valuable information about the accuracy or other aspects of the distribution of the estimator. It should be noted that ancillary statistics are not unique and alternative approximate ancillaries have been suggested in the literature. We will consider two ancillary statistics namely Efron and Hinkley's ancillary which in this (2,1)-CEM equals the affine-, or score-ancillary and a signed likelihood ratio statistic based on a test of the CEM against its full exponential embedding as we explain below.

Second, exponential tilting, which is a statistical way of deriving the saddlepoint approximation is closely related to exponential families. See e.g. Reid, (1986), or Barndorff-Nielsen and Cox, (1989, p105) or Durbin (1980) and Barndorff-Nielsen (1980) who specifically derive saddlepoint expansion and conditionality resolutions for CEMs. We will derive the saddlepoint approximations for the MLE using two different ancillary statistics. In both cases the associated saddlepoint approximation captures the bimodality perfectly, but the two distributions are quite different. We confirm a very high degree of accuracy of the approximations by extensive simulations.

Third, since the dimension of the minimal sufficient statistic is two and fixed for all sample sizes, we can represent the model graphically and give a graphical explanation for the bimodality, which we will do next.

3 Geometry and Bimodality

We fist consider the geometry of the model by considering the curved manifolds in the canonical parameter space and in the sample space. The model renders the two-dimensional parameter η in (10) as a function of a single parameter β. This vector function is $C^{\infty}$and defines a smooth manifold $\eta(\beta)$ embedded in the canonical parameter space $\mathcal{N}$. This canonical manifold is illustrated in Figure 1. For negative values of β, the curve starts at the bottom moving up and approaches the origin horizontally from the right as β approaches 1 from below. As β increases above 1 the graph moves to the left.

Since $\eta(\beta)$ is a smooth manifold in $\mathcal{N}$, we can calculate the embedding curvature if we endow it with

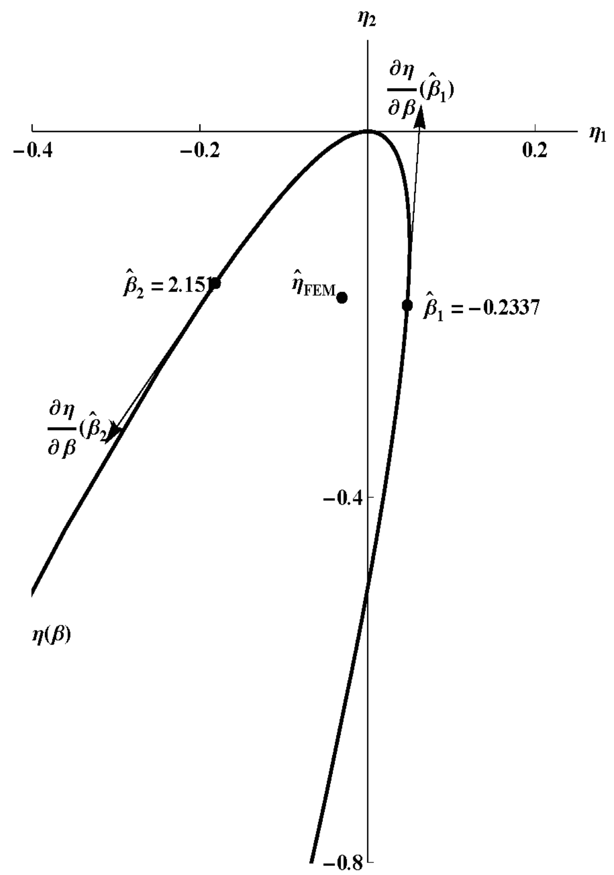

Figure 1: *Canonical parameter space of the embedding FEM,* $\eta(\beta)$ *is the manifold representing the CEM implied by Keynes' model,* $\hat{\eta}_{FEM}$ *is the unrestricted estimate of* $\eta = (\eta_1, \eta_2)'$ *of the embedding FEM for a typical realization with* $t = (-10.5, 69.6)'$, $\hat{\beta}_1 = -0.2337$, $\hat{\beta}_2 = 2.151$ *are the two values that each satisfy the first order conditions,* $\hat{\beta}_1$ *gives the global maximum.*

a metric and affine connection as in e.g. Amari (1985) or by using the original definition of the statistical curvature ϱ by Efron (1975). Straightforward calculations (see Appendix A) give the following result.

Proposition 1. The squared Efron curvature equals, for all $\beta \neq 1$:

$$\varrho_\beta^2 = 2\frac{s_{zz}\gamma^2}{n\left(2 + s_{zz}\gamma^2\right)^3}. \tag{13}$$

The Efron curvature does not depend on β, which is surprising and an interesting particularity of this model. It has various implications, including the fact that the Efron-Hinkley (1978) ancillary is much simpler than usual. The Efron curvature only depends on the strength of the instruments through $s_{zz}\gamma^2$ and on the sample size. It converges to zero at the usual rate for both weak and strong instrument scenarios, unless the average variation in the instruments increases or decreases with n. Surprisingly, in both cases the curvatures vanishes faster.

The expectation of t will also change as β changes and this generates a manifold inside the sample space for t which we denote by $\tau(\beta) = E_\beta[t]$ and is called the expectation manifold. It is shown in Figure 2. The expectation manifold here is smooth everywhere, apart from $\beta = 1$ but that point has been excluded from the model, and furthermore $\tau(\beta)$ has a singularity at $(-\gamma n s_{zz}, \gamma^2 n s_{zz})'$.

The left branch of the graph, where t_1 is negative and t_2 is large, corresponds with β close to, but larger than, 1 and goes down as β increases. As $\beta \to \infty$ the expectation moves to the limiting point $(-\gamma n s_{zz}, \gamma^2 n s_{zz})'$. This is actually the same limit as for $\beta \to -\infty$. When β increases from very negative values the curve, starting at $(-\gamma n s_{zz}, \gamma^2 n s_{zz})'$, first goes down and then moves up, crossing the vertical axis when $\beta = -1$ and then increases beyond all bounds as β approaches 1 from below. The expectation of t_1 is positive for $|\beta| < 1$.

The figures further give a graphical explanation of the MLE. The first order conditions for the MLE in CEMs are immediate from (9)

$$\frac{\partial \mathcal{L}}{\partial \beta} = (t - \tau(\beta))' \frac{\partial \eta}{\partial \beta} = 0, \tag{14}$$

where we have made use of the fact that in full exponential models the expectation of the sufficient statistic t can be obtained by differentiating the cumulant function: $\tau(\beta) = \left.\frac{\partial \kappa}{\partial \eta}\right|_{\eta=\eta(\beta)}$. In order for $\hat{\beta}$ to satisfy the first order conditions, the difference $(t - \tau(\hat{\beta}))$ needs to be orthogonal to $\partial\eta(\hat{\beta})/\partial\beta$ in the standard Euclidean sense. The relevant vectors are shown in the two graphs. The gradient to the canonical manifold evaluated at the MLE, $\partial\eta(\hat{\beta})/\partial\beta$, is shown in Figure 1 and also in Figure 2 together with the difference vector $(t - \tau(\hat{\beta}))$ to which it is orthogonal.

Note that as t changes, also $\hat{\beta}$ changes, unless t changes along $(t - \tau(\hat{\beta}))$. Hence there are straight lines going through $\tau(\hat{\beta})$ and orthogonal to $\partial\eta(\hat{\beta})/\partial\beta$ which all satisfy the first order conditions. Following Efron (1978) we can think of these points as the inverted MLE. From the graph it is clear that every t lies on two different such lines associated with two different $\hat{\beta}$, both satisfying the first order condition. The actual MLE out of these two points is the one that gives the highest likelihood value. See Figure 3 below showing the bimodality in the likeihood function. We will also see algebraically that there are two solutions and we give an analytical expression for this MLE.

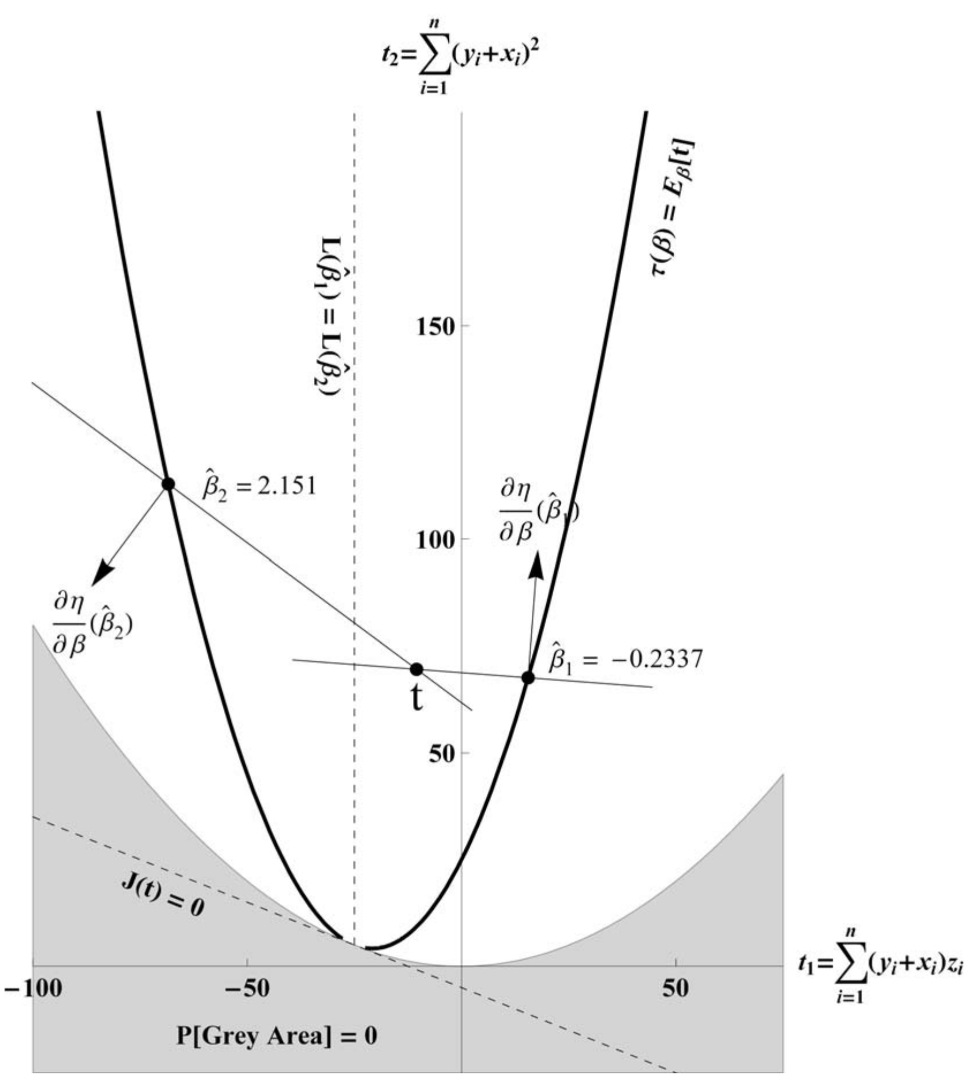

Figure 2: *Sample Space and Expectation Manifold for the minimal sufficient statistic.$\tau(\beta)$: Expectation manifold. t represents an arbitrary sample point and equals $t = (-10.5, 69.6)'$ here. $\hat{\beta}_1$ and $\hat{\beta}_2$ both satisfy the first order conditions for the MLE, but $\hat{\beta}_1 = -0.23$ gives the global maximum.*

There is a set (with probability content 0) where the likelihood value is the same for both β's. This is a vertical line $t_1 = -\gamma \cdot n \cdot s_{zz}$, which is illustrated in Figure 2 as a dashed line. To the right of this line, observations t are mapped to estimates $\hat{\beta}$ smaller than 1, and t's to the left give rise to ML estimates larger than 1. The ML estimate will never equal 1 as it would require t_2 to be larger than any finite bound ($\hat{\beta} \neq 1$ a.s. since the likelihood goes to $-\infty$ when $\beta \to 1$, either from above or below, for any given sample). We can combine the geometrical insights with the Gaussian distribution of $t_1 = \sum (y_i + x_i) z_i$

to investigate the importance of the two modes. The fact that

$$t_1 \sim N\left(\gamma n s_{zz} \frac{1+\beta}{1-\beta}, n s_{zz} \frac{4}{(1-\beta)^2}\right), \tag{15}$$

means, first of all, that for a given value of $\beta < 1$ it is always possible that t_1 is to the right of the critical line. This is analogues to the result by Phillips (2006) that the distribution is always bimodal. Second, using (15) it is simple to compute the probability of $\hat{\beta}$ being on the wrong side of 1:

Proposition 2. Wrong side of 1 probabilities equal:

$$P[\hat{\beta} > 1 \mid \beta < 1] = \Phi\left(-4\gamma\sqrt{n\ s_{zz}}\right) \tag{16}$$

$$P[\hat{\beta} < 1 \mid \beta > 1] = 1 - \Phi\left(4\gamma\sqrt{n\ s_{zz}}\right) \tag{17}$$

The probabilities of being on the wrong side of 1 only depend on the parameter γ that controls the strength of the instrument and on the variation in the instruments as measured by $\sum z_i^2$. The term $\gamma^2 n s_{zz}$ is the noncentrality parameter (Phillips, 2006 p.950). Further note that with a typical local to zero sequence $\gamma_n = d/\sqrt{n}$, as is regularly used in the weak instrument literature (see e.g. Phillips, 2006 p.955 for some references and discussion), this probability depends on the average variation in the instruments as given by s_{zz} and will not decrease as the sample size n increases. It is also interesting that the probability of being on the wrong side of 1 is independent of β. This holds even for arbitrary large values of β, since the result is exact.

We will be interested in the conditional distribution of $\widehat{\beta}$ given an ancillary a, and (16) is only a marginal probability. It is clear however, that when this probability is substantial that there must be values of a for which the conditional probability is also substantial since $P[\hat{\beta} > 1] = \int P[\hat{\beta} > 1|a]dF(a)$. We will show that the probability of being on the wrong side of 1 depends heavily on the value of the ancillary a actually observed. This obviously holds only for a suitable choice of ancillary statistic. There is no unique ancillary statistic, however, and it is not even known whether an exact ancillary exists in our model. Various general proposals have been put forward for statistics that are approximately ancillary and we should like to choose that statistic that is most informative about the degree of bimodality of the MLE.

3.1 Global versus Local Curvature

Following Efron (1978) and Amari (1982) we can think of a foliation of the sample space where $\hat{\beta}$ serves as the coordinate to index the folio, in this case the inverted MLE line, and the ancillary a as the other coordinate indicating the position of t on the folio. From a statistical point of view we want this coordinate system to be orthogonal in the sense that the distribution of a should not depend on β. If the distribution of a depended on β, then we could use a to increase the efficiency of the estimator. If the distribution of a does not depend on β, then a on its own cannot tell us anything about β, but together with $\hat{\beta}$, could tell us something about the accuracy of the estimator. This is the traditional argument. We generalize this to the notion where a can tell us something about the extend of bimodality. In this respect the graphs are informative again.

The dashed vertical line in Figure 3 is defined as those points t such that the two local maxima have the same likelihood value $\mathcal{L}\left(\hat{\beta}_{(1)} \mid t\right) = \mathcal{L}\left(\hat{\beta}_{(2)} \mid t\right)$. For these values of t we cannot decide between the two values of β on the basis of the likelihood. Both seem equally plausible. Values of t close to this vertical line will have two local maxima with similar likeihood values and deciding between them is difficult. In contrast, values t_1 well to the left of the expectation manifold are far more plausible for a positive β, and for large positive t_1 we can be much more confident that it is a realization from the model with a negative β. The ancillary should reflect this and show that for t close to the vertical dashed line, bimodality is much more important.

The bimodality of the distribution derives from the global geometry of the model. Given the way in which $\eta(\beta)$ curves inside $\mathcal{N}$ there will be $t's$ that are an equal likelihood distance away from $\tau(\beta^{(1)})$ and $\tau(\beta^{(2)})$ regardless of the local curvature at $\beta^{(1)}$ and $\beta^{(2)}$.

The local curvature can also lead to an indeterminacy for the MLE, namely when the observed information is 0, and the likelihood is locally flat. The observed information can be written as

$$\begin{aligned} J_\beta &= -\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = \frac{\partial \eta'}{\partial \beta}\frac{\partial^2 \kappa}{\partial \eta \partial \eta'}\frac{\partial \eta}{\partial \beta} - (t-\tau(\beta))'\frac{\partial^2 \eta}{\partial \beta^2} \\ &= i_\beta - (t-\tau(\beta))'\frac{\partial^2 \eta}{\partial \beta^2}, \end{aligned}$$

where i_β is the expected Fisher information (this follows immediately from the first line since the second term has expectation 0). Holding $\hat{\beta}$ constant, we can move t along the inverted MLE line so far that $|J_\beta| = 0$. For these points the MLE is not uniquely defined.

Hence we see that the local and global curvature lead to two different kinds of non-uniqueness. One due to the global geometry of the model with a likelihood that is not at all flat, and a second one due to the local curvature and a likelihood that is locally flat.

4 Saddlepoint Approximation

In order to show the effect and importance of conditioning, we need the conditional density of $\widehat{\beta}$ given an approximate ancillary statistic a. Bergstrom (1962) and Phillips and Wickens (1978) find the exact *marginal* distribution of the MLE. Finding the *conditional* distribution involves a number of additional complications. First, the approximate ancillary statistic a needs to be chosen and subsequently the distribution needs to be derived. For sensible choices of a that allow us to recover lost information, not only in the traditional sense but also to tell us something about the extend of the bimodality, we could not derive exact conditional distributions. Instead, we use approximations. The Edgeworth approximation is not appropriate because it will lead to a distribution that is essentially uni-modal (multimodality occurs due to unfortunate oscillations). Saddlepoint approximations can accommodate bimodality, as is known from the statistical literature, and we will show below that it captures the conditional bimodality in the current problem extremely well.

The saddlepoint approximation was pioneered by Daniels (1954) based on the steepest decent technique in asymptotic analysis. A statistical derivation via exponential tilting can be found in Barndorff-Nielsen & Cox (1979) and Reid (1988) gives a very clear review of the developments in the statistical literature including work by Durbin, Sargan, and Phillips. Phillips (1978) approximated the density of the MLE in the first order autoregression and showed that it outperformed the Edgeworth approximation, although in certain cases with moderate sample sizes and large parameter values performance breaks down. Durbin (1980a) derived a tilted distribution for exponential families and allowed for non-i.i.d. data, which he also applied to the AR(1) model Durbin (1980b). Related work also includes Lieberman (1994 a,b,c) and Larsson (1999).

In Econometrics the technique was pioneered by Phillips and Holly (1979) who derived the approximate density of the k-class estimator in a simultaneous equations system with two endogenous variables. The saddlepoint approximation was shown to outperform the Edgeworth expansion, especially in the tails where it is well know that the Edgeworth expansion can behave very poorly. Spady (1991) derives saddlepoint approximations for systems of estimating equations and applies this to the least absolute deviation estimator in a regression context. He shows that the approximation captures the multimodality of the estimator (four modes in that case) and shows that his Ψ-transformation measure is the key component in the explanation of multimodality. It should be noted however, that not only is the criterion function not smooth, but the disturbance that he uses is itself a mixture of normals and bimodal. In our model the disturbances are uni-modal. Sowell (2007) provides a saddlepoint approximation for GMM estimators and the statistics for testing the overidentification restrictions. The true distribution of the GMM estimator can have multiple modes and he shows that the approximation captures this important feature. For small samples he further shows in a Monte-Carlo study that it often outperforms the bootstrap. See also Sowell (2009).

We take the simple Keynesian model (1) and derive the saddlepoint approximation, conditional on an approximate ancillary statistic. Since the model is a CEM the saddlepoint approximation has a simple likelihood formulation, as was originally noted by Daniels (1958). See also Durbin (1980) and Barndorff-Nielsen (1980) who, inspired by results of Fisher, developed his p^* formula for the conditional distribution of the MLE $\hat{\beta}$ given an ancillary a. The p^* formula is:

$$p(\widehat{\beta};\beta \mid a) \approx c(a)\left|j\left(\hat{\beta};a\right)\right|^{1/2}\exp[l(\beta)-l(\hat{\beta})], \tag{18}$$

where $l(\beta)=l(\beta;y)$ is the log likelihood function for β given observations y, and $\widehat{j}=-\left.\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right|_{\beta=\hat{\beta}}$ is the observed information matrix, which can be expressed either as a function of the sufficient statistic $t=t(y)$ or in terms of $(\hat{\beta},a)$ if they are in one-to-one correspondence. The term $c(a)$ is a normalizing constant such that the right hand side of (18) integrates to 1. A closed form for $c(a)$ is generally not available but can be computed numerically.

In order to make the formula operational, we need the mapping between the minimal sufficient statistc $t(y)$ and the pair $\left(\hat{\beta},a\right)$, which in the current model is finite dimensional and can be chosen 1-1 since the model is a curved exponential model. It is straightforward to find an explicit expression for $\hat{\beta}$ in terms of t, and also for the various possible ancillaries a below. No analytical inverses for this mapping are available however. We cannot explicitly write t in terms of $\hat{\beta}$ and a. For *given* values of $\hat{\beta}$ and a, however, we can easily find t by solving a set of two linear equations and this will be exploited in the calculation of the saddlepoint approximations.

4.1 The MLE and Ancillaries

The first order conditions in (14) give:

$$\mathcal{L}'(\widehat{\beta}) = -\frac{n}{(1-\widehat{\beta})} - \frac{1}{2}t_1\widehat{\beta}\gamma + \frac{1}{4}t_2(1-\widehat{\beta}) - \frac{1}{4}n\ s_{zz}(1+\widehat{\beta})\gamma^2 = 0. \tag{19}$$

This is a quadratic equation in $\hat{\beta}$ and the likelihood function has two stationary points. Generally, one of these solutions gives the global maximum, the other is a local maximum. There are exceptional values for t discussed below where the MLE is not unique, but they have probability 0.[1] The global maximum cannot be determined using the second order derivative

$$\mathcal{L}''(\widehat{\beta}) = -\frac{n}{(1-\widehat{\beta})^2} - \frac{2t_1\gamma + t_2 + n\ s_{zz}\gamma^2}{4} \tag{20}$$

since it is negative for both solutions.

A likelihood function for a typical realization of the process is depicted in Figure 3 and it shows the local and the global maxima. The realization is the same one that lead to t in Figure 2. This explains the bimodal likelihood, since t is almost the same likelihood distance away from $\tau\left(\hat{\beta}_1\right)$ as from $\tau\left(\hat{\beta}_2\right)$. In between $\hat{\beta}_1$ and $\hat{\beta}_2$ the (log-) likelihood (7) will go to minus infinity as $\hat{\beta}$ approaches 1. The model is not defined for $\beta = 1$ however, and explicitly excludes it.

The root that maximizes the likelihood globally is found by direct comparison of the likelihood values for the two roots, resulting in the explicit solution:

Proposition 3. The MLE in the simple Keynesian model (1) and (2) is given by:

$$\widehat{\beta} = \frac{t_1\gamma + t_2 + sign(t_1 + ns_{zz}\gamma)\sqrt{4nt_2 + 8nt_1\gamma + 4n^2\gamma^2 s_{zz} + t_1^2\gamma^2 + 2ns_{zz}t_1\gamma^3 + n^2 s_{zz}^2\gamma^4}}{2t_1\gamma + t_2 + ns_{zz}\gamma^2},$$

$$t_1 + ns_{zz}\gamma \neq 0. \tag{21}$$

$$\hat{\beta} = 1 \pm 2\sqrt{n/\left(t_2 - \gamma^2 n\ s_{zz}\right)} \tag{22}$$

For $t_1 + ns_{zz}\gamma = 0$, the roots to the first order conditions are

$$\hat{\beta} = 1 \pm 2\sqrt{n/\left(t_2 - \gamma^2 n\ s_{zz}\right)}$$

[1] The two roots $\hat{\beta}_1$ and $\hat{\beta}_2$ have $\mathcal{L}_{\max}(\hat{\beta}_1) = \mathcal{L}_{\max}(\hat{\beta}_2)$ when $t_1 + ns_{zz}\gamma = 0$,see proof case i) of Lemma 1, Appendix B, but this is a line the sample space T and therefore dense in T with probability zero.

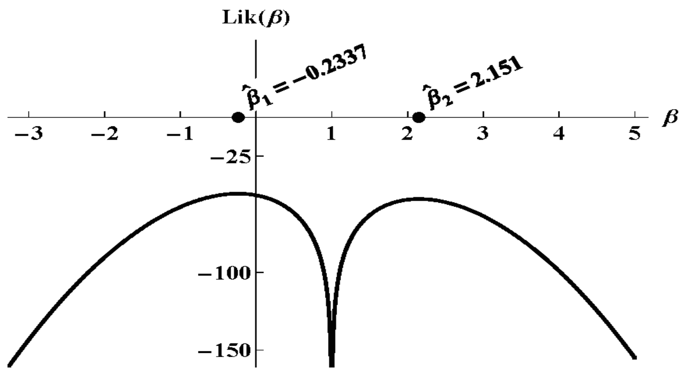

Figure 3: *Typical Log-likelihood given a sample of size $n = 10$ from the model with $t = (-10.5, 69.6)'$ as before. The two local maxima are indicated, but only $\hat{\beta}_1$ is associated with the global maximum for this sample.*

and are distinct, but give the same value of the likelihood.

We call the line $t_1 + ns_{zz}\gamma = 0$ the *critical set* based on the global curvature of the model and is illustrated in Figure 2 by the dashed vertical line. The MLE is not uniquely defined on this line. This critical set itself has zero probability since t_1 is continuously distributed, but we can observe values close to this line. In a neighbourhood of this critical set, we can have that a small change in the normally distributed t_1 can lead to a discrete and substantial change in the MLE $\hat{\beta}$. If t_1 changes from the right to the left of the line, then the MLE will jump from a value larger than 1 to a value smaller than 1.

There is a second critical set where MLE is not uniquely defined, consisting of those values of t for which the observed information $j(\widehat{\beta})$ is singular. The likelihood is locally flat and as a consequence there is no unique $\hat{\beta}$. The values of t in (20) must satisfy the first order conditions $(t - \tau(\beta))' \frac{\partial \eta}{\partial \beta}\Big|_{\beta = \hat{\beta}}$ and the t's therefore lie on the inverted MLE line. Points on this line can be indexed by a giving an invertible relation between t and $(\hat{\beta}, a)$ and we can express $t = \left(t_1(\widehat{\beta}, a), t_2(\widehat{\beta}, a)\right)'$ and the observed information, as a function of $(\widehat{\beta}, a)$ is:

$$j(\widehat{\beta}, a) = -\mathcal{L}''(\widehat{\beta}; \widehat{\beta}, a) = \frac{n}{(1 - \widehat{\beta})^2} + \frac{2\gamma \cdot t_1(\widehat{\beta}, a) + t_2(\widehat{\beta}, a) + s_{zz}\gamma^2}{4}, \tag{23}$$

There are various choices for the ancillary statistic a that could be employed in the formulas above and this we consider next.

4.1.1 Ancillary Statistics

In order to find the saddlepoint approximation we need a one-to-one transformation between the set of minimal statistics and the set of parameters and ancillary statistics. In our CEM (2,1) model that means that we need to find only one ancillary statistic to condition the MLE $\widehat{\beta}$ on. For a given model it is not obvious whether or not an exact ancillary statistic is available and even when this is the case it is still possible to have two (or more) ancillaries that yield different conditional distributions. Although in general there are no techniques to prove that exact ancillary statistics exist for a given model, it is still possible to construct approximate ancillary statistics. Sometimes these approximate ancillary statistics turn out to be exact, depending on the properties of the model. For our model we use three approximate ancillary statistics: the Efron-Hinkley (1978) ancillary, the Affine- or score-ancillary of Barndorff-Nielsen (1980) and a signed-LR ancillary.

The Jacobian term in the saddlepoint approximation (18) is written as a function of $(\widehat{\beta}, a)$, but the second derivative of the log-likelihood depends on β and (t_1, t_2). After substitution of the MLE, the observed information will depend on $\hat{\beta}$ and t and not yet explicitly on a. We can, however, with the choices of a below which render the relations invertible, write t as a function of $\hat{\beta}$ and $a : t(\widehat{\beta}, a)$. Using this functional relation for t, the saddlepoint approximation (18) for our model becomes:

$$\begin{aligned} f(\widehat{\beta}; \beta \mid a) \quad &\approx \quad c(a)[\frac{n}{(1-\widehat{\beta})^2} + \frac{2\gamma \cdot t_1(\widehat{\beta}, a) + t_2(\widehat{\beta}, a) + s_{zz}\gamma^2}{4}]^{1/2}[\frac{(1-\beta)^2}{(1-\widehat{\beta})^2}]^{n/2} \times \\ &\times \exp[\tfrac{1}{8}(\widehat{\beta}-\beta)(\gamma^2 s_{zz}(\widehat{\beta}+\beta+2) + (\widehat{\beta}+\beta-2) \cdot t_2(\widehat{\beta}, a) + 2\gamma(\widehat{\beta}+\beta) \cdot t_1(\widehat{\beta}, a)\ ]. \end{aligned} \tag{24}$$

In practice we need to determine the values $t(\widehat{\beta}, a)$ numerically, which is straightforward since for fixed $\hat{\beta}$ and a we have two linear equations in t_1 and t_2. This follows from the fact that all loglikelihood derivatives for given $\hat{\beta}$ are linear in $(t - \hat{\tau})$.

4.1.2 Efron-Hinkley Ancillary

The Efron-Hinkley (1978) statistic was derived based on the (local) geometry of the model as a way to improve the reporting of the accuracy of the MLE. They showed that the observed information was a more accurate measure of the conditional variance than the expected information.

$$\begin{aligned} a_{EH} &= (1-\hat{\jmath}/\hat{\imath})/\varrho \\ &= (t-\tau(\beta))'\frac{\partial^2\eta}{\partial\beta^2}/(\varrho i), \end{aligned}$$

with

$$i = n\frac{2+s_{zz}\gamma^2}{(1-\beta)^2}$$

and the Efron curvature ϱ given in Proposition 1 and constant over β, and depending only on γ, s_{zz} and n. This means that the Efron-Hinkley statistic is essentially the ratio of the observed information over the expected information.The Efron-Hinkley ancillary does not take the global structure of the model into account, but does change sign when crossing the expectation manifold and hence will distinguish points close to the critical set $t_1 = -\gamma\ n\ s_{zz}$ from those that are close to $\tau(\beta)$ the expectation manifold.

4.1.3 Affine or Score Ancillary

Barndorff-Nielsen (1980) proposed an approximate ancillary that, for a fixed value of $\hat{\beta}$, is an affine function of the minimal sufficient statistic. The affine ancillary a in curved exponential models, using our notation, is defined as:

$$a = \widehat{A}'(t-\widehat{\tau}), \tag{25}$$

where $\widehat{\tau} = \tau\left(\hat{\beta}\right)$ is the expectation of t evaluated at the MLE $\widehat{\beta}$, and $A = A(\beta)$ is a $k \times (k-d)$ normalizing matrix such that $A(\beta)'(t-\tau(\beta))$ has identity covariance matrix and mean 0, resulting in a statistic that is, at least approximately, ancillary in mean and variance. In the current model $(k-d) = 1$ and the affine ancillary and Efron and Hinkley's ancillary coincide.

4.1.4 Signed Likelihood Ratio Ancillary

An alternative interpretation of the affine ancillary is as a score test of the CEM against the embedding FEM. Similarly one can use a likelihood ratio statistic for testing the CEM against the FEM. This statistic will be approximately Chi-squared distributed with $(k-d)$ degrees of freedom. An obvious problem is however that such a statistic does not distinguish points close to the critical set $t_1 = -\gamma\ n\ s_{zz}$, and points on the other side (outside) of the expectation manifold that are an equal likelihood distance away from the restricted likelihood. In order to overcome this we attach a negative sign to observations on the inside of the expectation manifold and further take the square root of the LR statistic. Such signed LR statistics have been succesfully used in the literature before and often result in a distribution that is close to Gaussianity. Our sign, however, is based on the insight developed in Section 2 and is based on the global geometry of the model.

In order to construct the LR statistic we need to estimate the FEM. This is straightforward using the cumulant function for the FEM derived earlier, and estimate the unrestricted canonical parameters η in (10) using the first order conditions:

$$t - \partial\kappa(\eta)/\partial\eta = t - \tau(\eta) = 0,$$

resulting in the canonical ML estimates:

$$\hat{\eta}_1 = \frac{n\ t_1}{n\ s_{zz}\ t_2 - t_1^2}; \qquad \hat{\eta}_2 = -\frac{1}{2}\frac{n^2\ s_{zz}}{n\ s_{zz}\ t_2 - t_1^2}, \tag{26}$$

and for the maximized likelihood:

$$\mathcal{L}_{\max} = \mathcal{L}(\hat{\eta}) = -\frac{1}{2}n\left(1 + Log\,(2\pi) + Log\left(\frac{n\ s_{zz}\ t_2 - t_1^2}{n^2\ s_{zz}}\right)\right),$$

from which the signed-LR is given by:

$$SLR = sign(t) * \sqrt{2(\mathcal{L}(\hat{\eta}) - \mathcal{L}\left(\hat{\beta}\right))}$$

where $sign(t) = 1$ if observation t lies inside the expectation manifold and $sign\,(t) = -1$ if it lies on the outside (further away from the *critical set,* dashed line in Figure 3).

The saddlepoint approximation formula (18) has the same form for this ancillary, but the relation between $(\widehat{\beta}, a)$ and (t_1, t_2) is different and the normalizing constant $c\,(a)$ is also different.

5 Results

In this section we report the saddlepoint approximations and the results of a simulation study. The study has three goals. The first goal is to study the quality of the saddlepoint approximation. In particular how well it captures the bimodality in the estimation of β and how it depends on the value of a given ancillary, even when these values are large The second goal is to examine the ancillarity properties of the conditioning statistics used . Since asymptotic theory only establishes approximate ancillarity of the chosen statistics, there is no guarantee that in small samples, or with weak instruments, the distribution is still almost invariant to changes in the parameter β. We want to investigate whether the statistics are approximately ancillary in small samples. Finally, since it is trivial to find exact ancillaries that have nothing to say about the distribution of the MLE, the third goal is to investigate which ancillary is most useful or informative about the distribution of the MLE.

Below we also investigate and confirm the ancillarity properties of the statistics used in the conditional saddlepoint approximations. This avoids discussions about conditioning on statistics that still contain information about the parameters by themselves and are not ancillary.

In the simulations reported here we consider three different scenarios with alternative values for the sample size and for the γ parameter that controls the strength of the instrument. The average variation in the instrument s_{zz} is kept constant. The values considered are not extremely small, yet highlight the very substantial differences with asymptotic theory.

Scenario	n	γ	s_{zz}
I	25	0.02	1
II	250	0.02	1
III	25	0.2	1

Table 1: *Scenarios used in the simulation study*

The values for β were restricted to the interval $\beta \in [0, 1)$ since it represents the propensity to consume. We chose four different values $\beta = \{0, 0.6, 0.95, 0.99\}$. All simulations were based on 100.000 replications.

5.1 Saddlepoint Approximation | Efron-Hinkley Ancillary

The saddlepoint approximation for the conditional density given the Efron-Hinkley ancillary were examined for the three different scenarios and the different values of β. Figure 4 shows the conditional density for the Efron-Hinkley ancillary in Scenario I with $n = 25$, $\gamma = 0.02$ and $\beta = 0.6$. Results for other values of β and different scenarios are given in Appendix D, but the qualitative conclusions are the same.

A number of observations are immediate. The distribution is extremely bimodal. There is one mode at the true value for $\beta = 0.6$ and one at 1.4. In the other cases we also found one mode in a neigbourhood of the true value β and another mode almost symmetrically on the other side of 1, in a neighbourhood of $1 + (1 - \beta)$. The modes are perfectly separated by a region around $\widehat{\beta} = 1$ where the density is zero. This is explained by the fact that the model is not properly defined for $\beta = 1$ and the likelihood for any sample will go to $-\infty$ as $\hat{\beta}$ goes to 1.

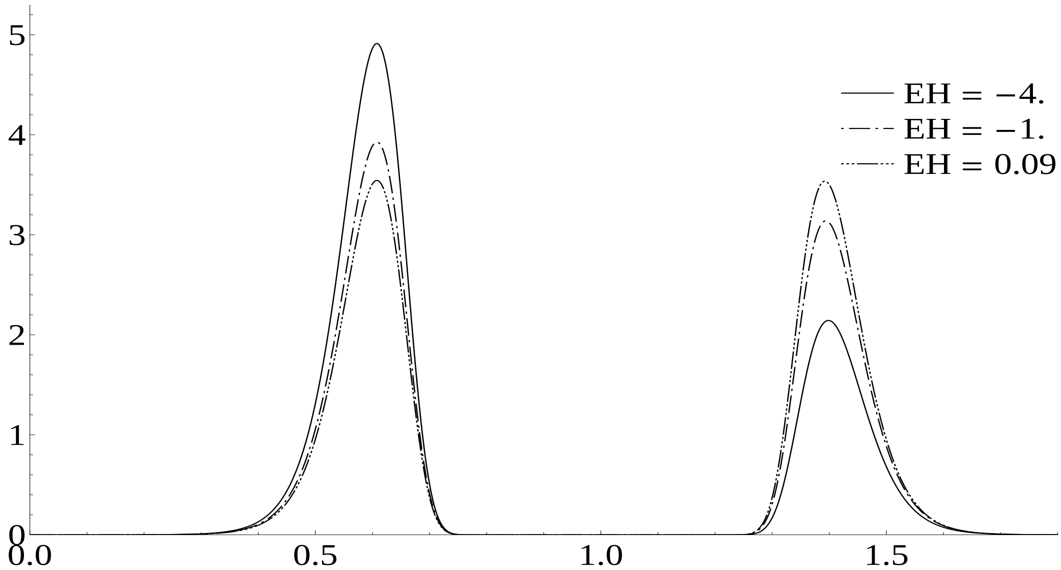

Figure 4: *Saddlepoint Approximation given Efron-Hinkley ancillary,* $\beta = 0.6$, $n = 25, \gamma = 0.02$, $s_{ZZ} = 1$

The conditional density is really very different for different values of the ancillary. When $a_{Eh} = 0.09 \approx a_{\max}$, which corresponds to observations t falling inside the expectation manifold close to the critical set $t_1 = -\gamma\ n\ s_{zz}$, the bimodality is more extreme than for the a_{EH} values -1 and -4 associated with observations t on the outside. The second mode around 1.4 increases markedly when a_{EH} increases. If we express the degree of bimodality as the probability of $\hat{\beta}$ being larger than 1, then $\text{Prob}\left[\hat{\beta} > 1 | a_{EH} = 0.09\right] = 0.499$ and $\text{Prob}\left[\hat{\beta} > 1 | a_{EH} = -4\right] = 0.3$. Other values for the probability of being on the wrong side of 1 are given in Table 2. The difference in bimodality is in line with the geometrical insight we developed earlier. The big variations in the conditional distributions also shows that a_{EH} contains valuable information on the distribution of the MLE, despite having a distribution that effectively does not depend on β and therefore having no information on β itself. This also means that the conditional distribution can be very different from the marginal distribution which averages over all possible values of the ancillary and ignores the fact that certain configurations of the sample are far more informative than others and observations close the critical set lead to far more bimodality than negative values of a_{EH}. When the true β is larger than 1, the picture is the other way round and for large a_{EH} the probability of observing a value much smaller than 1 is close to 0.5.

Conditional: $\mathrm{P}[\hat{\beta} > 1 \vert a_{EH}]$	Scen.I	Scen.II	Scen.III
$a_{EH} = -4$	0.309	0.062	0.000
$a_{EH} = -1$	0.445	0.303	0.022
$a_{EH} = a_{\max}$	0.499	0.497	0.495
Marginal: $\mathrm{P}[\hat{\beta} > 1]$	0.460	0.376	0.159

Table 2: *Probability of $\hat{\beta}$ on the wrong side of 1 based on the conditional Saddlepoint approximation given a_{EH} when true $\beta = 0.6$, and the marginal probability based on the exact distribution.*

Table 2 was also calculated for other values of β, but it hardly varies with β. Those results are therefore not reported. We already showed in equation (16) that the marginal probability of observing a $\hat{\beta} > 0$ was independent of β, and in the previous section we showed that a_{EH} was almost ancillary. That does not imply however, that the conditional probabilities are also invariant to β.

If we move to Scenario II by increasing the sample size to $n = 250$, then bimodality decreases. The probability of obtaining a $\hat{\beta} > 1$ is reduced, both conditionally and unconditionally (from 0.46 to 0.38) which can also be seen in the graph in Appendix D. For moderate values of a_{EH} the probability reduces significantly, but for very large ancillary values, the conditional probability is still 0.497. When we move to Scenario III where the instruments are strong and the sample size is small, the bimodality reduces even further, but for very large a_{EH} the probability is still 0.495. One would not often observe these very large values and the conditional distributions will not deviate much from the marginal in most cases. If we do have an exceptionally large value for a_{EH} , then the difference is all the more spectacular. The crucial point is that we can simply calculate and observe a_{EH} from the data and we know which distribution to use. We do not have to guess or average over all possible values of a_{EH} that might have occurred.

5.2 Saddlepoint Approximation | Signed Likelihood Ratio Ancillary

The effect of conditioning on the signed-LR statistic of testing the CEM (know to be true) versus the embedding FEM density gives a similar picture as when conditioning on the Efron-Hinkley ancillary with some minor differences. Of course the maximum values of the signed-LR statistic are different as is its distribution, but the conditional probabilities of being on on the wrong side of 1, given the signed-LR show the same patern. The conditional densities also look very similar and hence we show here the conditional density of $\hat{\beta}$ given similar values of signed-LR as for the Efron-Hinkley, but now for Scenario III with a small sample size but informative instruments. We see that the conditional distribution is changing far more with the ancillary than for the other two scenarios. For the smallest value of a_{SLR} the density is essentially unimodal and for the largest value of a_{SLR} it is almost symmetric around 1. This again means that the value of the ancillary statistic whould not be ignored when making inference. Confidence intervals with a large value of the ancillary and a standard coverage rate of 95% say and based purely on statistical properties, should consist of two disjoint regions, whereas for small values it should be a simply connected interval.

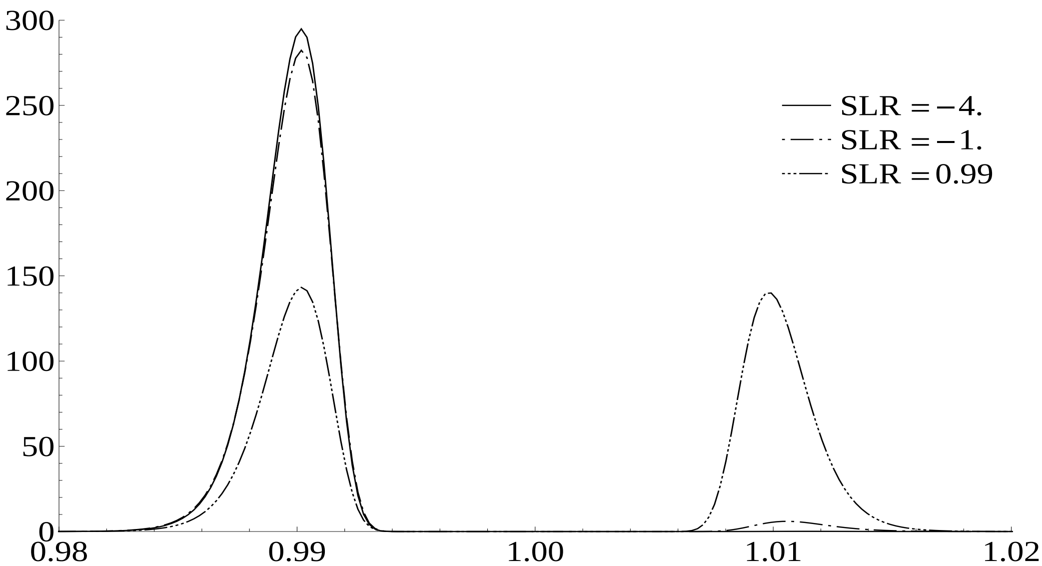

Figure 5: *Saddlepoint approximation given SLR ancillary* , $\beta = 0.99$, $n = 25, \gamma = 0.2$, $s_{ZZ} = 1$

We have chosen $\beta = 0.99$ to show that the same strict separation between the two sides of the density around β and $1+(1-\beta)$ exists, remarkedly so regardles of how close or how far β is from 1, including the extreme values 0 and 0.99. This is the case for both the signed LR and the Efron-Hinkley statistic. The conditional probability $P[\widehat{\beta} > 1 \mid a_{SLR} = a]$ depends on the value of a_{SLR}. Integrating out appropriately across all possible values of a_{SLR} gives the marginal density and probability that $\hat{\beta} > 1$. We have already shown that the marginal probability $P[\widehat{\beta} > 1]$ does not depend on β and the marginal distribution of a_{SLR} is also virtually free of β, yet the conditional distribution of $\hat{\beta}$ given a_{SLR} (or a_{EH}) does vary remarkedly with β as the locations of the modes around β and $2-\beta$ clearly show.

Conditional: P[$\hat{\beta} > 1 \vert a_{SLR}$]	Scen.I	Scen.II	Scen.III
a_{SLR} =-4	0.333	0.064	0.000
-1	0.445	0.303	0.023
$a_{\max}$	0.499	0.498	0.494
Marginal:P[$\hat{\beta} > 1$]	0.460	0.376	0.159

Table 3: *Probability of $\hat{\beta}_{MLE} > 1$ based on the conditional Saddlepoint approximation given a_{SLR} for Scenarios I-III and $\beta = 0.99$*

Finally, relating to discussions here and in the literature about the cause of the bimodality, note that because the conditional distributions are bimodal the marginal distribution must also be bimodal and vice versa.

5.3 Quality of the Saddlepoint Approximation

The quality of the saddlepoint approximations is very good for this simple, but inferentially difficult Keynesian model. It captures the bimodality and its dependence on the ancillaries perfectly. The accuracy of the approximation is not easy to assess. There is no closed form expression for the conditional density $f(\widehat{\beta}; \beta \mid a)$ and we resort to simulation based methods. These are hindered by the complication that we compare conditional distributions for different values of a, including exceptional ones. The number of relevant observations can be small, especially in the tails of the distributions of the MLE and for extreme values of the ancillary. Kernel based densitiy estimates therefore proved troublesome and could not be used to illustrate the quality of the approximation (the Kernel estimates proving less reliable than the Saddlepoint approximation). We therefore use a one-sample Kolmogorov-Smirnov test, since there is no difficulty in calculating the saddlepoint approximation, even when the density is very small, which we can use as null hypothesis. The results are given in Table 4. In none of the cases could the saddlepoint approximation be rejected at the 5% level and in most cases it was far below the critical value of 1.36

KS stats		Effron-Hinkley Ancillary					Signed LR Ancillary			
Scenario	β	-4	-2	0	$a_{\max}$		-4	-2	0	$a_{\max}$
I	0	0.60	0.75	0.57	0.83		0.55	0.61	1.11	1.02
$n = 25$,	0.6	1.03	0.72	0.72	0.72		0.73	0.88	0.91	0.90
$s_{zz} = 1$,	0.95	0.86	0.60	1.02	0.40		1.21	0.90	0.43	0.88
$\gamma = 0.02$	0.99	0.67	0.76	0.58	0.85		1.15	0.61	1.08	1.05
II	0	1.11	0.89	0.78	0.85		1.05	0.46	0.73	0.96
$n = 250$,	0.6	0.49	1.05	0.71	0.83		0.57	0.85	0.97	0.93
$s_{zz} = 1$,	0.95	0.66	0.76	1.10	0.82		0.97	1.01	1.06	0.97
$\gamma = 0.02$	0.99	0.53	0.89	1.04	0.99		1.20	0.95	0.75	0.51
III	0	0.75	0.89	0.73	0.44		1.05	1.13	0.66	0.94
$n = 25$,	0.6	1.11	1.20	0.74	0.84		1.15	1.05	0.69	0.92
$s_{zz} = 1$,	0.95	0.98	0.81	1.03	0.75		0.90	0.64	0.93	0.67
$\gamma = 0.2$	0.99	0.81	1.29	1.08	0.72		1.01	0.99	1.17	1.06

Table 4: *Kolmogorov-Smirnov one-sample statistics for testing the null hypothesis that the Saddlepoint density is the true density.*

We have chosen four different values of the ancillary statistics, namely $\{-4, -2, 0, a_{\max}\}$, where $a_{\max}$ is a value just below the theoretical maximum for every ancillary and each of the three scenarios. For the empirical distribution we had to use observations with values of the ancillary a in a neighbourhood of the stated values since a is continuously distributed. A small bandwidth of 10^{-3} was chosen and we ran the simulation until the required number of observations was achieved.

5.4 Ancillarity Properties

Exact ancillarity can be proved only by analytical methods and no results are available or could be obtained for our model.[2] Simulations can be used, however, to gather evidence in finite samples for both the Efron-Hinkley and the signed-LR statistics. Figure 6 shows the density of the Efron-Hinkley statisic based on these simulations for Scenario II and four different values of β, and similarly for the signed-LR in Scenario II. Ancillarity holds if the densities do not change when the parameter β changes. Figure 6 shows for both the Efron-Hinkley and the signed-LR statistic that the distribution hardly changes as we vary β from 0 to 0.99. For the empirical CDFs given in the appendix it is even harder to see the difference.

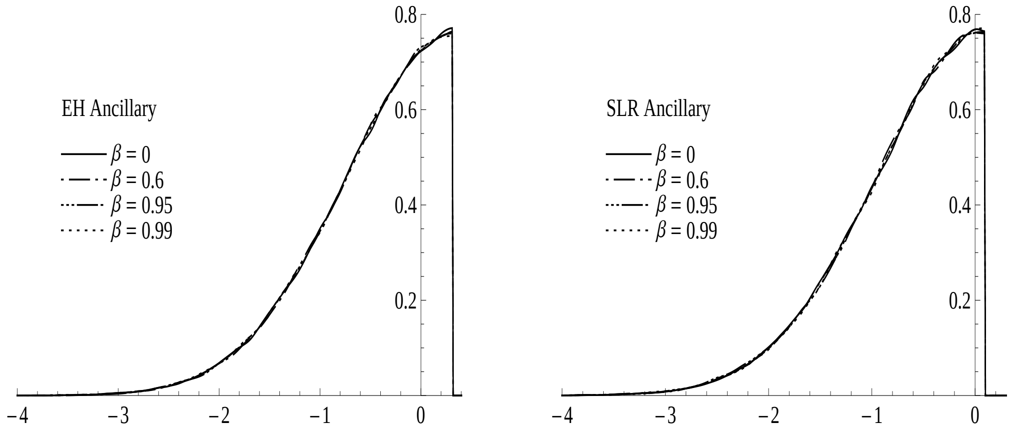

Figure 6: *Empirical Densities of approximate ancillaries for* $\beta = \{0, 0.6, 0.95, 0.99\}$.

(a: Left): Efron-Hinkley Ancillary, $n = 250, \gamma = 0.02$ *(b: Right): Signed LR Ancillary,* $n = 25, \gamma = 0.02$.

A more formal comparison is provided by the Kolmogorov-Smirnov test statistics reported in Table 5, which test the null hypothesis that the distribution of the ancillary statistic is the same for β_0 and β_1. The table only reports the upper triangle since the test is symmetric in β_0 and β_1, and the diagonal $\beta_0 = \beta_1$ is trivial (hence no row $\beta_0 = 0.99$ nor column $\beta_0 = 0$). The test statistics are well below the 5% critical value of 1.36, e.g. for testing if the distribution of signed-LR when β equals 0 versus 0.99 is 1.002, and in this sense the distributions are not significantly different for any of the parameter values in

[2] It was for instance not possible to show that the model constitutes a group family, which would have implied the existence of an exact ancillary.

the relevant range $\beta \in [0, 1)$. The Efron-Hinkley has lower values for the Kolmogorov-Smirnov statistics than the Signed-LR.

Efron-Hinkley					Signed-LR			
KS statistic	$\beta_1 = 0.6$	0.95	0.99		KS statistic	0.6	0.95	0.99
$\beta_0 = 0.0$	0.680	0.534	0.626		$\beta_0 = 0.0$	0.863	0.637	1.002
0.6	·	0.707	0.894		0.6	·	0.745	0.809
0.95	·	·	0.561		0.95	·	·	0.805
$n = 250, \gamma = 0.02, s_{zz} = 1$					$n = 25, \gamma = 0.02, s_{zz} = 1$			

Table 5: *Kolmogorov-Smirnov two-sample statistics for testing* $H_0 : pdf\,(a|\beta_0) = pdf(a|\beta_1)$, *based on* 100.000 *replications. Critical value at* 5% *level equals* 1.36.

Figures 6a and 6b show some further interesting features that are particular to the simple Keynesian model considered. The signed-LR has been shown in many regular problems to be closely approximated by a normal distribution, but in our model we see a clear cut-off to the right of 0 (here $a_{SLR} = 0.09$ and $a_{EH} = 0.31$)[3]. This cut-off is directly related to the critical set $t_1 = -\gamma\ n\ s_{zz}$ and is a consequence of the global curvature of this model as explained above. Using Figure 2 we showed that when t_1 crosses this line, the complementary solution for β to the first order condition becomes the global maximum. The signed-LR and Efron-Hinkley are calculated using this $\hat{\beta}$ and cannot be larger than the cut-off values. Although the cut-off points are independent of β, they change with other variables such as n, γ or s_{zz}, used to define the three scenarios.

The Efron-Hinkley and signed-LR both appear to have good ancillarity properties, even in very small samples and with weak instruments for a wide range of parameter values. They are therefore appropriate candidates for conditioning using standard arguments involving relevance of reference distribution, information recovery etc.. Below we show that, although these statistics have no information on the parameters by themselves, they do contain valuable information about the distribution of the MLE.

[3] The fact that the density slopes down, rather than going straight down is a consequence of the kernel estimation method used.

6 Conclusion

We have studied the simple Keynesian structural model à la Phillips (2006) Interest in this model that was first analysed by Bergstrom (1962) was rekindled because of questions concerning the effects of weak instruments and because of competing explanations for the bimodality of the marginal distribution of the MLE. This paper has contributed to this discussion first by providing a clear geometrical explanation and secondly by investigating conditional distributions rather than the marginal distribution and to show the importance of conditioning in this model. The geometrical representation proves instructive in understanding the ancillaries used. We investigate the well known Efron-Hinkley ancillary, which in this model equals the score- or affine ancillary, and a signed square root version of the likelihood ratio test, with the sign based on the geometry developed here. Both ancillaries are shown to be almost perfectly ancillary, even in very small samples and with weak instrument considered here. They are also shown to contain important information about the distribution of the MLE. The degree of bimodality not only depends on the strength of the instrument as Phillips (2006) has shown, but also depends heavily on the value of the ancillary statistic. In practice it does not seem to matter which of the two ancillaries is used. In some of the simulations there seems a slight preference for the Efron-Hinkley statistic, but this is not a general result. The important point is that the value of the ancillary should be used when making statements about the precision and properties of the MLE and inference should be carried out via the conditional distribution given the ancillary, rather than via the marginal distribution.

We have further shown that the saddlepoint approximation is extremely close to the actual conditional distribution using simulation. The saddlepoint approximation captures the bimodality of the distribution perfectly. the approximation was so good that it could not be distinguished from the true distribution and conclude that even in the cases considered here with small sample sizes, weak instruments and with large or little bimodality, the saddlepoint approximation is excellent.

7 Appendices

7.1 Appendix A

For one parameter Curved Exponential Models with canonical representation

$$p(x;\beta) = \exp[t'\eta(\beta) - \kappa(\beta) - h(t)],$$

with variance matrix

$$Var\,(t) = \Sigma_\beta = \Sigma_{\eta(\beta)} = \left.\frac{\partial \kappa(\eta)}{\partial \eta}\right|_{\eta=\eta(\beta)},$$

the Efron Curvature (Efron, 1975) was originally defined as:

$$\varrho_\beta = \left(\frac{|M_\beta|}{\upsilon_{20}^3(\beta)}\right)^{1/2},$$

where:

$$M_\beta = \begin{pmatrix} \upsilon_{20}(\beta) & \upsilon_{11}(\beta) \\ \upsilon_{11}(\beta) & \upsilon_{02}(\beta) \end{pmatrix} = \begin{pmatrix} \left(\frac{\partial\eta}{\partial\beta}\right)' \Sigma_\beta \left(\frac{\partial\eta}{\partial\beta}\right) & \left(\frac{\partial\eta}{\partial\beta}\right)' \Sigma_\beta \left(\frac{\partial^2\eta}{\partial\beta^2}\right) \\ \left(\frac{\partial\eta}{\partial\beta}\right)' \Sigma_\beta \left(\frac{\partial^2\eta}{\partial\beta^2}\right) & \left(\frac{\partial^2\eta}{\partial\beta^2}\right)' \Sigma_\beta \left(\frac{\partial^2\eta}{\partial\beta^2}\right) \end{pmatrix}.$$

In our model the canonical parameter is given by:

$$\eta = \begin{pmatrix} \eta_1(\beta) \\ \eta_2(\beta) \end{pmatrix} = \frac{1}{8} \begin{pmatrix} 2\gamma(1-\beta^2) \\ -(1-\beta)^2 \end{pmatrix},$$

and the covariance matrix of the canonical statistic is:

$$\Sigma_\beta = Var\,(t) = ns_{zz}\frac{4}{(1-\beta)^4} \begin{pmatrix} (1-\beta)^2 & 2\gamma(1-\beta^2) \\ 2\gamma(1-\beta^2) & 4(2/s_{zz} + (1+\beta)^2\gamma^2) \end{pmatrix}.$$

It is easy to see that:

$$\frac{\partial\eta}{\partial\beta} = \begin{pmatrix} -\frac{1}{2}\gamma\beta \\ \frac{1}{4}(1-\beta) \end{pmatrix}; \qquad \frac{\partial^2\eta}{\partial\beta^2} = \begin{pmatrix} -\frac{1}{2}\gamma \\ \frac{1}{4} \end{pmatrix}$$

and M becomes:

$$M_\beta = \begin{pmatrix} \frac{n(2+s_{zz}\gamma^2)}{(1-\beta)^2} & \frac{2n(1+s_{zz}\gamma^2)}{(1-\beta)^3} \\ \frac{2n(1+s_{zz}\gamma^2)}{(1-\beta)^3} & \frac{2n(1+2s_{zz}\gamma^2)}{(1-\beta)^4} \end{pmatrix}; \qquad |M_\beta| = \frac{2n^2 s_{zz}\gamma^2}{(1-\beta)^6}.$$

Therefore, in our model Effron Curvature equals:

$$\varrho_\beta = \left(\frac{\frac{2n^2 s_{zz}\gamma^2}{(1-\beta)^6}}{\frac{n^3(2+s_{zz}\gamma^2)^3}{(1-\beta)^6}}\right)^{1/2} = \left(2\frac{s_{zz}\gamma^2}{n(2+s_{zz}\gamma^2)^3}\right)^{1/2}.$$

7.2 Appendix B

LEMMA 1. The MLE in this model is given by:

$$\widehat{\beta} = \frac{t_2 + t_1\gamma + sign(t_1 + ns_{zz}\gamma)\sqrt{4nt_2 + 8nt_1\gamma + 4n^2\gamma^2 s_{zz} + t_1^2\gamma^2 + 2ns_{zz}t_1\gamma^3 + n^2 s_{zz}^2\gamma^4}}{2t_1\gamma + t_2 + ns_{zz}\gamma^2}, \quad t_1 + ns_{zz}\gamma \neq 0.$$

$$\hat{\beta} = 1 \pm 2\sqrt{n/\left(t_2 - \gamma^2 n\ s_{zz}\right)}, \quad t_1 + ns_{zz}\gamma = 0.$$

Comment: For $t_1 + ns_{zz}\gamma \neq 0$ there is a unique global maximizer and for $t_1 + ns_{zz}\gamma = 0$, their are two distinct roots to the first order conditions are

$$\hat{\beta} = 1 \pm 2\sqrt{n/\left(t_2 - \gamma^2 n\ s_{zz}\right)}$$

Proof. Let $\widehat{\beta}_1$ and $\widehat{\beta}_2$ be the roots of the first order condition for the MLE, equation (19):

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{t_2 + t_1\gamma + \sqrt{4nt_2 + 8nt_1\gamma + 4n^2\gamma^2 s_{zz} + t_1^2\gamma^2 + 2ns_{zz}t_1\gamma^3 + n^2 s_{zz}^2\gamma^4}}{2t_1\gamma + t_2 + ns_{zz}\gamma^2}, \\
\widehat{\beta}_2 &= \frac{t_2 + t_1\gamma - \sqrt{4nt_2 + 8nt_1\gamma + 4n^2\gamma^2 s_{zz} + t_1^2\gamma^2 + 2ns_{zz}t_1\gamma^3 + n^2 s_{zz}^2\gamma^4}}{2t_1\gamma + t_2 + ns_{zz}\gamma^2},
\end{aligned} \tag{27}$$

Define the likelihood difference for these two solutions:

$$\Delta\mathcal{L}(t_1; t_2, s_{zz}, \gamma, n) = \mathcal{L}(\widehat{\beta}_1) - \mathcal{L}(\widehat{\beta}_2), \tag{28}$$

which, after simplification, can be rewritten as:

$$\begin{aligned}
\Delta\mathcal{L}(t_1; t_2, s_{zz}, \gamma, n) &= \frac{1}{2(t_2 + \gamma(2t_1 + ns_{zz}\gamma)} \times \\
&\quad [\gamma(t_1 + ns_{zz}\gamma)\sqrt{\gamma^2(t_1 + ns_{zz}\gamma)^2 + 4n(t_2 + \gamma(2t_1 + ns_{zz}\gamma))} + \\
&\quad n(t_2 + \gamma(2t_1 + ns_{zz}\gamma) \times \\
&\quad Log(\frac{\sqrt{\gamma^2(t_1 + ns_{zz}\gamma)^2 + 4n(t_2 + \gamma(2t_1 + ns_{zz}\gamma))} + \gamma(t_1 + ns_{zz}\gamma)}{\sqrt{\gamma^2(t_1 + ns_{zz}\gamma)^2 + 4n(t_2 + \gamma(2t_1 + ns_{zz}\gamma))} - \gamma(t_1 + ns_{zz}\gamma)})].
\end{aligned} \tag{29}$$

Since t_2, s_{zz}, γ and $n > 0$ and

$$\begin{aligned}
t_2 + \gamma(2t_1 + ns_{zz}\gamma) &= \sum_{t=1}^{n}(x_t + y_t)^2 + 2\gamma\sum_{t=1}^{n}(x_t + y_t)z_t + \gamma^2\sum_{t=1}^{n} z_t^2, \\
&= \sum_{t=1}^{n}(x_t + y_t + \gamma z_t)^2 > 0,
\end{aligned} \tag{30}$$

we have three possibilities, depending on the value of $(t_1 + ns_{zz}\gamma)$:

(i) $t_1 + n s_{zz}\gamma = 0$, which implies:

$$\frac{\sqrt{\gamma^2(t_1+ns_{zz}\gamma)^2+4n(t_2+\gamma(2t_1+ns_{zz}\gamma)}+\gamma(t_1+ns_{zz}\gamma)}{\sqrt{\gamma^2(t_1+ns_{zz}\gamma)^2+4n(t_2+\gamma(2t_1+ns_{zz}\gamma)}-\gamma(t_1+ns_{zz}\gamma)}=1, \tag{31}$$

and hence that $\Delta\mathcal{L}(t_1; t_2, s_{zz}, \gamma, n) = 0$, since the numerator and denominator inside the Log function in (29) are equal. This implies:

$$\mathcal{L}(\widehat{\beta}_1) = \mathcal{L}(\widehat{\beta}_2) \tag{32}$$

for the two distinct roots $1 \pm 2\sqrt{n/\,(t_2 - \gamma^2 n\ s_{zz})}$.

(ii) $t_1 + n s_{zz}\gamma > 0$, which implies:

$$\frac{\sqrt{\gamma^2(t_1+ns_{zz}\gamma)^2+4n(t_2+\gamma(2t_1+ns_{zz}\gamma)}+\gamma(t_1+ns_{zz}\gamma)}{\sqrt{\gamma^2(t_1+ns_{zz}\gamma)^2+4n(t_2+\gamma(2t_1+ns_{zz}\gamma)}-\gamma(t_1+ns_{zz}\gamma)}>1, \tag{33}$$

and

$$\Delta\mathcal{L}(t_1; t_2, s_{zz}, \gamma, n) > 0, \tag{34}$$

as can be seen from (29). Hence $\mathcal{L}(\widehat{\beta}_1) > \mathcal{L}(\widehat{\beta}_2)$ and $\widehat{\beta}_1$ maximizes the likelihood globally.

(iii) $t_1 + n s_{zz}\gamma < 0$ analogous to (ii) but leading to $\mathcal{L}(\widehat{\beta}_1) < \mathcal{L}(\widehat{\beta}_2)$ and $\widehat{\beta}_2$ maximizing the likelihood globally.

The global solution(s) therefore depend on $sign(t_1 + n s_{zz}\gamma)$ as stated in the Lemma.

7.2.1 The Affine or Score Ancillary

Barndorff-Nielsen (1980) proposed the following matrix for A in his Affine ancillary:

$$A = \left(\frac{\partial\tau}{\partial\beta}\right)_{\perp} \Lambda^{-1/2}. \tag{35}$$

where $\Lambda = (\frac{\partial\tau}{\partial\beta})'_{\perp}\Sigma_t(\frac{\partial\tau}{\partial\beta})_{\perp}$ is a $(k-d)\times(k-d)$ matrix, with $(\frac{\partial\tau}{\partial\beta})_{\perp}$ denoting a $k\times(k-d)$ matrix whose column-vectors are linearly independent and orthogonal to the column vectors of $(\frac{\partial\tau}{\partial\beta})$ and Σ_t is the covariance matrix of the minimal sufficient statistic t.

In a CEM (2,1) we have in particular:

$$(\frac{\partial\tau}{\partial\beta})_{\perp} = \left|\Sigma_t^{-1}\right|(-\frac{\partial\tau_2}{\partial\beta}, \frac{\partial\tau_1}{\partial\beta})'. \tag{36}$$

Writing the 1×2 matrix $A\left(\hat{\beta}\right) = (\widehat{A}_1, \widehat{A}_2)$ which is fixed for given $\hat{\beta}$, we have:

$$a = \widehat{A}_1(t_1 - \widehat{\tau}_1) + \widehat{A}_2(t_2 - \widehat{\tau}_2). \tag{37}$$

Using this expression and the equation that defines the estimate we can get expressions for t_1 and t_2 in terms of $\widehat{\beta}$ and a:

$$t_1 = \frac{a - 4\widehat{A}_2 n - a\widehat{\beta}(2-\widehat{\beta}) - \widehat{A}_2 s_{zz}\gamma^2(1-\widehat{\beta}^2) + (1-\widehat{\beta})^2(\widehat{A}_1\widehat{\tau}_1 + \widehat{A}_2\widehat{\tau}_2)}{(1-\widehat{\beta})(\widehat{A}_1(1-\widehat{\beta}) + 2\widehat{A}_2\widehat{\beta}\gamma)}. \tag{38}$$

$$t_2 = \frac{4\widehat{A}_1 n + \widehat{A}_1 s_{zz}\gamma^2(1-\widehat{\beta}^2) + 2\gamma\widehat{\beta}(1-\widehat{\beta})(a + \widehat{A}_1\widehat{\tau}_1 + \widehat{A}_2\widehat{\tau}_2)}{(1-\widehat{\beta})(\widehat{A}_1(1-\widehat{\beta}) + 2\widehat{A}_2\widehat{\beta}\gamma)}. \tag{39}$$

Equation (19) remains valid, and since $\widetilde{\beta}, \widetilde{\sigma}^2$ (or $\hat{\eta}_1$ and $\hat{\eta}_2$) and $\widehat{\beta}$ can be expressed in terms of t_1 and t_2, we can use (26) to get this transformation. The main problem with this approach is that now we have a system of simultaneous non-linear equations that can be difficult to solve in closed form. Alternatively, for a given pair $(\widehat{\beta}, a_1)$ we can find an equivalent affine ancillary and use (38) and (39) to find t_1 and t_2. We have used this technique to find values for t_1 and t_2 for a given value of a_1 and different values for $\widehat{\beta}$.

With these remarks, the saddlepoint approximation when conditioning on the signed likelihood ratio ancillary similarly becomes:

$$\begin{aligned} f(\widehat{\beta};\beta \quad | \quad a_{SLR}) &\cong c(a_{SLR})[\frac{n}{(1-\widehat{\beta})^2} + \frac{2t_1\gamma + t_2 + s_{zz}\gamma^2}{4}]^{1/2}[\frac{(1-\beta)^2}{(1-\widehat{\beta})^2}]^{n/2} \\ &\times \exp\left[\frac{1}{8}(\widehat{\beta}-\beta)(\gamma^2 s_{zz}(\widehat{\beta}+\beta+2) + t_2(\widehat{\beta}+\beta-2) + 2\gamma t_1(\widehat{\beta}+\beta)\right]. \end{aligned} \tag{40}$$

Since we are using again the affine ancillary to compute t_1 and t_2 we may think that the conditional distribution results are also equivalent. But this is only true if there is a one-to-one transformation between the two ancillaries. In fact, this is not the case.

7.3 Appendix C : Conditional Saddlepoint Approximations

7.4 Efron-Hinkley statistic

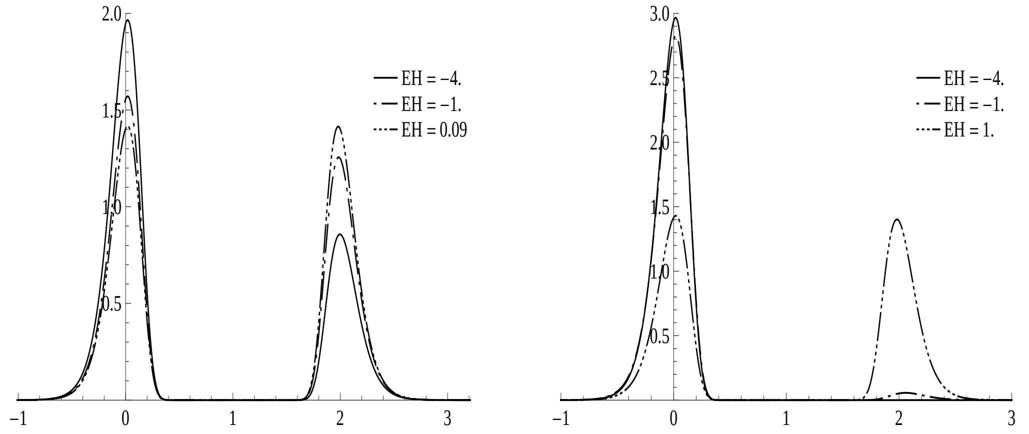

Figure 7: *Saddlepoint Approximation for the MLE Density conditional on the Efron Hinkley Ancillary,* $n = 25$, $\gamma = 0.02$, $szz = 1, \beta = 0$. *(a: Left)* $\gamma = 0.02$ *and (b: Right)* $\gamma = 0.2$

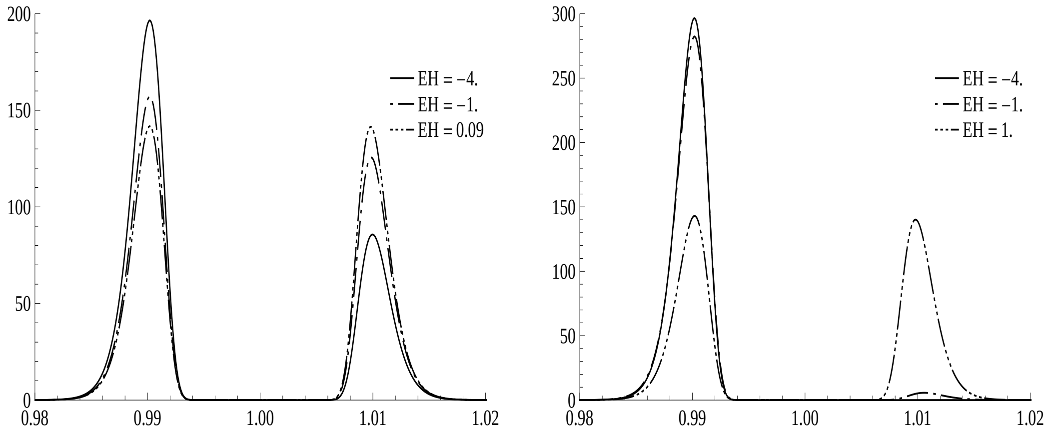

Figure 8: *Saddlepoint Approximation for the MLE Density conditioned on the Efron Hinkley Ancillary,* $n = 25$, $\gamma = 0.02$, $szz = 1, \beta = 0.99$. *(a: Left)* $\gamma = 0.02$ *and (b: Right)* $\gamma = 0.2$

7.5 Signed Likelihood Ratio statistic

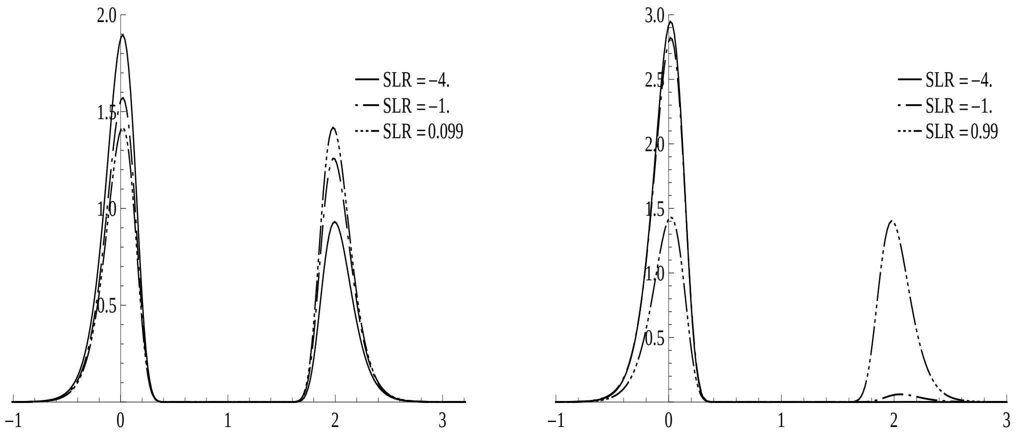

Figure 9: *Saddlepoint Approximation for the MLE Density conditioned on the Signed Likelihood Ratio Ancillary, $n = 25$, $\gamma = 0.02$, $szz = 1, \beta = 0$. (a: Left) $\gamma = 0.02$ and (b: Right) $\gamma = 0.2$*

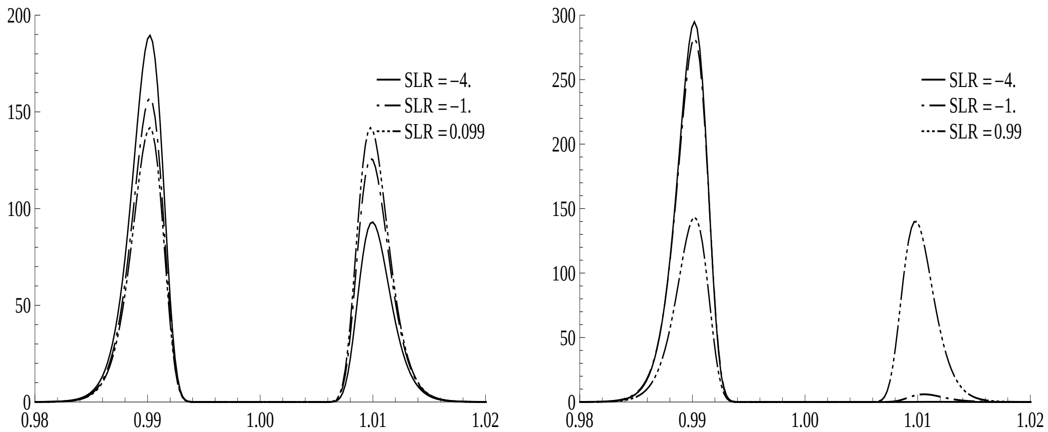

Figure 10: *Saddlepoint Approximation for the MLE Density conditioned on the Signed Likelihood Ratio Ancillary, $n = 25$, $\gamma = 0.02$, $szz = 1, \beta = 0$. (a: Left) $\gamma = 0.02$ and (b: Right) $\gamma = 0.2$*

7.6 Appendix D: Ancillarity Properties

7.6.1 Efron Hinkley Ancillary

Scenario I

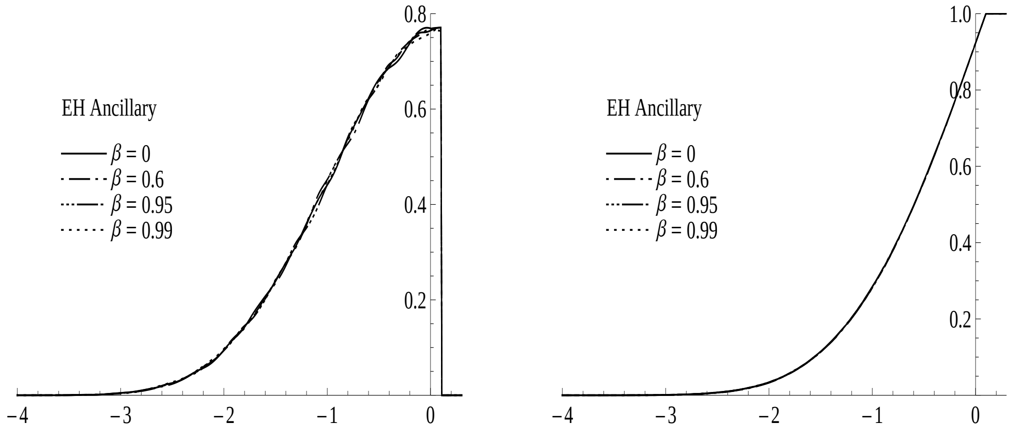

Figure 11: *(a: Left), Empirical Density; (b: Right), Cumulative Distribution Efron - Hinkley Ancillary, Scenario I;* $n = 25$, $\gamma = 0.02$, $szz = 1$, $\beta = \{0, 0.6, 0.9, 0.99\}$

KS statistic	$\beta = 0.6$	$\beta = 0.95$	$\beta = 0.99$
$\beta = 0$	0.771	0.85	0.597
$\beta = 0.6$	·	0.87	0.995
$\beta = 0.95$	·	·	1.044

Table 6: *Kolmogorov-Smirnov statistics for the EH ancillary in Scenario I Critical value at* $5\% = 1.36, Rep =$ 100.000

Scenario III

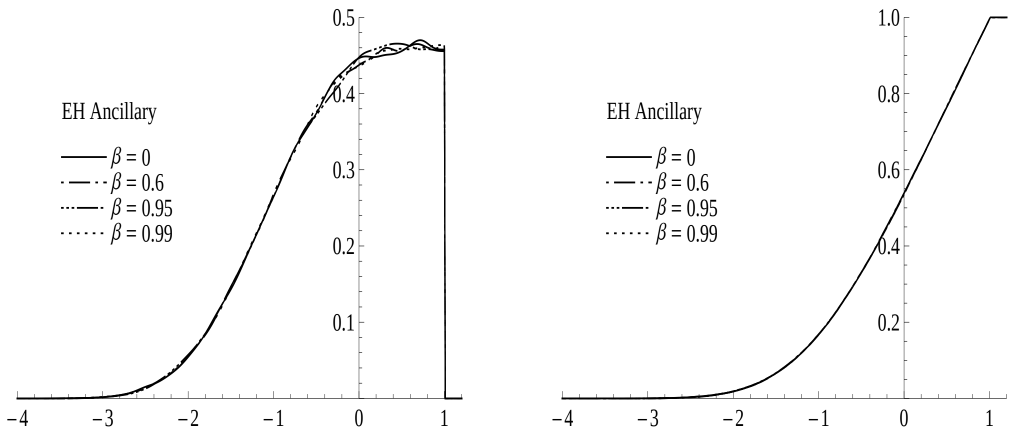

Figure 12: *(a: Left), Empirical Density (b: Right), Cumulative Distribution Efron Hinkley Ancillary ScenarioIII: n = 25,* $\gamma = 0.2$, $szz = 1$, $\beta = \{0, 0.6, 0.9, 0.99\}$

KS statistic	$\beta = 0.6$	$\beta = 0.95$	$\beta = 0.99$
$\beta = 0$	0.546	0.821	0.796
$\beta = 0.6$	·	1.04	0.416
$\beta = 0.95$	·	·	1.111

Table 7: *Kolmogorov-Smirnov statistics for the EH ancillary in scenario III,Rep=100.000,Criticalvalueat5%=1.36*

7.6.2 Signed Likelihood Ratio Ancillary

Scenario II

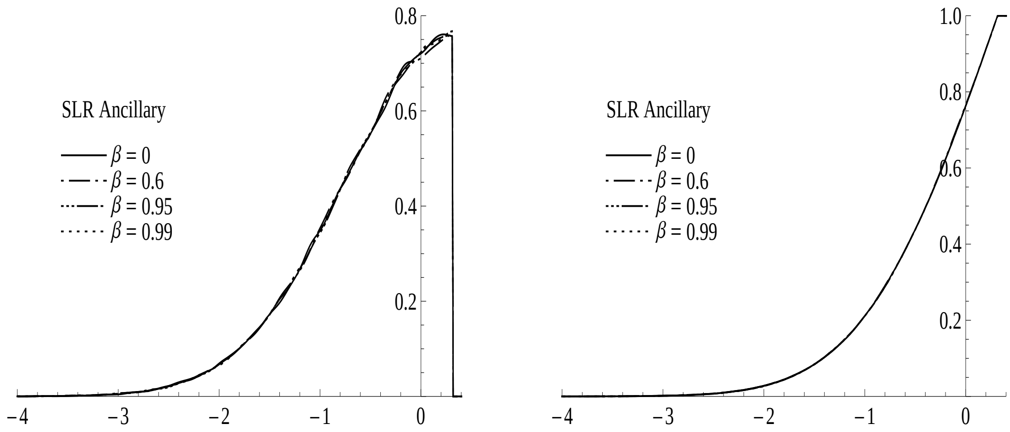

Figure 13: *(a: Left), Empirical Density (b: Right), Cummulative Distribution Signed LR Ancillary* $n = 250$, $\gamma = 0.02, s_{zz} = 1$, $\beta = \{0, 0.6, 0.9, 0.99\}$

KS statistic	$\beta = 0.6$	$\beta = 0.95$	$\beta = 0.99$
$\beta = 0$	1.093	0.561	0.590
$\beta = 0.6$	·	1.129	0.899
$\beta = 0.95$	·	·	0.635

Table 8: *Kolmogorov-Smirnov statistics for the SLR ancillary,* $n = 250$, $\gamma = 0.02, s_{zz} = 1$ *, based on* 100.000 *replications Critical value at*

8 REFERENCES

Barndorff-Nielsen, O.E. (1980) Conditionality resolutions. *Biometrika* **67,** 343-365

Barndorff-Nielsen, O.E. (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70,**343-365

Barndorff-Nielsen, O.E. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41,** 279-312

Barndorff-Nielsen, O.E. and D.R. Cox. *Inference and Asymptotics.* Chapman and Hall, 1994.

Bergstrom, A.R. (1962) The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. *Econometrica* **30,**480-490

Forchini, G. (2006), On the bimodality of the exact distribution of the TSLS estimator. *Econometric Theory* **Vol. 22,** 932-946

Hall, P., S.J. Sheather, M.C. Jones, and J.S. Marron (1991), On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **Vol. 78, No. 2,** 263-269

Hillier, G.H. (2006), Yet more on the exact properties of IV estimators. *Econometric Theory* **Vol. 22,** 913-931

Larsson, R. (1999), Approximation of the asymptotic distribution of the log likelihood ratio test for cointegration. *Econometric Theory* **Vol. 15,** 789-813

Lieberman, O. (1994) Saddlepoint Approximation for the Distribution of a Ratio of Quadratic Forms in Normal Variables. *Journal of the American Statistical Association*, **Vol. 89**, No. 427, 924-928

Lieberman, O (1994) Saddlepoint Approximation for the Least Squares Estimator in First-Order Autoregression. *Biometrika*, **Vol. 81**, No. 4 , 807-811 Amendments and corrections: (1996) **Vol. 83**, No. 1, p. 247

Lieberman, O (1994)On the Approximation of Saddlepoint Expansions in Statistics *Econometric Theory*, **Vol. 10**, No. 5 , 900-916

Maddala, G. and Jeong, J. (1992), On the exact small sample distribution of the instrumental variable estimator. *Econometrica* **Vol. 60,** No. 1, 181-183

Nelson, Ch. and Startz, R. (1990), The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *The Journal of Business* **Vol. 63,** No. 1, 125-140