

NONLINEAR INSTRUMENTAL VARIABLE ESTIMATION OF AN AUTOREGRESSION¹

Peter C. B. Phillips
Cowles Foundation, Yale University
University of Auckland and University of York

Joon Y. Park
School of Economics
Seoul National University

and

Yoosoon Chang
Department of Economics
Rice University

April 2002

¹Computations were performed by the authors in GAUSS and the paper was typed by the authors in SW2.5. Phillips thanks the NSF for research support under Grants SBR 94-22922, SBR97-30295 and SES 0092509. Park and Chang thank the Cowles Foundation for hospitality during a visit in June-July 1998. Park also gratefully acknowledges the financial support from the Korea Research Foundation.

Abstract

Instrumental variable (IV) estimation methods that allow for certain nonlinear functions of the data as instruments are studied. The context of the discussion is the simple unit root model where certain advantages to the use of nonlinear instruments are revealed. In particular, certain classes of IV estimators and associated t-tests are shown to have simpler (standard) limit theory in contrast to the least squares estimator, providing an opportunity for the study of optimal estimation in certain IV classes and furnishing tests and confidence intervals that allow for unit root and stationary alternatives. The Cauchy estimator studied in recent work by So and Shin (1999) is shown to have such an optimality property in the class of certain IV procedures with bounded instruments.

Key words and phrases: Cauchy estimator, instrumental variable autoregression, nonlinear instruments, sojourn time, unit root.

JEL Classification Numbers: C22, C25

1. Introduction

Instrumental variable (IV) estimation is a traditional tool of econometric analysis and has found extensive uses throughout the subject, covering both microeconomic and time series applications. With regard to the latter, the dominant class of models for which instrumental variable methods have been advocated are stationary, linear systems, a subject on which there is now a vast literature in both econometrics and statistics, a classic study being that of Deistler and Hannan (1988). IV techniques have been used in some nonstationary models, including cointegrating system estimation (Phillips and Hansen, 1990), but even in those cases linear instruments have been used.

The present paper studies the use of nonlinear IV techniques in which the instruments are nonlinear functions of integrated processes. The context of our discussion is the simple unit root model and it is shown that here there are certain advantages to the use of nonlinear instruments. The reason is that nonlinear functions of integrated processes offer a very wide range of alternative time series behavior, allowing for both attenuation and exaggeration of the effects of nonstationarity. This range of behavior in potential (nonlinear) instruments in turn gives rise to a variety of possible limit behavior for IV estimators of the autoregressive coefficient. Among these possibilities are included some interesting cases where the IV estimator and its associated t-ratio have similar asymptotic behavior in unit root and stationary autoregressive models. Such cases provide an opportunity for developing new methods of testing and confidence interval construction that allow for both unit root and stationary alternatives.

In recent work, So and Shin (1999) suggested the use of the Cauchy (1836) estimator, which uses the sign function as an instrumental variable, in place of the ordinary least squares (OLS) estimator in autoregressions that included both stationary and nonstationary cases. Our framework extends the analysis of these authors, providing a more general analysis of IV estimation in potentially nonstationary autoregressions and showing that the Cauchy estimator has an optimality property in the class of certain IV procedures. Our analysis is limited to the AR(1) model but it is hoped that the results are sufficiently interesting to warrant further investigation of these techniques.

The paper is organised as follows. Section 2 lays out the model, assumptions and gives some background theory. Section 3 classifies classes of IV estimators according to the nature of the instrument functions and provides a limit theory for the various cases. Section 4 considers the issue of optimality and derives optimal estimators for two classes of instrument functions. Section 5 discusses testing and confidence interval construction. Some finite sample simulations are reported in Section 6. Section 7 concludes the paper and derivations are collected together in Section 8. Notation is summarized in Section 9.

2. Assumptions and Background Theory

We consider the unit root autoregression

$$y_t = \alpha y_{t-1} + u_t, \quad \alpha = 1 \quad (1)$$

with arbitrary initial condition $y_0 = O_p(1)$. The regression error u_t is assumed to have zero mean and satisfy one of the following conditions.

2.1 Assumption $\{u_t\}$ is a martingale difference sequence with $\sigma^2 = \mathbf{E}(u_t^2 | \mathcal{F}_t)$ and $\sup_{t \geq 1} \mathbf{E}(|u_t|^p | \mathcal{F}_t) < \infty$ a.s. for some $p > 2$, where $\mathcal{F}_t = \sigma(\{u_s\}_{s \leq t})$.

2.2 Assumption $\{u_t\}$ is an iid $(0, \sigma^2)$ sequence of random variables with $\mathbf{E}|u_t|^p < \infty$ for some $p > 4$, and its distribution is absolutely continuous with respect to Lebesgue measure and has characteristic function φ such that $\lim_{\lambda \rightarrow \infty} \lambda^r \varphi(\lambda) = 0$ for some $r > 0$.

Both Assumptions 2.1 and 2.2 imply that y_{t-1} is uncorrelated with u_t and regression (1) is correctly specified. Though our theory may be extended to allow for more general linear processes that are serially correlated, we will consider this simple case in the present paper for the brevity of exposition. For some of our subsequent results, the martingale difference assumption in Assumption 2.1 is sufficient. For others, however, we need more restrictive independence and distributional assumptions as in Assumption 2.2.

We also consider time series with nonzero mean and deterministic trend in the paper. For these cases, we could consider regression (1) with demeaned or detrended y_t . However, conventional demeaning and detrending invalidates the asymptotics for nonlinear transformations of unit root processes that is developed in the paper. This is because the usual procedure utilizes present and future as well as past observations in demeaning and detrending, with the consequence that transformations of the demeaned and detrended series end up being correlated with the regression errors. The problem is overcome by employing a recursive demeaning and detrending that relies only on past observations. This approach is discussed in a later section.

Define $B_n(r) = n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} u_t$. Then an invariance principle holds under Assumption 2.1, i.e.,

$$B_n \rightarrow_d B, \quad (2)$$

where B is Brownian motion with variance σ^2 . We will sometimes be convenient to write $B = \sigma W$, where W is standard Brownian motion with unit variance. Our asymptotic theory also relies on the Brownian local time process (e.g Chung and Williams, 1990). In general, the local time $L_M(t, s)$ of a continuous semimartingale M is defined as

$$L_M(t, s) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t 1_{\{|M(r) - s| < \epsilon\}} d[M](r),$$

where $[M]$ is the quadratic variation process of M . The local time $L_M(t, s)$ thus measures in the unit of $[M]$ how long M stays in the neighborhood of level s up to time t . If $L_B(t, s)$ and $L_W(t, s)$ are the local times of B and W , respectively, then we may write

$$L_B(t, s) = (1/\sigma)L_W(t, s/\sigma).$$

It is often more convenient to use the scaled version $L(t, s)$ defined by

$$L(t, s) = (1/\sigma^2)L_B(t, s) = (1/\sigma^3)L_W(t, s/\sigma),$$

which is called *chronological local time* (Phillips and Park, 1998) because it refers to the amount of calendar time spent by the process in the vicinity of a spatial point s over the time interval $[0, t]$. In terms of chronological local time, we may represent the integral of any locally bounded function T of Brownian motion as follows

$$\int_0^t T(B) dr = \int_0^t T(\sigma W) dr = \int_{-\infty}^{\infty} T(s)L(t, s) ds, \quad (3)$$

a formula known as the *occupation times formula*.

We now introduce two classes of transformations in \mathbf{R} which play important roles in the subsequent development of our theory. First, we define the functions that will be referred to as *regularly integrable* in the paper.

2.3 Definition *A transformation T on \mathbf{R} is said to be regularly integrable if T is a bounded integrable function such that for some constants $c > 0$ and $k > 6/(p - 2)$ with $p > 4$ given in Assumption 2.2 $|T(x) - T(y)| \leq c|x - y|^k$ on each piece A_i of its support $A = \bigcup_{i=1}^m A_i \subset \mathbf{R}$.*

The second class of functions is the class of regular functions introduced in Park and Phillips (2001). We say that a transformation T is *regular* if and only if (a) it is continuous in a neighborhood of infinity, and (b) for any compact subset K of \mathbf{R} given, there exist for each $\varepsilon > 0$ continuous functions $\underline{T}_\varepsilon, \overline{T}_\varepsilon$ and $\delta_\varepsilon > 0$ such that $\underline{T}_\varepsilon(x) \leq T(y) \leq \overline{T}_\varepsilon(x)$ for all $|x - y| < \delta_\varepsilon$ on K , and such that $\int_K (\overline{T}_\varepsilon - \underline{T}_\varepsilon)(x) dx \rightarrow 0$ as $\varepsilon \rightarrow 0$. It is easy to see that a piecewise continuous function has this property. The reader is referred to Park and Phillips (2001) for a more detailed discussion on this class of regular functions.

It is also necessary to define the order of a function $S : \mathbf{R} \times \mathbf{R}_+ \rightarrow \mathbf{R}$ with two arguments (x, λ) , in terms of a function $\nu : \mathbf{R}_+ \rightarrow \mathbf{R}$ with argument λ . We say that S is of order smaller than ν if and only if we may write: either

$$(a) S(x, \lambda) = a(\lambda)P(x) \text{ with}$$

$$\limsup_{\lambda \rightarrow \infty} a(\lambda)/\nu(\lambda) = 0$$

and where P is locally bounded with exponentially bounded tails (i.e., $|P(x)| \leq e^{c|x|}$ for all large $|x|$, where $c > 0$ is a constant); or

(b) $S(x, \lambda) = b(\lambda)Q(x)$ with

$$\limsup_{\lambda \rightarrow \infty} b(\lambda)/\nu(\lambda) < \infty,$$

and where Q is bounded and vanishes at infinity (i.e., $Q(x) \rightarrow 0$ as $|x| \rightarrow \infty$).

2.4 Definition A transformation T on \mathbf{R} is said to be asymptotically homogeneous if it can be written as

$$T(\lambda x) = \kappa(\lambda)H(x) + R(x, \lambda),$$

where H is regular and R is of order smaller than κ . We call κ the asymptotic order and H the limit homogeneous function of F .

Roughly speaking, regularly integrable transformations are integrable functions that are reasonably smooth on each part of their supports. The required smoothness depends on the moment condition of the innovation sequence u_t . Let p be the maximum order of the existing moments. If $p > 8$, any piecewise Lipschitz continuous function is allowed. For the indicator function on a compact interval to be regularly integrable, on the other hand, it is sufficient to have $p > 4$.

Asymptotically homogeneous transformations are functions that behave asymptotically as homogeneous functions. Of course, homogeneous functions are asymptotically homogeneous and, therefore, functions such as the sign function, $\text{sgn}(x)$, and the power function, $|x|^k$, are asymptotically homogeneous. A much wider class of transformations, however, are asymptotically homogeneous. For instance, the functions $\log(x)$ and $\arctan(x)$ are asymptotically homogeneous, with respective asymptotic orders $\nu(\lambda) = \log \lambda$ and 1, and limit homogeneous functions $H(x) = 1$ and $(\pi/2)\text{sgn}(x)$. Also, all distribution function-like transformations are asymptotically homogeneous with the same asymptotic order $\nu(\lambda) = 1$ and the same limit homogeneous function $H(x) = 1\{x \geq 0\}$.

The asymptotic behavior of nonlinear functions of an integrated time series y_t generated as in (1) are analyzed by Park and Phillips (1999, 2001). In particular, they provide the relevant asymptotic theory for the sample moments $\sum_{t=1}^n T(y_t)$ and $\sum_{t=1}^n T(y_{t-1})u_t$, which are referred to as *mean and covariance asymptotics*, respectively, for various types of function T including the regularly integrable and asymptotically homogeneous functions considered here.

The present paper considers IV estimation of the autoregression (1) with an instrument given by $z_t = F(y_{t-1})$ for a nonlinear function F . The transformation F will be called the *instrument generating function* (IGF) throughout the paper. Our theory therefore concentrates on the covariance asymptotics for the IGF F , and the mean asymptotics for the function uF given by

$$uF(x) = xF(x), \tag{4}$$

i.e., the product of F with identity. Note that we need to analyze the product moment of z_t and u_t , and that of z_t and y_{t-1} , for the IV estimation.

If F is asymptotically homogeneous, then so is ιF , and vice versa. In particular, if we let F be an asymptotically homogeneous function with asymptotic order κ and limit homogeneous function H , then the associated ιF function defined in (4) becomes asymptotically homogeneous with asymptotic order $\iota\kappa$ and limit homogeneous function ιH , where $\iota\kappa$ and ιH are defined respectively as $\iota\kappa(\lambda) = \lambda\kappa(\lambda)$ and $\iota H(x) = xH(x)$. This is not necessarily so when F is regularly integrable, and the function ιF in (4) may not be regularly integrable. However, a regularly integrable F often has a corresponding ιF function that is regularly integrable. For instance, if F is a regularly integrable function such that $\int_{-\infty}^{\infty} sF(s)ds < \infty$ and is piecewise Lipschitz continuous with $p > 8$, then ιF defined in (4) above is regularly integrable. By convention in this paper, we will call the IGF F a regularly integrable function, only when both F and the corresponding ιF are regularly integrable. This convention will be made throughout the paper.

In the development of the asymptotic theory for IV estimation with an asymptotically homogeneous IGF, Assumption 2.1 will be invoked throughout the paper. The condition is fairly weak and is just enough to guarantee the invariance principle in (2). To develop the asymptotic theory for IV estimation with a regularly integrable IGF, however, we require the stronger conditions in Assumption 2.2. Roughly speaking, this is because we need weak convergence to and invariance of the sample local time, as well as the invariance principle (2), in order to develop asymptotics for integrable transformations of integrated time series.

3. IV Estimation and Limit Theory

In the autoregression (1), we consider the IV estimator of α given by

$$\hat{\alpha} = \frac{\sum_{t=1}^n F(y_{t-1}) y_t}{\sum_{t=1}^n F(y_{t-1}) y_{t-1}}. \quad (5)$$

Here, $\hat{\alpha}$ is an IV estimator in which the instrument is generated by the IGF F . In its general form, the class of IV estimators that can be represented by (5) includes, of course, the conventional OLS estimator as a special case with the linear IGF $F(x) = x$. However, this paper will concentrate on IV estimators constructed with various nonlinear IGF's.

One important example is the IV estimator with the asymptotically homogeneous IGF

$$F(x) = \text{sgn}(x).$$

The resulting estimator is the so-called *Cauchy estimator*, which derives from an original suggestion by Cauchy (1836), and which has recently been investigated by So and Shin (1999). Another is the IV estimator using as the IGF the regularly integrable function

$$F(x) = 1\{|x| \leq 1\},$$

or more generally any indicator function on a compact interval. One may easily see that, in this case, the corresponding IV estimator is simply a *trimmed OLS*

estimator, i.e., the OLS estimator which uses only those observations taking values in some compact interval. The behavior of IV estimators in the general class of (5) depends on the properties of the IGF, as we now investigate.

First, we present mean and covariance asymptotics for regularly integrable and the asymptotically homogeneous IGF F . For this purpose, it is convenient to define some additional notation in the case of a regularly integrable IGF F . Let

$$\overline{IF}(x) = \int_{-\infty}^x IF(y)dy,$$

where IF is defined in (4), and for a transformation T on \mathbf{R} define

$$I(T) = \int_{-\infty}^{\infty} T(x)dx, \quad \text{and} \quad I_s(T) = \int_{-\infty}^{\infty} T(x)^2 dx.$$

These notations will be used repeatedly in the rest of the paper without further reference.

3.1 Lemma *Let F be regularly integrable, and let Assumption 2.2 hold. Then*

$$n^{-1/4} \sum_{t=1}^n F(y_{t-1}) u_t \rightarrow_d \sigma [I_s(F)L(1,0)]^{1/2} U(1), \quad (6)$$

where U is standard Brownian motion independent of L . We also have

$$n^{-1/2} \sum_{t=1}^n F(y_t) y_t \rightarrow_d I(IF)L(1,0), \quad (7)$$

$$n^{-1/2} \sum_{t=1}^n F(y_t)^2 \rightarrow_d I_s(F)L(1,0). \quad (8)$$

Moreover, if $I(IF) = 0$, then

$$n^{-1/4} \sum_{t=1}^n F(y_t) y_t \rightarrow_d (2/\sigma) [I_s(\overline{IF})L(1,0)]^{1/2} V(1), \quad (9)$$

where V is standard Brownian motion independent of L , and V has correlation coefficient $\rho = -I(F\overline{IF})/[I_s(F)I_s(\overline{IF})]^{1/2}$ with U , as introduced above.

3.2 Lemma *Let F be asymptotically homogeneous with asymptotic order κ and limit homogeneous function H , and let Assumption 2.1 hold. Then we have*

$$n^{-1/2} \kappa(n^{1/2})^{-1} \sum_{t=1}^n F(y_{t-1}) u_t \rightarrow_d \int_0^1 H(B) dB, \quad (10)$$

and

$$n^{-3/2} \kappa(n^{1/2})^{-1} \sum_{t=1}^n F(y_t) y_t \rightarrow_d \int_0^1 H(B) B dr, \quad (11)$$

$$n^{-1} \kappa(n^{1/2})^{-2} \sum_{t=1}^n F(y_t)^2 \rightarrow_d \int_0^1 H(B)^2 dr, \quad (12)$$

where B is Brownian motion with variance σ^2 .

Of the asymptotic results presented in Lemmas 3.1 and 3.2, those in (6) – (8) and (10) – (12) can be obtained as special cases of Park and Phillips (1999, 2001). The result given in (9) is new.

For some detailed discussion on the asymptotics here, the reader is referred to Park and Phillips (1999, 2001) and Phillips and Park (1998). For a regularly integrable function F , the covariance asymptotics yield mixed normal limiting distributions with a mixing variate depending upon the local time L of the limit Brownian motion, as well as the integral of the square of the transformation function. Note that the Brownian motion V in the limit variate (9) is independent of the local time L . In contrast, the covariance asymptotics for an asymptotically homogeneous function F involve limiting distributions that are generally non-Gaussian except for some special cases.

3.3 Theorem (a) *Let IGF F be regularly integrable and let Assumption 2.2 hold. If $I(\iota F) = 0$, we have*

$$\hat{\alpha} - 1 \rightarrow_d \frac{\sigma^2 I_s(F)^{1/2} U(1)}{2 I_s(\overline{\iota F})^{1/2} V(1)} \quad (13)$$

where U and V are standard Brownian motions defined in Lemma 2.5. If, on the other hand, $I(\iota F) \neq 0$, then

$$n^{1/4}(\hat{\alpha} - 1) \rightarrow_d \frac{\sigma I_s(F)^{1/2} U(1)}{I(\iota F)L(1,0)^{1/2}} \quad (14)$$

(b) *Let IGF F be asymptotically homogeneous and let Assumption 2.1 hold. Then we have*

$$n(\hat{\alpha} - 1) \rightarrow_d \frac{\int_0^1 H(B) dB}{\int_0^1 H(B) B dr} \quad (15)$$

From this theorem it is apparent that IV estimators with instruments generated from a broad class of functions are consistent. If they are constructed from asymptotically homogeneous IGF's, their convergence rate is n . This is well known for the case of the linear IGF $F(x) = x$, which yields the OLS estimator. Theorem 3.3 shows that this result applies across all IV estimators constructed under the broad class of asymptotically homogeneous IGF's. The limiting distributions of IV estimators in this class are generally non-Gaussian, just as in the well studied case of the OLS estimator. In contrast, if regularly integrable functions are used to generate the instruments the corresponding IV estimators converge at the reduced rate $n^{1/4}$, and their limiting distributions are mixed normal.

However, there are cases when IV estimators of α are inconsistent. In particular, when the instrument is generated by a regularly integrable function F for which

$$\int_{-\infty}^{\infty} xF(x) dx = 0 \quad (16)$$

then the resulting IV estimator becomes inconsistent, as evidenced in Theorem 3.3(a). The condition (16) implies that the IGF F is in this case orthogonal to the regression function, x , in the Hilbert space $L^2(\mathbf{R})$ of square integrable functions. We recall from the theory of standard stationary regression that an instrument is invalid and the resulting IV estimator is inconsistent if the instrument fails the ‘relevance’ condition and is asymptotically uncorrelated with the regressor. We see from the present result that such instrument failure can also arise in nonstationary regression with an integrated regressor when the instrument function is orthogonal to the regression function.

Under the conditions in Assumption 2.1 or 2.2, the conditional variance of the numerator of the estimation error

$$\hat{\alpha} - \alpha = \frac{\sum_{t=1}^n F(y_{t-1}) u_t}{\sum_{t=1}^n F(y_{t-1}) y_{t-1}}$$

is given by $\sigma^2 \sum_{t=1}^n F(y_{t-1})^2$. Correspondingly, the asymptotic variation of $\hat{\alpha}$ depends on the asymptotic behavior of

$$\sigma^2 \frac{\sum_{t=1}^n F(y_{t-1})^2}{(\sum_{t=1}^n F(y_{t-1}) y_{t-1})^2}.$$

The ratio

$$P_n(F) = \frac{(\sum_{t=1}^n F(y_{t-1}) y_{t-1})^2}{\sum_{t=1}^n F(y_{t-1})^2}$$

therefore provides a measure of the conditional precision of the IV estimator with IGF F . The following result gives the asymptotic behavior of P_n , which we will take as a measure of asymptotic precision in estimation.

3.4 Theorem (a) *Let IGF F be regularly integrable and let Assumption 2.2 hold. If $I(\iota F) = 0$, then*

$$P_n(F) \rightarrow_d \frac{4I_s(\overline{\iota F})V(1)^2}{\sigma^2 I_s(F)}. \quad (17)$$

If $I(\iota F) \neq 0$, then

$$n^{-1/2} P_n(F) \rightarrow_d \frac{I(\iota F)^2 L(1, 0)}{I_s(F)}. \quad (18)$$

(b) *Let the IGF F be asymptotically homogeneous with limit homogeneous function H , and let Assumption 2.1 hold. Then*

$$n^{-2} P_n(F) \rightarrow_d \frac{\left(\int_0^1 H(B) B \, dr\right)^2}{\int_0^1 H(B)^2 \, dr}. \quad (19)$$

Not surprisingly, it is apparent that IV estimators in unit root regressions with nonlinear instruments have precision that depends crucially on the type of functions

being used to generate the instruments. For instruments generated by asymptotically homogeneous functions, IV precision increases at a rate of $O_p(n^2)$ according to the signal from the instrument. If regularly integrable functions are used, then IV precision is weaker with the reduced rate of divergence $O_p(n^{1/2})$, corresponding to a weaker signal. If the IGF is orthogonal to the identity function in the space $L^2(\mathbf{R})$ of square integrable functions, then precision is of stochastic order $O_p(1)$, as in the case of instrument failure mentioned earlier.

4. Optimal IV Estimators

In this section, we construct some ‘optimal’ estimators in the class of IV estimators generated by various IGF’s. Here we interpret ‘optimality’ in terms of the precision quantity $P_n(F)$ introduced above. IV estimators with regularly integrable IGF’s cannot be optimal against wider classes because they have precision that is of a smaller order of magnitude compared to that of asymptotically homogeneous IGF’s. We therefore concentrate on estimators in the IV class with asymptotically homogeneous IGF’s.

To analyze the precision of IV estimators with asymptotically homogeneous IGF’s, we first note that the representation

$$\frac{\left(\int_0^1 H(B)B dr\right)^2}{\int_0^1 H(B)^2 dr} = \frac{\left(\int_{-\infty}^{\infty} sH(s)L(1, s)ds\right)^2}{\int_{-\infty}^{\infty} H(s)^2L(1, s)ds}, \quad (20)$$

which follows immediately from the occupation times formula (3). For the purpose of our discussion, an optimal IV estimator will be defined as an IV estimator generated by an asymptotically homogeneous function with limit homogeneous function H that maximizes the asymptotic precision as given by (20). The limit homogeneous function H which yields an optimal IV estimator is given only up to scalar multiplication, since the maximization problem is invariant with respect to the scalar multiplication.

4.1 Lemma *Let Assumption 2.1 hold.*

(a) *The asymptotic precision is maximized a.s. for IV estimators with asymptotically homogeneous IGF’s having limit homogeneous function*

$$H(x) = x$$

in the class of all IV estimators with asymptotically homogeneous IGF’s.

(b) *In the restricted class of IV estimators with asymptotically homogeneous and bounded IGF’s having limit homogeneous function*

$$H(x) = a1\{x > 0\} - b1\{x < 0\},$$

asymptotic precision is maximized a.s. when

$$a^*/b^* = \frac{\int_0^{\infty} |s|L(1, s) ds}{\int_{-\infty}^0 |s|L(1, s) ds} \bigg/ \frac{\int_0^{\infty} L(1, s) ds}{\int_{-\infty}^0 L(1, s) ds}.$$

Part (a) of Lemma 4.1 shows that the OLS estimator, i.e., the IV estimator with the instrument generated by the linear IGF $F(x) = x$, is optimal in the sense that it has highest precision as we have defined it. This is so, irrespective of the actual realization of the local time $L(1, \cdot)$ and is a result that may be expected given our definition of precision. As is well known, the OLS estimator is the optimal IV estimator for the classical regression with homoskedastic errors when we condition on the regressors. Our definition of precision is based on the conditional variance of the (score) component $\sum_{t=1}^n F(y_{t-1}) u_t$, which has the same form as it does in the classical regression model with fixed regressors. Accordingly, the asymptotic precision of the IV estimator with $H(x) = x$ is given by

$$\int_{-\infty}^{\infty} s^2 L(1, s) ds,$$

as is immediately apparent from (20).

Part (b) of Lemma 4.1 considers IV estimation in a restricted class. In particular, the IV class is restricted to instruments generated by transformations in the class \mathcal{F}_B of *bounded functions*. From the proof of part (b) of Lemma 4.1, we may deduce that the asymptotic precision of IV estimators in this class is bounded by

$$\frac{(\int_0^{\infty} |s| L(1, s) ds)^2}{\int_0^{\infty} L(1, s) ds} + \frac{(\int_{-\infty}^0 |s| L(1, s) ds)^2}{\int_{-\infty}^0 L(1, s) ds}.$$

This bound is attained when the IGF has limit homogeneous function as given in part (b) of Lemma 4.1 is used. Note that this IGF is path-dependent. In particular, it is dependent upon the realized value of the local time $L(1, \cdot)$ via the ratio a^*/b^* . Therefore, the optimal IV estimator in this class is not feasible, although a feasible version of the procedure may be constructed by estimating this ratio.

Subsequent sections concentrate on the particular IV estimator whose IGF is asymptotically homogeneous with limit homogeneous function $H(x) = \text{sgn}(x)$. This IV estimator is the Cauchy estimator, mentioned earlier, and it is the optimal bounded estimator in the special case where $\int_0^{\infty} L(1, s) ds = \int_{-\infty}^0 L(1, s) ds$ and $\int_0^{\infty} |s| L(1, s) ds = \int_{-\infty}^0 |s| L(1, s) ds$. These conditions hold on average since the distribution of $L(1, \cdot)$ is symmetric about origin, i.e., the distributions of $L(1, x)$ and $L(1, -x)$ are the same for all $x \in \mathbf{R}$. Therefore, this IV estimator may be regarded as an optimal bounded estimator which does not use any additional sample information. Clearly, the function $F(x) = \text{sgn}(x)$ is an asymptotically homogeneous function having itself as the limit homogeneous function. Therefore, the Cauchy estimator, the IV estimator with instrument $\text{sgn}(y_{t-1})$, is an optimal IV in this sense. Any other asymptotically homogeneous functions with limit homogeneous function $H(x) = \text{sgn}(x)$ also yield optimal IV estimators. Some obvious choices include functions such as $F(x) = \Phi(x) - 1/2$ with any distribution function Φ , or $F(x) = \arctan(x)$, among many others. In what follows, we call *asymptotic sign functions* all asymptotically homogeneous functions whose limit homogeneous function is $H(x) = \text{sgn}(x)$.

The bounded optimal IV estimator with asymptotic sign IGF has some nice properties that the conventional OLS estimator does not have. The estimator yields a t-ratio that has a standard normal limit distribution when $\alpha = 1$, as well as when $|\alpha| < 1$. This enables us to construct and interpret the confidence interval for α in a conventional way. On the other hand, of course, the t-ratio based on the OLS estimator has a limit normal distribution only when $|\alpha| < 1$ and its limit distribution is non-Gaussian when $\alpha = 1$. The fact that the t-ratio based on the Cauchy estimator has a limit normal distribution has implications for tests of a unit root. These properties are explored in So and Shin (1999), where the Cauchy estimator was first suggested. We provide a brief overview in the next section.

5. Unit Root Tests and Confidence Intervals

Testing for a unit root in (1) can be based on the IV t-ratio

$$Z_{IV} = \frac{\hat{\alpha} - 1}{s(\hat{\alpha})} \quad (21)$$

where $s(\hat{\alpha})$ is the standard error of the IV estimator $\hat{\alpha}$ given by

$$\begin{aligned} s(\hat{\alpha})^2 &= \hat{\sigma}^2 \left[\left(\sum_{t=1}^n F(y_{t-1})y_{t-1} \right) \left(\sum_{t=1}^n F(y_{t-1})^2 \right)^{-1} \left(\sum_{t=1}^n y_{t-1}F(y_{t-1}) \right) \right]^{-1} \\ &= \hat{\sigma}^2 \frac{\sum_{t=1}^n F(y_{t-1})^2}{\left(\sum_{t=1}^n F(y_{t-1})y_{t-1} \right)^2}. \end{aligned} \quad (22)$$

with the usual error variance estimator $\hat{\sigma}^2 = \sum_{t=1}^n \hat{u}_t^2$, with $\hat{u}_t = y_t - \hat{\alpha}y_{t-1}$. It is natural here to use the IV estimate $\hat{\alpha}$ in fitting the residual \hat{u}_t . However, other estimates of α can be used as long as they yield consistent estimates of σ^2 .

5.1 Theorem (a) *If the IGF is regularly integrable, then*

$$Z_{IV} \rightarrow_d \mathbf{N}(0, 1)$$

under Assumption 2.2.

(b) *If the IGF is asymptotically homogeneous, then*

$$Z_{IV} \rightarrow_d \frac{\int_0^1 H(B) dB}{\sigma \left(\int_0^1 H(B)^2 dr \right)^{1/2}}$$

under Assumption 2.1.

The limiting distribution of the t-ratio Z_{IV} is standard normal for a regularly integrable IGF but not generally so when the IGF is an asymptotically homogeneous function. However, there is an important exception. The IV estimator with the asymptotic sign IGF is distributed as normal asymptotically. This is the only exception,

and no other IV estimator with asymptotically homogeneous IGF has asymptotic normality. The following corollary extends the asymptotic normality obtained for the Cauchy estimator by So and Shin (1999) to all IV estimators with the asymptotic sign IGF.

5.2 Corollary *Let Assumption 2.1 hold. Then we have*

$$Z_{IV} \rightarrow_d \mathbf{N}(0, 1)$$

when the IGF is the asymptotic sign function.

Thus, t-ratios based on IV estimators with regularly integrable and asymptotic sign IGF's are asymptotically standard normal under the null. This is in contrast to the usual unit root distribution of the OLS t-ratio, which is asymmetric and skewed to the left (c.f Fuller, 1996). The standard normal IV t-ratios are convenient and useful in unit root testing as discussed in So and Shin (1999).

First, under the alternative of stationarity, i.e., $|\alpha| < 1$, we have

$$n^{1/2}(\hat{\alpha} - \alpha) \rightarrow_d \mathbf{N}(0, v^2),$$

where

$$v^2 = \text{plim}_{n \rightarrow \infty} \frac{n^{-1} \sum_{t=1}^n F(y_{t-1})^2}{(n^{-1} \sum_{t=1}^n F(y_{t-1})y_{t-1})^2}.$$

Consequently, if we let

$$Z_{IV}(\alpha) = \frac{\hat{\alpha} - \alpha}{s(\hat{\alpha})}, \quad (23)$$

where $s(\hat{\alpha})$ is defined in (22), then

$$Z_{IV}(\alpha) \rightarrow_d \mathbf{N}(0, 1). \quad (24)$$

Therefore, the IV t-ratios are normally distributed asymptotically with any choice of IGF, for all $|\alpha| < 1$.

The asymptotic normality in (24) also applies to the standard t-ratio based on the OLS estimator for all $|\alpha| < 1$. However, when $\alpha = 1$, the unit root limit distribution is non-Gaussian and leans towards the alternative distributions that apply under stationarity which adversely affects the power of the unit root test. In contrast, the IV t-ratios considered here have no such aspect and there is no discontinuity in the asymptotics, since they are asymptotically normal and symmetric around the true value of α under both the null and alternative hypothesis. It is therefore reasonable to expect that they may have some advantage over conventional OLS based procedures that rely on unit root limit theory. The simulations in So and Shin (1999) support this view and they further indicate that confidence intervals constructed from the Cauchy estimate and its limit distribution have shorter average lengths than those based on the median unbiased estimates suggested in Andrews (1993).

For testing a unit root in time series with nonzero mean or deterministic trend, we may use demeaned or detrended y_t . In this event, as mentioned earlier, we must

use recursive demeaning or detrending to maintain the martingale property and the subsequent Gaussian limit theory for the IV t-ratios. In the case of a fitted mean, this implies using

$$y_t - \frac{1}{t-1} \sum_{i=1}^{t-1} y_i \quad \text{and} \quad y_{t-1} - \frac{1}{t-1} \sum_{i=1}^{t-1} y_i \quad (25)$$

in place of y_t and y_{t-1} , respectively, for $t \geq 2$. In (25), both y_t and y_{t-1} are demeaned using the lagged recursive mean μ_{t-1} , as in So and Shin (1999) - see also Moon and Phillips (2000), where both forward-recursive and backward-recursive detrending is used. In the presence of a linear time trend, we may use

$$\begin{aligned} y_t - \frac{1}{n-t+1}(y_n - y_{t-1}) + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} iy_i \\ y_{t-1} + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} iy_i \end{aligned} \quad (26)$$

respectively for y_t and y_{t-1} , $t \geq 2$. It is not difficult to see that the recursive detrending in (26) indeed effectively detrends both y_t and y_{t-1} . Also, the recursively demeaned or detrended y_t is a martingale with respect to the filtration \mathcal{F}_t introduced in Assumption 2.1. By contrast, the usual demeaning formula $y_t - (1/n) \sum_{i=1}^n y_i$ yields a time series that is non-martingale. Usual detrending has a similar problem.

With the recursive demeaning and detrending introduced above, the previous results continue to apply for demeaned and detrended y_t . That is, if we denote by \bar{Z}_{IV} the IV t-ratio defined from recursively demeaned or detrended y_t in the same way as Z_{IV} in (21), then we have the following result.

5.3 Corollary *If the IGF is a regularly integrable function or the asymptotic sign function, then we have*

$$\bar{Z}_{IV} \rightarrow_d \mathbf{N}(0, 1)$$

under Assumption 2.2 or 2.1 respectively.

Continuity of the distribution across α of the t-ratio $Z_{IV}(\alpha)$ defined in (23) also allows us to construct confidence intervals for α in a convenient way from IV estimators. Again the approach follows So and Shin (1999). Since $Z_{IV}(\alpha) \rightarrow_d \mathbf{N}(0, 1)$ for all values of α including unity, when the IGF is a regularly integrable function or the asymptotic sign function, we can construct 100 $(1 - \tau)$ % confidence limits for α as

$$\left[\hat{\alpha} - z_{\tau/2} s(\hat{\alpha}), \quad \hat{\alpha} + z_{\tau/2} s(\hat{\alpha}) \right] \quad (27)$$

where $z_{\tau/2}$ is the two-tailed 100 τ % percentage point of the standard normal distribution. Obviously, the IV estimator with highest precision has the shortest confidence interval in (27).

6. Simulations

We report the outcome of a simulation exercise investigating the finite sample performances of various IV estimators and IV based t-ratios for the autoregression (1). For each set of simulations, the innovations u_t were drawn as zero mean independent normals with sample sizes $n = 100$ and 500 and with 10,000 replications. The variance of the innovations is set $\sigma^2 = 0.1$. Note that our IV method is not invariant with respect to σ^2 . However, it only has a scaling effect on the IGF's. It is indeed not difficult to see that the results of our simulations would remain unchanged for different values of σ^2 if we adjust the scales of the IGF's accordingly. For testing $H_0 : \alpha = 1$ against stationary alternatives, we explore the size of the IV based t-ratios in relation to their nominal size from the $\mathbf{N}(0, 1)$ limit distribution, and report some power comparisons against the Dickey-Fuller (DF) test. We also examine the coverage probabilities and the average lengths of the confidence intervals of α computed from the IV based t-ratios.

For estimation of α , we took six IV estimators generated by various bounded IGF's. Three of the IV estimators were constructed from IGF's that are asymptotically equivalent to the sign function, and the rest from integrable IGF's. The table below lists these IV estimators and the associated IGF's. We denote by IVh the IV estimators for which the IGF is asymptotically the sign function, and by IVi those corresponding to integrable IGF's. The truncation parameter K appearing in the definitions of the instruments for IVi1, IVi2 and IVh2 are set at 2 for the sample size $n = 100$ and at 3 for $n = 500$. The shapes of these IV generating functions are pictured in Figure 1.

IV Estimators	Instrument Generating Functions
IVh1	$\text{sgn}(x)$
IVh2	$x 1\{ x \leq K\} + \text{sgn}(x)K 1\{ x > K\}$
IVh3	$\arctan(x)$
IVi1	$\text{sgn}(x) 1\{ x \leq K\}$
IVi2	$x 1\{ x \leq K\}$
IVi3	$xe^{- x }$

Figures 2–5 present kernel density estimates of the finite sample distributions of the OLS and these six IV estimators of the centered unit root coefficient and the null distributions of the t-ratios based on those estimators. The density estimates from the simulations with the sample sizes $n = 100$ and $n = 500$ for the model without an intercept are given in Figures 2 and 3; and those for the model with an intercept are given in Figures 4 and 5.

Tables 1 and 2 present for the models without and with an intercept, respectively, the biases, variances and mean-squared-error's (MSE's) of the OLS and the IV estimators. Table 3 reports the finite sample sizes and the local powers of the DF test and the IV based t-ratios. The nominal sizes of the tests are set at 5%. The

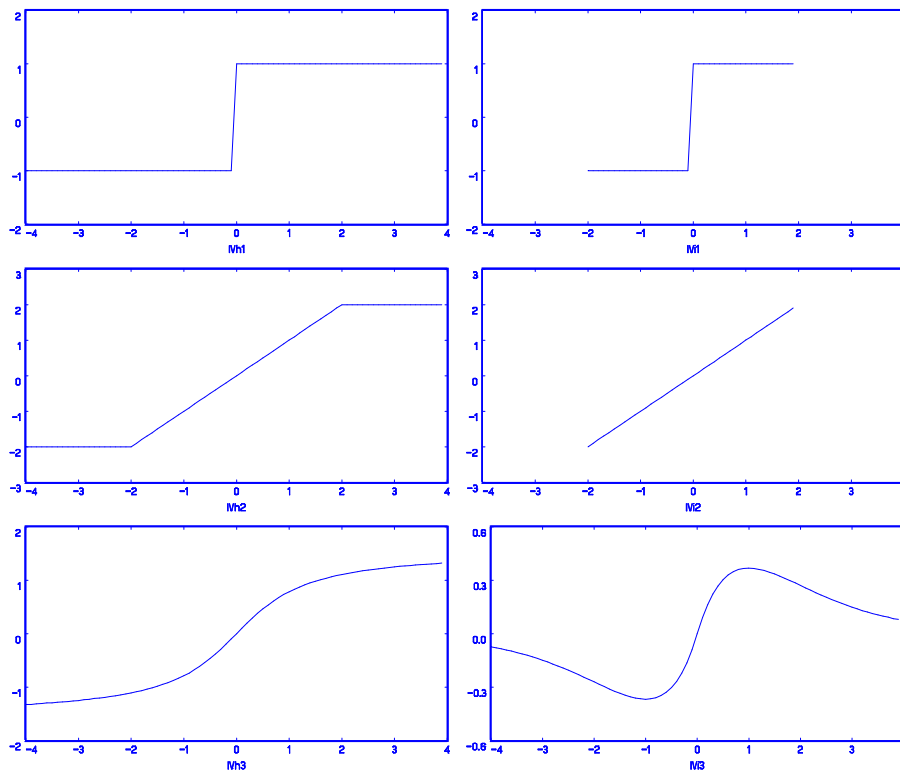


Figure 1: Functional Forms of Various IGF's

critical values for the DF tests are obtained from Fuller (1985): -1.95 and -2.86 for the models without and with an intercept, respectively. The critical values for the IV based unit root tests are all set at -1.645 from their standard normal limit distributions.

For the IV based t-ratios, we also construct 90% confidence intervals (CI's) for α according to (27) for various values of α , and compute their average lengths and their empirical coverage probabilities (CP's). The confidence intervals for the DF tests are not computed due to the discontinuity of their limit distribution at the null. The percent CP's and the average length (a.l.) of the CI's are computed and are listed in Table 4. For the purpose of comparison, we also list in Table 4 the percent CP's and the average length of the CI's for the exactly median unbiased (EMU) estimator of Andrews (1993) for the model with an intercept for $n = 100$.

The finite sample performance of the IV estimators is largely as expected from the limit theory. The OLS estimator for the unit root coefficient is known to converge at the n rate to a non-Gaussian limit distribution that is skewed to the right, but is an optimal IV estimator in the unrestricted class of IV generating functions. The IVh estimators converge at the same convergence rate n , but their asymptotic distri-

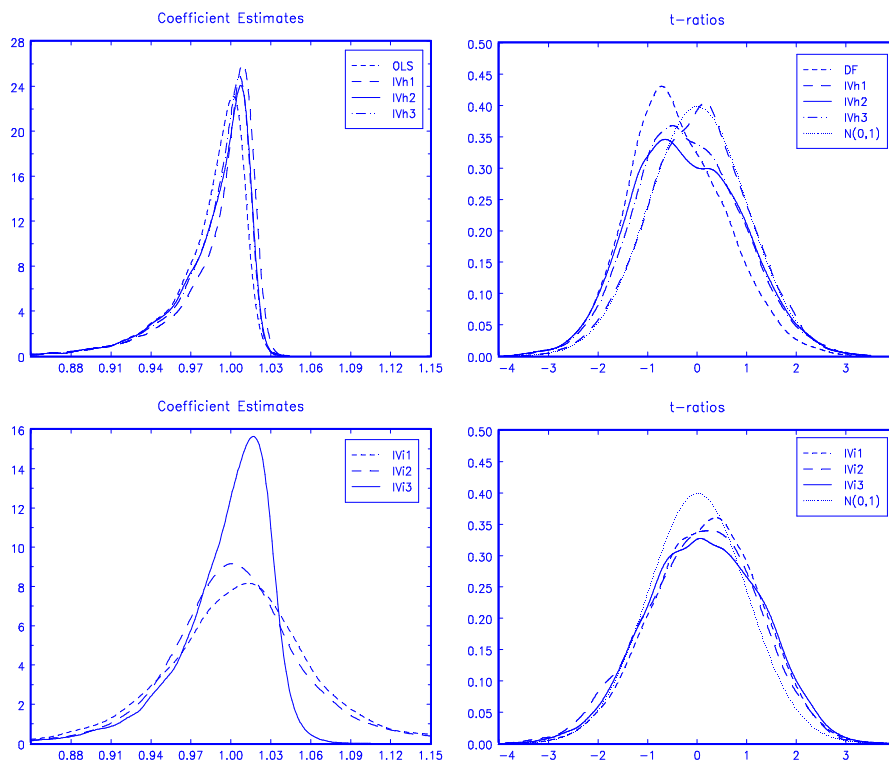


Figure 2: Densities for Models without Constant, $n = 100$

butions are contrastingly mixed normal. The IV_i estimators also have mixed normal limit distributions. However, they have slower $n^{1/4}$ convergence rates and lower their precision rates than the IV_h 's, and thus are suboptimal.

From Figures 2–5, we can see that the empirical distributions of the OLS and the IV_h estimators do look nonstandard, but are more concentrated than those of the IV_i 's, reflecting their faster rates of convergence. The finite sample distributions of OLS are skewed with a long left tail, and skewness becomes more apparent in the model with an intercept. The density estimates of the IV_i 's, though more dispersed, are better centered and more symmetric around zero. Tables 1 and 2 confirm these observations. The OLS and IV_h estimators have larger finite sample variations, but have noticeably smaller variation than the IV_i 's. Of the OLS and the IV_h 's, the IV_h 's seem less biased, but have larger finite sample variation than OLS. For the model with an intercept, the bias reduction in IV_h 's is large enough to offset the larger variances, and consequently their MSE's are smaller than those of OLS. Among the IV_h 's, IV_{h1} (the Cauchy estimator) has much smaller bias but has significantly larger variation and MSE than IV_{h2} and IV_{h3} .

The simulation study for the sampling behavior of the t-ratios also confirms our

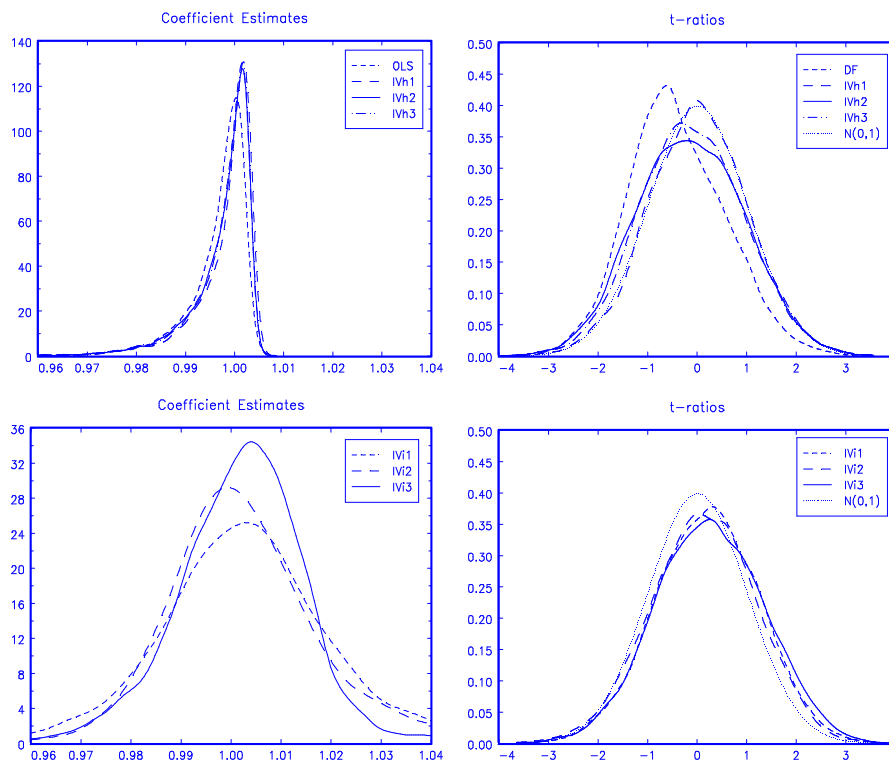


Figure 3: Densities for Models without Constant, $n = 500$

theoretical findings. The limit distribution of the DF test is known to be non-Gaussian and skewed toward the stationary distributions. In contrast, the limiting distributions of the IV based t-ratios are standard normal. Thus, we expect that the finite sample sizes of the IV based unit root tests, computed by using critical values from the standard normal distribution, to be close to nominal size, and to be more powerful than the DF test, at least when the true value of α is near unity.

As is apparent from Figures 2–5, the empirical distributions of the DF tests are skewed to the left, and the skewness is much more serious in the model with a fitted intercept. The finite sample distributions of all of our IV based t-ratios, however, are much better centered and are symmetric around zero, approximating reasonably well their limit $\mathbf{N}(0, 1)$ distributions. The density estimates of the IVi based t-ratios better approximate the limit $\mathbf{N}(0, 1)$ distribution, though are noticeably more dispersed than those based on IVh's. The approximations do improve as the sample size increases.

We therefore expect that the IVi based t-ratios suffer less size distortions and may be relatively more powerful in testing against local alternatives. Also expected is that the IVh based t-ratios have higher discriminatory powers with more noticeable size distortions. We can see from Table 3 that the IV based t-ratios do, in general,

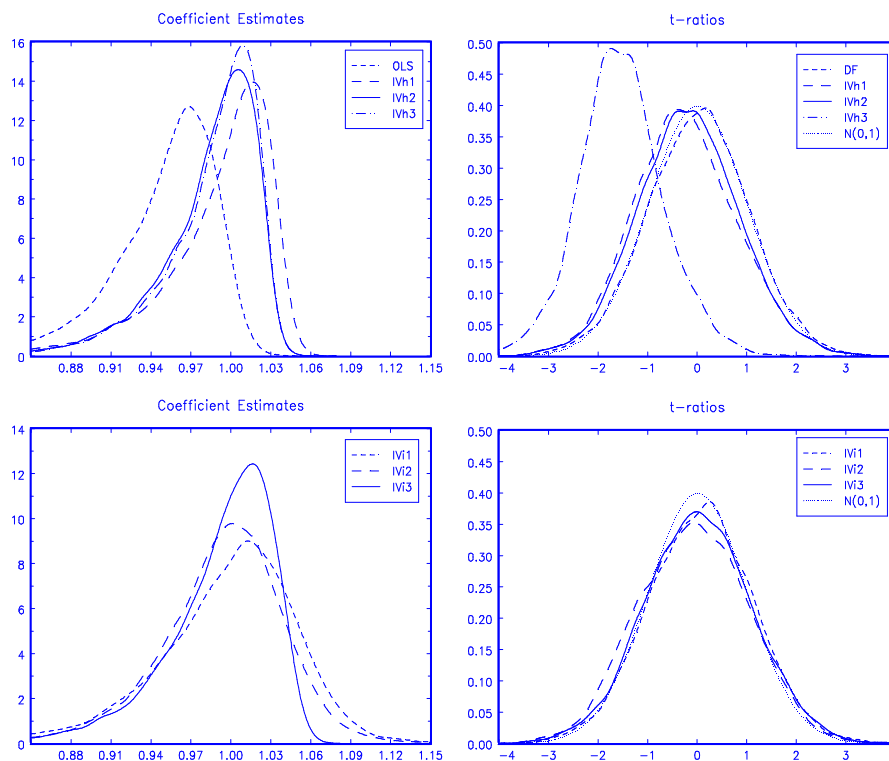


Figure 4: Densities for Models with Constant, $n = 100$

suffer from size distortions. The t-ratios based on the IVh's tend to have larger size distortions than their IVi counterparts, except for those based on the IV's constructed from the sign function. The t-ratios based on IVh1 turn out to have the least size distortions. Size distortions in all cases seem to decrease as the sample size increases.

As far as power is concerned, the t-ratios based on the IVh's are more powerful than their IVi counterparts. Compared to the DF test, the IV based t-ratios are relatively more powerful in the model with a fitted intercept, where the distributions of the DF tests are more seriously skewed toward the alternative distributions. All of the IV based t-ratios are noticeably more powerful than the DF test when α is near unity. The IVh based t-ratios, except for the IVh1 based t-ratio, remain more powerful than the DF test for all alternatives for both small and large samples. The IVi based t-ratios are also more powerful than the DF test when α is near unity, but become less powerful as α moves away from one. For the model without the constant, the IV based t-ratios also seem more powerful for α near one in small samples. However, except for the IVh2 based t-ratio, they all become less powerful than the DF tests as α moves away from one and the sample size increases. The t-ratio based on the IVh2 turns out to be slightly more powerful than the DF tests

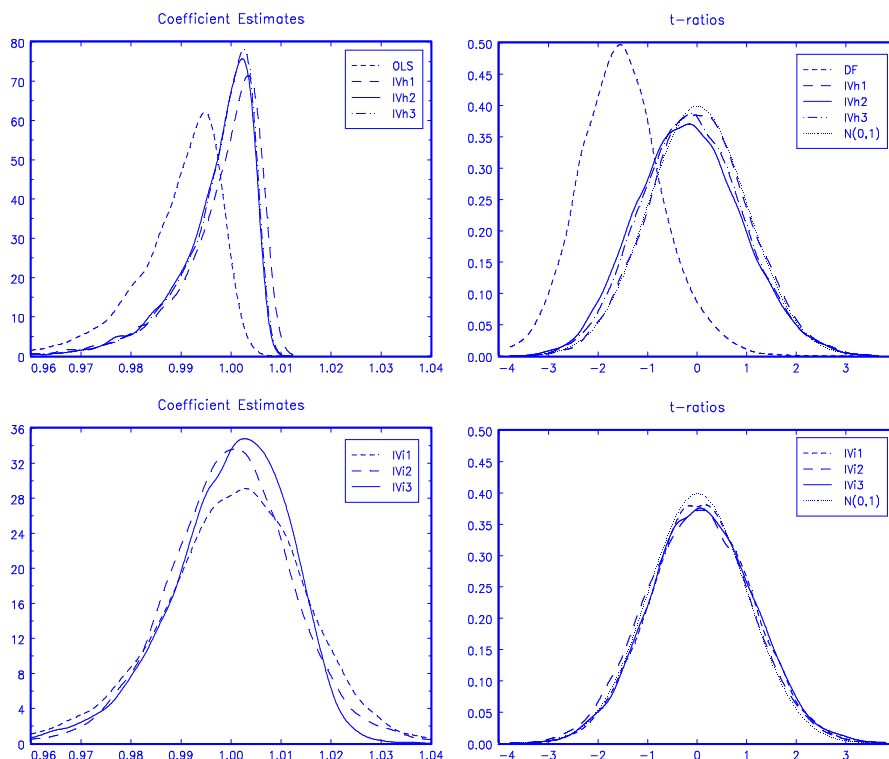


Figure 5: Densities for Models with Constant, $n = 500$

in both small and large samples for all alternatives.

As can be seen from Table 4, the empirical coverage probabilities of the CI's for α constructed from all of our IV based t-ratios quite well approximate the nominal 90% coverage for various values of α . The average lengths of the CI's from the model without the constant are somewhat shorter and their empirical CP's seem to better approximate the nominal 90% than those from the model with a constant. The empirical CP's of the CI's constructed from the IVh based t-ratios are closer to the nominal 90% than those for the CI's constructed from the IVi's when α is near unity, but the differences between two CP's become negligible as α moves away from one. The average lengths of the IVh based CI's are shorter than those of the IVi based CI's when α is near unity, but again they become closer as α moves away from one. As the sample size increases both CP's become closer to the nominal 90% and the average lengths of the IV based CI's become shorter. The performance of our IV based CI's and their CP's are overall quite satisfactory as one can see from comparing them with the performance of the CI's based on Andrews' (1993) exactly median unbiased estimator using Table 4. Our IV based CI's have slightly smaller CP's, while the CI based on Andrews' (1993) EMU estimator has slightly bigger CP's than the nominal

CP. However, the average lengths of our IV based CI's are in general shorter than that of the EMU based CI's.

7. Conclusions

This paper introduces a class of new autoregressive coefficient estimators that have some good asymptotic and finite sample properties. The estimators are asymptotically median unbiased and can be used to produce simple symmetric confidence intervals. There are no problems of discontinuity in the confidence intervals in the transition from stationary to nonstationary cases. The optimal IV estimator is shown to be the Cauchy estimator for which the instrument is simply $\text{sgn}(y_{t-1})$. These good properties confirm earlier findings by So and Shin (1999) where the Cauchy estimator was first suggested.

8. Technical Appendix and Proofs

8.1 Proof of Lemma 3.1 The results in (6) - (8) are shown in Park and Phillips (1999, 2001). To show the result in (9), note that

$$n^{-1/4} \sum_{t=1}^n F(y_t)y_t =_d n^{3/4} \int_0^1 \iota F \left(n^{1/2} B_n(r) \right) dr,$$

which can be approximated by

$$S_n = n^{3/4} \int_0^1 \iota F \left(n^{1/2} B(r) \right) dr + o_{a.s.}(1),$$

as shown in Phillips and Park (1998).

Let

$$M_n = -n^{1/4} \int_0^1 \overline{\iota F} \left(n^{1/2} B(r) \right) dB(r).$$

Then it follows from Ito's formula that

$$S_n = (2/\sigma^2)M_n + o_{a.s.}(1).$$

The stated result can now be deduced, since

$$[M_n] = n^{1/2}\sigma^2 \int_0^1 \overline{\iota F} \left(n^{1/2} B(r) \right)^2 dr \rightarrow_{a.s.} \sigma^2 I_s(\overline{\iota F})L(1, 0),$$

where $[M_n]$ is the quadratic variation of M_n , and

$$M_n \rightarrow_d V \left(\sigma^2 I_s(\overline{\iota F})L(1, 0) \right),$$

as shown in Park and Phillips (1999, 2001)

To see that the Brownian motions U and V defined respectively in (6) and (8) have the given correlation coefficient, we let

$$N_n = n^{1/4} \int_0^1 F(n^{1/2}B(r)) dB(r).$$

As shown in Park and Phillips (1999, 2001), N_n has the same limiting distribution as

$$n^{-1/4} \sum_{t=1}^n F(y_{t-1})u_t.$$

The quadratic covariation between M_n and N_n is

$$[M_n, N_n] = -n^{1/2}\sigma^2 \int_0^1 (F\overline{tF}) (n^{1/2}B(r)) dr \rightarrow_{a.s.} -\sigma^2 I(F\overline{tF})L(1, 0),$$

and then the stated result follows by straightforward calculation.

8.2 Proof of Lemma 3.2 For the proofs of the results in (10) - (12), see Park and Phillips (1999, 2001).

8.3 Proof of Theorem 3.3 We may easily deduce (13) from (6) and (9). The result in (14) follows similarly from (6) and (7). Likewise, (15) is immediate from (10) and (11).

8.4 Proof of Theorem 3.4 The stated results follow readily from Lemma 3.1. The result in (17) can be deduced from (8) and (9). Similarly, (18) follows from (7) and (8). Finally, we may get (19) from (11) and (12).

8.5 Proof of Lemma 4.1 Part (a) is easily deduced from the Cauchy-Schwarz inequality

$$\left(\int fg d\mu \right)^2 \leq \left(\int f^2 d\mu \right) \left(\int g^2 d\mu \right).$$

Set $f(x) = H(x)$, $g(x) = x$, define the measure

$$d\mu = L(1, s) ds$$

pathwise, and apply the Cauchy-Schwarz inequality to get

$$\frac{\left(\int_{-\infty}^{\infty} sH(s)L(1, s)ds \right)^2}{\int_{-\infty}^{\infty} H(s)^2L(1, s)ds} \leq \int_{-\infty}^{\infty} s^2L(1, s)ds.$$

Equality holds iff

$$H(x) = x$$

a.s. with respect to the measure $L(1, s)ds$, up to scalar multiplication. The stated result follows immediately.

For the proof of part (b), we let

$$H(x) = a1\{x > 0\} - b1\{x < 0\}.$$

The limit homogeneous functions of bounded asymptotically homogeneous functions can be represented in this way. Obviously, we may let $a, b > 0$. Moreover, since the maximization of (20) is invariant with respect to scalar multiplication, we may set

$$\int_{-\infty}^{\infty} H(s)^2 L(1, s) ds = a^2 \int_0^{\infty} L(1, s) ds + b^2 \int_{-\infty}^0 L(1, s) ds = 1$$

without loss of generality. Now we simply maximize

$$\int_{-\infty}^{\infty} sH(s)L(1, s) ds = a \int_0^{\infty} |s|L(1, s) ds + b \int_{-\infty}^0 |s|L(1, s) ds$$

with respect to $a, b > 0$ to get the stated result.

For the actual derivation of the result, we consider the Lagrangian

$$\begin{aligned} \mathcal{L}(a, b, \lambda) &= a \int_0^{\infty} |s|L(1, s) ds + b \int_{-\infty}^0 |s|L(1, s) ds \\ &\quad + \lambda \left(1 - a^2 \int_0^{\infty} L(1, s) ds - b^2 \int_{-\infty}^0 L(1, s) ds \right) \end{aligned}$$

and have the first order conditions

$$\begin{aligned} \int_0^{\infty} |s|L(1, s) ds - 2\lambda a^* \int_0^{\infty} L(1, s) ds &= 0 \\ \int_{-\infty}^0 |s|L(1, s) ds - 2\lambda b^* \int_{-\infty}^0 L(1, s) ds &= 0 \end{aligned}$$

from which a^*/b^* can be easily obtained as given. It is tedious, but straightforward to check that the solution satisfies the second order condition for maximum. Moreover, it is clear that the asymptotic precision depends only upon the ratio of a and b .

8.6 Proof of Theorem 5.1

We have

$$\begin{aligned} Z_{IV} &= \frac{\hat{\alpha} - 1}{s(\hat{\alpha})} = \frac{\sum_{t=1}^n F(y_{t-1})u_t}{\sum_{t=1}^n F(y_{t-1})y_{t-1}} \bigg/ \frac{\hat{\sigma} (\sum_{t=1}^n F(y_{t-1})^2)^{1/2}}{|\sum_{t=1}^n F(y_{t-1})y_{t-1}|} \\ &= \frac{\sum_{t=1}^n F(y_{t-1})u_t}{\hat{\sigma} (\sum_{t=1}^n F(y_{t-1})^2)^{1/2}}, \end{aligned}$$

from which the stated results follow immediately, as shown in Park and Phillips (1999, 2001).

8.7 Proof of Corollary 5.2 In the particular case that $H(x) = \text{sgn}(x)$, we have

$$\frac{\int_0^1 H(B) dB}{\sigma \left(\int_0^1 H(B)^2 dr \right)^{1/2}} = \frac{\int_0^1 \text{sgn}(B) dB}{\sigma} = \int_0^1 \text{sgn}(W) dW$$

We now define a continuous martingale M_t by

$$M_t = \int_0^t \text{sgn}(W_s) dW_s.$$

Then $[M]_t = t$, and it follows directly from Levy's characterization theorem that M_t is a Brownian motion. We immediately have $M_1 =_d \mathbf{N}(0, 1)$, as was to be shown.

8.8 Proof of Corollary 5.3 It is obvious that all our previous results hold with B replaced by

$$\bar{B}(r) = B(r) - \int_0^r B(s) ds$$

when the recursive demeaning is applied. To prove the result for the case of recursive detrending, we first write

$$y_t = \mu + \nu t + y_t^0$$

where $y_t^0 = y_{t-1}^0 + u_t$. Note that

$$y_{t-1} + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i = y_{t-1}^0 + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i^0 - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i^0$$

Moreover,

$$\begin{aligned} & y_t - \frac{1}{n-t+1} (y_n - y_{t-1}) + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i \\ &= y_{t-1} + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i + u_t - \left[\frac{1}{n-t+1} (y_n - y_{t-1}) - \nu \right] \\ &= y_{t-1}^0 + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i^0 - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i^0 + u_t - \frac{1}{n-t+1} \sum_{i=t}^n u_i \end{aligned}$$

Now it is not difficult to see that all our previous results hold with B replaced by

$$\bar{B}(r) = B(r) + \frac{2}{r} \int_0^r B(s) ds - \frac{6}{r^2} \int_0^r s B(s) ds$$

in this case.

Let

$$\bar{y}_{t-1}^0 = y_{t-1}^0 + \frac{2}{t-1} \sum_{i=1}^{t-1} y_i^0 - \frac{6}{t(t-1)} \sum_{i=1}^{t-1} i y_i^0$$

and

$$v_{n,t} = \frac{1}{n-t+1} \sum_{i=t}^n u_i$$

Then we have

$$\sqrt{n}(\hat{\alpha} - 1) = P_n/Q_n$$

where

$$P_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \text{sgn}(\bar{y}_{t-1}^0) u_t + \frac{1}{\sqrt{n}} \sum_{t=1}^n \text{sgn}(\bar{y}_{t-1}^0) v_{n,t}$$

and

$$Q_n = \frac{1}{n} \sum_{t=1}^n |\bar{y}_{t-1}^0|$$

However,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \text{sgn}(\bar{y}_{t-1}^0) v_{n,t} \rightarrow_p 0$$

since

$$\frac{1}{n} \sum_{t=1}^n v_{n,t}^2 = O_p(n^{-1})$$

The stated result now follows immediately.

9. Notation

$\rightarrow_{a.s.}$	almost sure convergence	\rightarrow_p	convergence in probability
\rightarrow_d	weak convergence	$=_d$	distributional equivalence
$o_{a.s.}(1)$	tends to zero almost surely	$[\cdot]$	integer part of

10. References

- Cauchy, A. (1836). "On a new formula for solving the problem of interpolation in a manner applicable to physical investigations." *Philosophical Magazine*, 8, 459-468.
- Hannan E. J. and M. Deistler (1988). *The statistical theory of linear systems*. New York: Wiley.
- Moon H. R. and P. C. B. Phillips (2000). "Estimation of Autoregressive roots near unity using panel data". *Econometric Theory*, 16, 927-997.
- Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory* **15**:269-298.
- Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear regressions with integrated time series," *Econometrica* **69**: 1452-1498.

- Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* 57, 99–125.
- Phillips, P. C. B. and J. Y. Park (1998). "Nonstationary density estimation and kernel autoregression," Cowles Foundation Discussion Paper #1181, Yale University.
- So, B.S. and D.W. Shin (1999). "Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals," *Econometric Theory* **15**:165-176.

Table 1. Biases, Variances and MSE's: Model without Constant

	$n=100$				$n=500$			
	α	$n \times \text{bias}$	$n \times \text{var}$	$n \times \text{MSE}$	α	$n \times \text{bias}$	$n \times \text{var}$	$n \times \text{MSE}$
DF	1.000	-1.694	0.053	0.082	1.000	-1.750	0.011	0.017
	0.990	-1.758	0.070	0.101	0.990	-2.021	0.030	0.038
	0.975	-1.821	0.097	0.130	0.980	-1.906	0.049	0.056
	0.950	-1.867	0.143	0.178	0.970	-1.916	0.069	0.076
	0.900	-1.750	0.230	0.261	0.960	-1.980	0.088	0.096
IVh1	1.000	-1.082	0.080	0.092	1.000	-1.104	0.016	0.019
	0.990	-1.081	0.106	0.117	0.990	-1.186	0.046	0.049
	0.975	-1.094	0.147	0.159	0.980	-1.146	0.076	0.078
	0.950	-1.176	0.220	0.233	0.970	-1.140	0.106	0.109
	0.900	-0.947	0.356	0.365	0.960	-1.232	0.137	0.140
IVh2	1.000	-1.435	0.054	0.075	1.000	-1.337	0.011	0.015
	0.990	-1.542	0.071	0.095	0.990	-1.753	0.031	0.037
	0.975	-1.675	0.098	0.126	0.980	-1.768	0.049	0.056
	0.950	-1.796	0.143	0.176	0.970	-1.848	0.069	0.075
	0.900	-1.736	0.230	0.260	0.960	-1.948	0.088	0.096
IVh3	1.000	-1.369	0.055	0.074	1.000	-1.277	0.012	0.015
	0.990	-1.424	0.073	0.093	0.990	-1.505	0.033	0.037
	0.975	-1.498	0.100	0.123	0.980	-1.434	0.052	0.056
	0.950	-1.580	0.147	0.172	0.970	-1.490	0.072	0.077
	0.900	-1.500	0.234	0.256	0.960	-1.590	0.092	0.097
IVi1	1.000	0.977	0.303	0.312	1.000	2.570	0.183	0.197
	0.990	0.204	0.233	0.233	0.990	-0.420	0.082	0.082
	0.975	-0.464	0.212	0.214	0.980	-0.816	0.094	0.095
	0.950	-0.901	0.249	0.258	0.970	-0.962	0.116	0.118
	0.900	-0.892	0.364	0.372	0.960	-1.149	0.142	0.144
IVi2	1.000	0.729	0.202	0.207	1.000	2.265	0.127	0.137
	0.990	-0.123	0.157	0.158	0.990	-0.763	0.057	0.058
	0.975	-0.871	0.143	0.151	0.980	-1.203	0.063	0.066
	0.950	-1.340	0.165	0.183	0.970	-1.465	0.076	0.080
	0.900	-1.608	0.236	0.262	0.960	-1.724	0.092	0.098
IVi3	1.000	-0.480	0.084	0.086	1.000	0.978	0.070	0.072
	0.990	-0.709	0.099	0.104	0.990	-0.561	0.065	0.066
	0.975	-0.927	0.126	0.135	0.980	-0.736	0.084	0.085
	0.950	-1.127	0.174	0.187	0.970	-0.906	0.104	0.105
	0.900	-1.102	0.263	0.275	0.960	-1.073	0.124	0.126

Table 2. Biases, Variances and MSE's: Model with Constant

	$n=100$				$n=500$			
	α	$n \times \text{bias}$	$n \times \text{var}$	$n \times \text{MSE}$	α	$n \times \text{bias}$	$n \times \text{var}$	$n \times \text{MSE}$
DF	1.000	-5.119	0.099	0.361	1.000	-5.345	0.021	0.078
	0.990	-5.223	0.121	0.394	0.990	-4.927	0.041	0.090
	0.975	-5.010	0.149	0.400	0.980	-4.451	0.059	0.099
	0.950	-4.608	0.190	0.402	0.970	-4.232	0.077	0.113
	0.900	-4.159	0.271	0.444	0.960	-4.215	0.096	0.132
IVh1	1.000	-1.104	0.159	0.171	1.000	-1.156	0.033	0.036
	0.990	-1.355	0.196	0.214	0.990	-1.292	0.065	0.068
	0.975	-1.430	0.242	0.262	0.980	-0.693	0.093	0.094
	0.950	-1.216	0.309	0.324	0.970	-0.275	0.123	0.123
	0.900	-0.812	0.442	0.449	0.960	-0.254	0.153	0.153
IVh2	1.000	-1.740	0.102	0.132	1.000	-1.630	0.022	0.027
	0.990	-2.010	0.125	0.165	0.990	-1.805	0.042	0.048
	0.975	-2.065	0.154	0.196	0.980	-1.350	0.059	0.063
	0.950	-1.862	0.195	0.230	0.970	-1.057	0.078	0.080
	0.900	-1.472	0.279	0.300	0.960	-0.951	0.097	0.099
IVh3	1.000	-1.553	0.104	0.128	1.000	-1.461	0.023	0.027
	0.990	-1.799	0.128	0.160	0.990	-1.537	0.044	0.049
	0.975	-1.841	0.157	0.191	0.980	-1.014	0.063	0.065
	0.950	-1.635	0.199	0.226	0.970	-0.700	0.082	0.083
	0.900	-1.259	0.282	0.298	0.960	-0.609	0.101	0.102
IVi1	1.000	-0.356	0.243	0.245	1.000	0.210	0.103	0.103
	0.990	-0.941	0.252	0.261	0.990	-0.852	0.094	0.096
	0.975	-1.157	0.278	0.291	0.980	-0.437	0.109	0.109
	0.950	-1.060	0.329	0.340	0.970	-0.127	0.131	0.131
	0.900	-0.779	0.448	0.454	0.960	-0.188	0.157	0.157
IVi2	1.000	-0.741	0.157	0.163	1.000	-0.062	0.070	0.070
	0.990	-1.363	0.163	0.181	0.990	-1.093	0.063	0.065
	0.975	-1.590	0.178	0.203	0.980	-0.842	0.071	0.073
	0.950	-1.564	0.209	0.234	0.970	-0.723	0.085	0.086
	0.900	-1.381	0.283	0.302	0.960	-0.763	0.101	0.102
IVi3	1.000	-0.886	0.134	0.141	1.000	-0.158	0.066	0.066
	0.990	-1.254	0.158	0.173	0.990	-0.844	0.079	0.081
	0.975	-1.377	0.187	0.206	0.980	-0.398	0.096	0.097
	0.950	-1.235	0.230	0.245	0.970	-0.134	0.115	0.115
	0.900	-0.927	0.315	0.323	0.960	-0.140	0.134	0.134

Table 3. Sizes and Adjusted Local Powers

Model without Constant							
$n=100$	c.v.	size	e.c.v.	$\alpha=.99$	$\alpha=.975$	$\alpha=.95$	$\alpha=.90$
DF	-1.950	0.047	-1.922	0.080	0.148	0.344	0.786
IVh1	-1.645	0.052	-1.661	0.074	0.133	0.268	0.537
IVh2	-1.645	0.088	-1.915	0.080	0.149	0.344	0.786
IVh3	-1.645	0.076	-1.845	0.080	0.148	0.340	0.774
IVi1	-1.645	0.047	-1.605	0.073	0.129	0.267	0.552
IVi2	-1.645	0.075	-1.883	0.079	0.145	0.331	0.762
IVi3	-1.645	0.057	-1.701	0.080	0.147	0.313	0.705
$n=500$	c.v.	size	e.c.v.	$\alpha=.99$	$\alpha=.98$	$\alpha=.97$	$\alpha=.96$
DF	-1.950	0.048	-1.933	0.331	0.770	0.970	0.999
IVh1	-1.645	0.048	-1.627	0.257	0.553	0.776	0.902
IVh2	-1.645	0.080	-1.869	0.334	0.773	0.971	0.999
IVh3	-1.645	0.066	-1.800	0.308	0.720	0.943	0.996
IVi1	-1.645	0.040	-1.539	0.203	0.488	0.743	0.898
IVi2	-1.645	0.052	-1.665	0.261	0.638	0.897	0.986
IVi3	-1.645	0.042	-1.530	0.232	0.520	0.786	0.933
Model with Constant							
$n=100$	c.v.	size	e.c.v.	$\alpha=.99$	$\alpha=.975$	$\alpha=.95$	$\alpha=.90$
DF	-2.860	0.054	-2.916	0.058	0.070	0.116	0.316
IVh1	-1.645	0.053	-1.673	0.080	0.122	0.206	0.426
IVh2	-1.645	0.081	-1.889	0.079	0.135	0.255	0.595
IVh3	-1.645	0.074	-1.833	0.078	0.137	0.256	0.594
IVi1	-1.645	0.052	-1.664	0.077	0.120	0.203	0.425
IVi2	-1.645	0.079	-1.885	0.078	0.132	0.252	0.589
IVi3	-1.645	0.059	-1.725	0.079	0.136	0.252	0.560
$n=500$	c.v.	size	e.c.v.	$\alpha=.99$	$\alpha=.98$	$\alpha=.97$	$\alpha=.96$
DF	-2.860	0.052	-2.878	0.122	0.296	0.593	0.848
IVh1	-1.645	0.047	-1.618	0.218	0.420	0.626	0.803
IVh2	-1.645	0.078	-1.851	0.250	0.574	0.854	0.976
IVh3	-1.645	0.064	-1.766	0.245	0.549	0.822	0.964
IVi1	-1.645	0.043	-1.574	0.193	0.383	0.602	0.796
IVi2	-1.645	0.061	-1.749	0.224	0.504	0.787	0.951
IVi3	-1.645	0.044	-1.569	0.210	0.419	0.663	0.855

Table 4: Coverage Probabilities and Average Length of CI's

Model without Constant											
		$n=100$					$n=500$				
	α	1.00	0.99	0.95	0.90	0.80	1.00	0.99	0.98	0.97	0.96
CP (%)	IVh1	89.5	89.9	89.8	89.6	89.9	89.9	90.5	89.9	90.0	89.6
	IVh2	86.6	87.3	88.9	89.9	90.3	87.0	88.7	89.1	89.8	90.3
	IVh3	88.0	88.7	88.8	89.6	90.0	88.6	89.8	89.5	89.9	90.0
	IVi1	87.2	87.9	89.5	89.5	89.8	87.6	89.4	89.7	89.9	89.5
	IVi2	85.2	86.3	87.9	89.3	90.3	87.2	88.3	89.1	89.1	89.8
	IVi3	85.2	86.3	88.8	89.1	89.7	85.2	89.0	89.8	90.2	89.3
CI (a.l.)	IVh1	.082	.098	.149	.193	.255	.016	.030	.040	.047	.054
	IVh2	.069	.081	.121	.155	.204	.014	.025	.032	.038	.043
	IVh3	.069	.082	.122	.156	.205	.014	.026	.033	.039	.044
	IVi1	.167	.151	.160	.195	.255	.057	.042	.045	.050	.055
	IVi2	.139	.126	.131	.157	.204	.048	.035	.037	.040	.044
	IVi3	.091	.099	.134	.166	.213	.037	.037	.042	.047	.051
Model with Constant											
		$n=100$					$n=500$				
	α	1.00	0.99	0.95	0.90	0.80	1.00	0.99	0.98	0.97	0.96
CP (%)	IVh1	89.1	90.5	89.2	88.4	89.1	89.8	89.7	89.2	88.6	89.6
	IVh2	87.9	89.3	89.1	88.4	89.1	87.4	88.5	88.5	89.0	89.4
	IVh3	88.6	90.0	88.9	88.4	88.6	88.9	88.9	88.9	88.5	89.1
	IVi1	88.2	89.8	89.0	88.3	89.1	89.1	89.8	89.2	88.8	89.7
	IVi2	86.5	88.2	88.0	88.2	89.1	88.1	88.6	88.6	88.5	89.3
	IVi3	87.4	89.9	88.7	88.1	88.5	88.5	89.4	89.4	88.6	88.9
	EMU	92.4	91.7	90.3	91.2	91.2					
CI (a.l.)	IVh1	.123	.139	.178	.215	.271	.025	.036	.044	.051	.057
	IVh2	.099	.111	.142	.171	.215	.020	.029	.035	.041	.045
	IVh3	.100	.113	.143	.172	.216	.021	.030	.036	.042	.046
	IVi1	.157	.161	.185	.217	.271	.046	.045	.048	.053	.058
	IVi2	.127	.130	.148	.172	.215	.038	.037	.039	.042	.046
	IVi3	.116	.127	.155	.182	.225	.037	.041	.045	.050	.054
	EMU	.155	.159	.175	.197	.236					