The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models

Charles R. Nelson* and Richard Startz

Department of Economics
University of Washington

First draft May 3, 2002
This draft February 15, 2006

JEL *classifications*: C120, C220, C300

Abstract

In many econometric models the asymptotic variance of a parameter estimate depends on the value of another structural parameter in such a way that the data contain little information about the former when the latter is close to a critical value. This paper introduces the Zero-Information-Limit-Condition (*ZILC*) to identify such models where 'weak identification' leads to spurious inference. We find that standard errors tend to be underestimated in these cases, but the size of the asymptotic *t*-test may either be too great (the intuitive case emphasized in the 'weak instrument' literature) or too small as in two cases illustrated here.

The authors gratefully acknowledge support from the Ford and Louisa Van Voorhis and the Cecil and Jane Castor Professorships, respectively, at the University of Washington. We thank the editor and referee of this volume as well as Jinyong Hahn, Andrew Harvey, Kundan Kishor, Sharon Kozicki, Jun Ma, James Nason, Les Oxley, Michael Perlman, Robert Rasche, Douglas Steigerwald, James Stock, Tao Zha, and Eric Zivot for helpful comments, but responsibility for errors is entirely the authors'. Able research assistance was provided by Jinho Bae.

* Corresponding author:
Department of Economics
Box 353330
University of Washington
Seattle, WA 98195
cnelson@u.washington.edu.

1. Introduction.

To introduce this paper, consider as an example the nonlinear regression model $y_i = \gamma \bullet (x_i + \beta \bullet z_i) + \varepsilon_i$ where β is the parameter of interest with least squares estimate $\hat{\beta}$. The model is identified if $\gamma \neq 0$ and under normality the asymptotic variance of $\hat{\beta}$ is proportional to γ^{-2}. Thus γ controls the amount of information in the data for given sample size and the data contain relatively little information about β when γ is small. In practice γ is unknown so the standard errors produced by econometric software and relied upon in empirical research are necessarily based on sample estimates. This distinction matters for inference. For example, when sample size is 100 and x, and z and ε are uncorrelated standard normal, the median reported standard error from a Monte Carlo experiment agrees closely with the asymptotic value (.10) when $\gamma = 1.0$, but is only 2.2 compared to the asymptotic value of 10 when $\gamma = 0.01$. This downward bias in estimation of the asymptotic standard deviation reflects an upward bias in estimating γ^2 that is large relative to the true value when γ is small. But too-small standard errors do not necessarily imply too-large *t*-statistics; rather the opposite occurs here. The actual size of the *t*-test at the nominal .05 level is .05 for γ =1, but is only .01 when γ =.01. Understanding why this seeming paradox occurs is one of the objectives of this paper. A key finding is that this nonlinear regression model is something more than an interesting case; indeed, a wide class of models can be recast in this format, and the analog to the correlation between x and z determines whether the *t*-test is undersized or oversized.

Other cases of spurious inference—systematically wrong standard errors and *t*-test size in finite samples—in the literature include the ‘weak instrument problem’ in *IV* estimation, and ARMA models with near parameter redundancy. These are further examples of *weak identification*: cases where the model is identified and asymptotic theory holds, but the data contain relatively little information. In the weak *IV* case, standard errors are too small, but *t*-statistics may be either too large (the outcome stressed in the literature) or too small (as shown in this paper). However, weak identification is not always associated with spurious inference, an example being multicollinear classical regression with highly but not perfectly co-linear regressors where inference is exact. This paper attempts to address the following questions: What links together models where weak identification is associated with spurious inference? Are standard errors underestimated in these cases, and why? What conditions determines whether the *t*-test is undersized or oversized?

In Section 2 we define the Zero-Information-Limit-Condition (*ZILC*) to hold when the limit of the inverse of the asymptotic variance of $\hat{\beta}$ goes to zero as an identifying parameter goes to a critical value. We argue that estimated standard errors tend to underestimate the asymptotic standard deviation in models where the *ZILC* holds, for given sample size, if identification is weak enough. We then derive a common linear representation or approximation for such models, and show that the parameter of interest may be expressed as the ratio of regression coefficients in the linear representation. It follows that *t*-tests may be oversized or undersized, in spite of underestimated standard errors, depending on the particular data generating process. Section 3 discusses three examples, the ARMA (1,1) model with near parameter redundancy and cases of nonlinear regression and weak instruments where the true null hypothesis is rarely rejected in spite of underestimated standard errors. Section 4 concludes.

That Wald-based standard errors and test statistics are problematical in weakly identified models is clear from the widely influential work of Dufour (1997). Valid confidence intervals with coverage probability $(1-\alpha)$ must be unbounded with probability $(1-\alpha)$ and Dufour proves that that for locally almost unidentified (LAU) parameters valid confidence intervals must have a non-zero probability of being unbounded. This phenomenon is illustrated in the context of weak instruments in Zivot, Startz, and Nelson (1998) where valid confidence intervals based on inversion of *LR*, *LM*, and *AR* test statistics are unbounded with increasing probability as the instrument becomes weaker, reaching $(1-\alpha)$ at the point of un-identification. Dufour shows that Wald-based intervals cannot meet this standard since they are always bounded and thus cannot be valid. Further, Dufour and Staiger and Stock (1997) show that Wald-based confidence intervals and test statistics are not pivotal since their distribution depends on unknown nuisance parameters. The simulation results in this paper provide clear examples of that dependence in contexts other than *IV*.

2. *The Zero-Information-Limit Condition (ZILC).*

2.1 *Definition and implications for estimated information.*

Consider a model with scalar parameters β and γ where β is the parameter of interest for hypothesis testing. The asymptotic variance of estimator $\hat{\beta}$ is assumed to have a representation as a function of β and γ, scale parameter σ, sample size *T*, and exogenous data **X**, and that function is denoted here by $V_{\hat{\beta}}(\beta,\gamma,\sigma,T,\mathbf{X})$. For example, in the model in the opening paragraph, for the maximum likelihood estimator one has $V_{\hat{\beta}}(\beta,\sigma,T)=\left(1/\gamma^2\right)\left(1+\beta^2\right)/T$. The inverse of the variance, sometimes called the "precision," is a natural measure of information associated with $\hat{\beta}$ that is convenient to work with, and we will use the notation:

$$I_{\hat{\beta}}(\beta,\gamma,\sigma,T,\mathbf{X})\equiv[V_{\hat{\beta}}(\beta,\gamma,\sigma,T,\mathbf{X})]^{-1}\geq 0\,. \qquad (2.1.1)$$

This paper is concerned with models and estimators where the information in $\hat{\beta}$ depends on the value of γ and diminishes toward zero as γ approaches a critical value denoted γ_0. We note that the value of γ_0 is typically implied by model specification and it is treated here as known. Thus:

Definition 1: The *Zero-Information-Limit Condition* ('*ZILC*') holds for estimator $\hat{\beta}$ if there is a value of γ, say γ_0, such that:

$$\lim_{\gamma\to\gamma_0} I_{\hat{\beta}}(\beta,\gamma,\sigma,T,\mathbf{X})=0\,. \qquad (2.1.2)$$

The main results of the paper flow from the following result:

Theorem 1: If *ZILC* holds for a given model and estimator, then to a second order approximation $I_{\hat{\beta}}(\beta,\gamma,\sigma,T,\mathbf{X})$ is proportional to the square of $(\gamma-\gamma_0)$:

$$I_{\hat{\beta}}(\beta,\gamma,\boldsymbol{\sigma},\mathbf{X}) \cong (\gamma-\gamma_0)^2 \bullet I''_{\hat{\beta}}(\beta,\gamma_0,\boldsymbol{\sigma},\mathbf{X})/2 \,. \qquad (2.1.3)$$

Proof: Note that second derivative I'' is taken with respect to γ and evaluated at $\gamma=\gamma_0$. The non-negativity of *I* implies it must be weakly concave from above in the neighborhood of the *ZILC* point γ_0 as illustrated in Figure 2.1. Assuming that *I* is continuous and twice differentiable in γ, then (i) the zero order derivative at the *ZILC* point is zero, which simply restates the *ZILC,* and (ii) the first order derivative is zero as well, following from the non-negativity of *I* implying zero must be a minimum.

In practice the true parameter values are not available for computing the asymptotic standard errors and asymptotic *t*-statistics reported by econometric software packages and relied upon in empirical research. Replacing true with estimated parameter values, and recalling that the value of γ_0 is typically implied by model specification, the theorem shows that estimated information is approximated by the product of two functions of sample statistics:

$$I_{\hat{\beta}}(\hat{\beta},\hat{\gamma},\hat{\sigma},T,\mathbf{X}) \cong (\hat{\gamma}-\gamma_0)^2 \bullet I''(\hat{\beta},\gamma_0,\hat{\sigma},T,\mathbf{X})/2 \,. \qquad (2.1.4)$$

If $\hat{\gamma}$ is unbiased, then the expectation of the first factor in (2.1.4) is given by:

$$E(\hat{\gamma}-\gamma_0)^2 = (\gamma-\gamma_0)^2 + V_{\hat{\gamma}} \,. \qquad (2.1.5)$$

Further, if γ is itself well-identified (in the sense that $V_{\hat{\gamma}}$ does not depend on γ, for example when it is a classical regression coefficient), then bias in estimating $(\gamma - \gamma_0)^2$ will be *large relative to its true value* when γ is close enough to γ_0, given sample size. This suggests a trade-off between γ and sample size, for example when $\hat{\gamma}$ is a regression coefficient with sampling variance proportional to 1/*T*, combinations of γ^2 and *T* that have a constant product will imply the same ratio of estimated value to true value of $(\gamma - \gamma_0)^2$. This ratio bias is a natural measure of how well the model is identified and is what matters for the magnitude of the asymptotic *t*-statistic:

$$t_{\hat{\beta}}{}^2 = (\hat{\beta} - \beta_0) \bullet V_{\hat{\beta}}^{-1}(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, T, \mathbf{X}) = (\hat{\beta} - \beta_0) \bullet I_{\hat{\beta}}(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, T, \mathbf{X}) . \tag{2.1.6}$$

The upward bias in estimating $(\gamma - \gamma_0)^2$ suggests that estimated *I* will be upward biased, but does not constitute a proof because the second factor in (2.1.4) is also a random variable. However, it is our *hypothesis* that in finite samples the estimated information measure will be too large - standard errors will be too small—and that this bias will become larger relative to actual information in the model the closer the identifying parameter is to the *ZILC* point. We express this conjecture as follows:

Definition 2: For a model which satisfies the *ZILC*, the *Zero-Information-Limit-Condition-Hypothesis* (‘*ZILCH*’) states that for a given sample **X**:

$$\lim_{\gamma \to \gamma_0} \left[prob\left(\frac{I_{\hat{\beta}}[\hat{\beta}, \hat{\gamma}, \hat{\sigma}, T, \mathbf{X}]}{I_{\hat{\beta}}[\beta, \gamma, \sigma, T, \mathbf{X}]} \right) > M \right] = 1; \forall\, M . \tag{2.1.7}$$

The direct implication of *ZILCH* is that one factor of the asymptotic *t*-statistic is too large. It does not follow, however, that the resulting *t*-statistic tends to be too large as well. Section 2.3 shows that the direction of bias in *t*-statistics depends on the specific data generating process and cases of negative bias are readily found.

2.2 A common linear representation for models in which ZILC holds.

In this section we ask whether there are common features shared by models in which *ZILC* holds. We show that *ZILC* places restrictions on their function form which is consistent with their having a common linear approximation or representation in which the coefficients are γ and the product $\beta\gamma$. Thus, inference for β can be thought of as estimating the ratio of regression coefficients. Further, although estimated information for $\hat{\beta}$ will be upwardly biased, the analysis shows why the *t*-statistic may either be too large or too small, and what conditions determine the direction.

We develop these results for models which have two parameters and a single equation representation of the form

$$y_i = f(\beta, \gamma, x_i) + \varepsilon_i; i = 1,..,T \ . \tag{2.2.1}$$

that can be estimated by Gauss-Newton, where ε is a random error. The linear approximation to the model takes the form:

$$\begin{aligned} y_i &\simeq f(\beta_*, \gamma_*, x_i) + (\beta - \beta_*)\bullet f_{\beta,i} + (\gamma - \gamma_*)\bullet f_{\gamma,i} + e_i \\ y_{*,i} &\simeq \beta\bullet f_{\beta,i} + \gamma\bullet f_{\gamma,i} + e_i \end{aligned} \tag{2.2.2}$$

where the first derivatives denoted by $f_{\beta,i} \equiv \dfrac{\partial f(\beta,\gamma,x_i)}{\partial\beta}; f_{\gamma,i} \equiv \dfrac{\partial f(\beta,\gamma,x_i)}{\partial\gamma}$ are evaluated at $\beta = \beta_*; \gamma = \gamma_*$, and in the second line the known terms are combined with the dependent variable so that: $y_{*,i} \equiv y_i - f(\beta_*, \gamma_*, x_i) + \beta_*\bullet f_{\beta,i} + \gamma_*\bullet f_{\gamma,i}$.

Gauss-Newton estimation iterates on the parameter values to minimize the sum of squares of residuals e_i. The estimated asymptotic variance of least squares estimates in the final iteration provides the first expression for *I* in (2.2.3) below. Equation (2.1.4) implies the second expression in (2.2.3) for *I* in models where *ZILC* holds (simplifying notation by making the *ZILC* point zero):

$$I_{\hat{\beta}} \simeq \sigma^{-2} \bullet (\sum f_\beta^2 - \frac{(\sum f_\beta f_\gamma)^2}{\sum f_\gamma^2})$$
$$I_{\hat{\beta}} \simeq \gamma^2 \bullet \frac{I_{\hat{\beta}}^{"}(\hat{\beta}, \gamma_0, \hat{\sigma}, T, \mathbf{X})}{2} \qquad (2.2.3)$$

Solving (2.2.3) for $I_{\hat{\beta}}^{"}$ we obtain:

$$I_{\hat{\beta}}^{"} \simeq 2 \bullet \sigma^{-2} \bullet \frac{(\sum f_\beta^2 - \frac{(\sum f_\beta f_\gamma)^2}{\sum f_\gamma^2})}{\gamma^2} \qquad (2.2.4)$$

Recall that the second derivative $I_{\hat{\beta}}^{"}$ from (2.1.4) was evaluated at the *ZILC* point γ_0 so it is not a function of γ. A sufficient condition for γ to drop out of (2.2.4) is that $f(.)$ and its derivatives take the following forms:

$$f(\beta, \gamma, x_i) = \gamma \bullet g(\beta, x_i)$$
$$f_\beta = \gamma \bullet g_\beta(\beta, x_i) \qquad (2.2.5)$$
$$f_\gamma = g(\beta, x_i).$$

where $g(\beta, x_i)$ is not a function of γ, and we define $g_\beta(\beta, x_i) \equiv \partial g(\beta, x_i) / \partial\beta$. Among the examples discussed in Section 3 the non-linear regression model takes this form directly, and the ARMA and IV models indirectly as well.

Rewriting (2.2.2) using (2.2.5) gives:

$$y_{*,i} \simeq \beta \bullet \gamma \bullet g_\beta(\beta, x_i) + \gamma \bullet g(\beta, x_i) + e_i$$
$$y_{*,i} \simeq \lambda \bullet g_\beta(\beta, x_i) + \gamma \bullet g(\beta, x_i) + e_i ; \text{ where } \lambda \equiv \beta\gamma \qquad (2.12)$$

It is useful to think of the first as the structural equation and the second as the reduced form, with the parameter of interest emerging as the ratio of regression coefficients:

$\beta = \lambda / \gamma$. Since β is just identified for $\gamma \neq 0$, estimation of either equation produces the same least squares estimate. This suggests that it is useful to think of inference for β as inference about the ratio of regression coefficients.

2.3 Implications of the ratio representation of $\hat{\beta}$ for size of t-statistics.

Now we use the representation of β as a ratio of regression coefficients to establish properties of its *t*-ratio. We wish to estimate $\beta = \lambda / \gamma$ where λ and γ are each well identified linear regression coefficients not subject to *ZILC*. The asymptotic variance of the estimator $\hat{\beta} = \hat{\lambda} / \hat{\gamma}$ is expressed here as $V_{\hat{\beta}}(\lambda, \gamma, \sigma, \mathbf{x})$, being a function in general of the true reduced form regression coefficients, the variance of the regression errors, and the data. Straightforward application of the ‘delta method’ for the asymptotic variance of a ratio gives the following result:

$$V_{\hat{\beta}}(\lambda, \gamma, \sigma, \mathbf{x}) = \frac{1}{\gamma^2} \bullet [V_{[\hat{\lambda} - \beta\hat{\gamma}]}(\beta, \sigma, \mathbf{x})] \tag{2.3.1}$$

Note that the second term is the variance of the linear combination of the regression coefficients $[\hat{\lambda} - \beta\hat{\gamma}]$ and is thus a function of the true β . Since the inverse, $I_{\hat{\beta}}$, is proportional to γ^2 we have:

Theorem 2: The *ZILC* holds for a ratio of regression coefficients.

The standard error produced by a general non-linear estimation routine will estimate (2.3.1) using estimated parameter values without regard to the particular null hypothesis the investigator has in mind. For the null hypothesis $\beta = \beta_0$ the result is the first expression for the *t*-statistic below:

$$t_{\hat{\beta}}^2 = \frac{(\hat{\beta}-\beta_0)^2}{V_{\hat{\beta}}(\hat{\lambda},\hat{\gamma},\hat{\sigma},\mathbf{x})}$$

$$= \left[\left(\frac{\hat{\lambda}-\beta_0\hat{\gamma}}{\hat{\gamma}}\right)^2 \Bigg/ \left(\frac{V_{[\hat{\lambda}-\beta\hat{\gamma}]}(\hat{\beta},\hat{\sigma},\mathbf{x})}{\hat{\gamma}^2}\right) \right]. \qquad (2.3.2)$$

$$= \frac{(\hat{\lambda}-\beta_0\hat{\gamma})^2}{V_{[\hat{\lambda}-\beta\hat{\gamma}]}(\hat{\beta},\hat{\sigma},\mathbf{x})}$$

Substituting using (2.3.1) in the second line, the square of $\hat{\gamma}$ appears in both numerator and denominator. Cancellation leaves the ratio of the linear combination of regression coefficients, $(\hat{\lambda}-\beta_0\hat{\gamma})$, over its estimated variance. The notation for that variance, as a function of the estimate $\hat{\beta}$, and not its value under the null hypothesis, again recognizes how standard errors are calculated in practice and reported in the literature. Of course it is possible to compute the standard error while imposing the null hypothesis; a well known example is the exact test of Anderson and Rubin (1949) in the context of *LIML*.

Now the estimated variance of $(\hat{\lambda}-\beta_0\hat{\gamma})$ may be written as follows, where C denotes covariance, and hats that V or C is evaluated using estimated parameters:

$$\hat{V}_{[\hat{\lambda}-\beta\hat{\gamma}]} = \hat{V}_{\hat{\lambda}} - 2\bullet\hat{\beta}\bullet\hat{C}_{\hat{\lambda},\hat{\gamma}} + \hat{\beta}^2\bullet\hat{V}_{\hat{\gamma}} \qquad (2.3.3)$$

This makes clear that whether the *t*-statistic tends to be 'too large' or 'too small' depends on this covariance, as well as on the distribution of $\hat{\beta}$.

To see this, consider the case where the true covariance is zero, so within sampling error:

$$\hat{V}_{(\hat{\lambda}-\beta\hat{\gamma})} \approx \hat{V}_{\hat{\lambda}} + \hat{\beta}^2\bullet\hat{V}_{\hat{\gamma}} = [\hat{V}_{\hat{\lambda}} + \beta^2\bullet\hat{V}_{\hat{\gamma}}] + (\hat{\beta}^2-\beta^2)\bullet\hat{V}_{\hat{\gamma}} \qquad (2.3.4)$$

The first term in brackets would be a proper standard error for the numerator of (2.3.2) because it evaluates (2.3.3) using the correct value of β under the null hypothesis. However, it is augmented by a second term which for the case β=0 gives:

$$t_{\hat{\beta}}^{2} \approx \frac{\hat{\lambda}^{2}}{\hat{V}_{\hat{\lambda}} + \hat{\beta}^{2} \bullet \hat{V}_{\hat{\gamma}}} \leq t_{\hat{\lambda}}^{2}. \qquad (2.3.5)$$

Since the additional term in denominator is always greater than zero, and indeed will be large if β is poorly estimated, the *t*-statistic will be smaller in absolute value than the *t*-statistic for λ. The size of the *t*-test for β will therefore be too small, in spite of the fact that the standard error of $\hat{\beta}$ is too small. This counterintuitive conjunction of undersized *t*-tests and underestimated standard errors occurs in two of the practical examples considered in the next section.

The case of strong correlation between the two regression coefficients is also an important one in practice. If the correlation is perfect then (2.3.3) becomes

$$\hat{V}_{[\hat{\lambda}-\beta\hat{\gamma}]} = \left(\sqrt{\hat{V}_{\hat{\lambda}}} - \hat{\beta} \bullet \sqrt{\hat{V}_{\hat{\gamma}}}\right)^{2} = \left(\sqrt{\hat{V}_{\hat{\lambda}}} - q \bullet \sqrt{\hat{V}_{\hat{\gamma}}}\right)^{2} = \left(\sqrt{\hat{V}_{\hat{\lambda}}} - q \bullet \left(\sqrt{\hat{V}_{\hat{\lambda}}} \Big/ q\right)\right)^{2} = 0 \qquad (2.3.6)$$

where perfect (positive) correlation between the regression coefficients implies that $\hat{\beta} = \hat{\lambda} / \hat{\gamma} = q$ is a positive constant and $\sqrt{\hat{V}_{\hat{\gamma}}} = \sqrt{\hat{V}_{\hat{\lambda}}} \Big/ q$. Thus the expression in parenthesis in (2.3.6) approaches zero when the regression coefficients are very highly correlated. An important practical example of positive correlation is *IV* with strong endogeneity studied by Nelson and Startz (1990a, b) who found that estimated standard errors are too small, *t*-statistics are too large, and rejections too frequent.

3. The Zero-Information-Limit-Condition in three weakly identified models.

3.1. The ARMA (1,1) model with near cancellation.

Ansley and Newbold (1980) reported that confidence intervals in ARMA models are too narrow when there is near parameter redundancy. To see whether this may be related to *ZILCH* we focus on the ARMA (1,1) model which may be written:

$$\begin{aligned}&(1-\phi L)y_t = (1-\theta L)\varepsilon_t;\ t = 1,..,T;\\ &\varepsilon_t \sim \text{i.i.d. } N(0,\sigma_\varepsilon^2), \quad |\phi|\langle 1;\ |\theta|\langle 1\end{aligned} \quad . \tag{3.1.1}$$

We impose the assumptions of stationarity and invertibility, and note that the AR and MA coefficients are identified if $\phi \neq \theta$, with asymptotic covariance matrix:

$$V_{\hat{\phi},\hat{\theta}}(\phi,\theta) = T^{-1}\frac{(1-\phi\theta)}{(\theta-\phi)^2}\begin{bmatrix}(1-\phi^2)(1-\phi\theta) & (1-\phi^2)(1-\theta^2)\\ (1-\phi^2)(1-\theta^2) & (1-\theta^2)(1-\phi\theta)\end{bmatrix}, \tag{3.1.2}$$

Note that when the absolute difference between ϕ and θ is small, sampling variance is large so identification is weak. A common situation in practice is testing for the order of the model and assuming it is the MA order that is of interest, the null hypothesis is that θ=0. Re-parameterize the model in terms of θ and the difference $\gamma=\phi-\theta$ and one has:

$$(1-(\theta+\gamma)L)y_t = (1-\theta L)\varepsilon_t\,. \tag{3.1.3}$$

The covariance matrix $V_{\hat{\theta},\hat{\gamma}}(\theta,\gamma)$ for this model is easily obtained by rearranging (3.1.2), and inverting the element for θ gives the information measure:

$$I_{\hat{\theta}}(\theta,\gamma) = \frac{\gamma^2 T}{(1-\theta^2-\theta\gamma)^2(1-\theta^2)}. \tag{3.1.4}$$

Since this function approaches zero as γ as goes to zero, *ZILC* holds for the ARMA (1,1) model, γ being the identifying parameter with *ZILC* point $\gamma_0 = 0$. We note that the 'inverted' or pure AR form of the model takes the form implied by (2.2.5), the product of the identifying parameter and a function of the MA parameter only:

$$y_t = \gamma \bullet [(1-\theta L)^{-1} y_{t-1}] + \varepsilon_t \,.$$

Estimated information $I_{\hat{\theta}}(\hat{\theta},\hat{\gamma})$ depends primarily on the estimate of γ^2 since (3.1.4) is relatively insensitive to θ, and when $\theta = 0$ it reduces to $\gamma^2 T$. The bias in estimating γ^2 is the variance of $\hat{\gamma}$ given by:

$$V_{\hat{\gamma}}(\theta,\gamma) = T^{-1}[1-(\theta^2+\theta\gamma)^2] \cong T^{-1}(1-\theta^4)\,, \tag{3.1.5}$$

Thus the upward bias in estimating $I_{\hat{\theta}}$ will be large if γ^2 is small relative to T^{-1}. For example, with T = 1,000, $\theta = 0$, and $\gamma = .01$ we have:

$$E(\hat{\gamma}^2) = \gamma^2 + V(\hat{\gamma}) = \gamma^2 + T^{-1} = .0001 + .001 = .0011 \tag{3.1.6}$$

which is an order of magnitude too large in this case.

To calibrate spurious inference in the ARMA(1,1) we ran the Monte Carlo (MC) experiments reported in Table 3.1.1 The data generating process is AR (1) and the models fitted are both AR(1), which is correctly specified and well identified, and ARMA(1,1) which is also correctly specified but is weakly identified if the difference $\gamma = \phi - \theta$ is small. Since the true value of θ is zero, $\gamma (= \phi)$ is both the identifying parameter and the AR coefficient, which we set at values of .01, .05, .10 and .20. Series length T is 1,000, and the number of replications is 1,000. Estimation is done within EViews™. Individual replications were discarded if either stationarity or invertibility was violated; this occurred twice with γ values of .01 and .05 only. Standard errors computed by EViews™

come out of its non-linear estimation algorithm. Alternatively, plugging coefficient estimates into the asymptotic formula, we find little difference in the resulting *t*-tests.

Results for estimating the AR(1) model confirmed what is well known: bias in the coefficient estimate is small when the true value is small. The asymptotic standard deviation of $\hat{\phi}$, roughly $T^{-0.5}$ or .032, corresponded closely to the sample standard deviation in the MC sample as well as to the median of estimated standard errors . The *t*-test has correct size. Thus, asymptotic theory works well for the well identified AR(1).

In contrast to ϕ in the AR(1) estimation, both ϕ and θ in the ARMA(1,1) estimation are subject to *ZILC*. The first panel of Table 3.1.1 compares the (true) asymptotic standard deviation of $\hat{\theta}$ with its empirical standard deviation in the MC sample and with the median estimated standard error in the MC sample. Note that the actual standard deviation cannot be larger than one because coefficient estimates are bounded, as required by stationarity and invertibility, within the interval (-1,1). Since the asymptotic formula ignores this constraint, it overstates the standard deviation for small enough $\phi(=\gamma)$ as we see in Table 3.1.1 under .01. Nevertheless, the downward bias in estimated standard error is so strong that the median standard error is well below the actual standard deviation; 0.359 versus .659 respectively.

Table 3.1.1:
Inference for θ in ARMA(1,1)
DGP is AR (1); T=1,000

	True value of $\phi(=\gamma)$:			
	0.01	0.05	0.10	0.20
Std Dev. of $\hat{\theta}$:				
Asymptotic (true)	3.162	0.632	0.316	0.158
In MC sample	0.659	0.556	0.381	0.180
MC median Std Error	0.359	0.333	0.256	0.152
Information measure $I_{\hat{\theta}}$:				
Asymptotic (true)	0.10	2.50	10.00	40.00
MC median of est. *I*.	7.77	9.01	15.29	43.41
$E\left[\hat{\gamma}^2 T\right]$	1.1	3.5	11	41
Size of tests of null hypothesis θ=0 at nominal .05 level:				
EViews *t*-test.	0.457	0.358	0.245	0.118
Asy SE using $\hat{\phi},\hat{\theta}$.	0.371	0.341	0.229	0.116
Likelihood ratio test.	0.178	0.142	0.103	0.073
Frequency ARMA(1,1) selected over AR(1):				
AIC	0.380	0.337	0.251	0.180
SIC	0.040	0.026	0.019	0.014

Correspondingly, the estimated information measure, denoted $\hat{I}_{\hat{\theta}}$, the inverse of the variance estimate from EViews, is substantially overestimated, and much more than can be attributed to upward bias in estimating $\gamma^2 T$ alone. However, as larger values of ϕ(=γ) are considered, the *ZILC* effect diminishes until, for ϕ=.2, there is little difference between asymptotic and the median of $\hat{I}_{\hat{\theta}}$.

The third panel of Table 3.1.1 reports the results for versions of the Wald *t*-test using alternative standard errors as well as the likelihood ratio test, all at a nominal 0.05

level. The first *t*-test uses the EViews standard error and is greatly oversized. In case of weakest identification the actual size of the *t*-test is about .46 rather than .05. The second *t*-test uses standard error computed from (3.1.2) and, again, size is excessive.

In contrast, asymptotic theory works fairly well for the identifying parameter γ. The empirical standard deviation and the median of estimated standard errors were close to the asymptotic standard error, though standard errors are a bit low for $\gamma = .01, .05$ where the empirical size of the *t*-test is about .13.

The size of the likelihood ratio test is also too large, though less excessive than for the *t*-test. As the identifying difference between coefficients becomes larger the size distortion diminishes, though the size of both tests is still excessive with $\phi = .2$.

What does work well is the information criterion approach to model selection. The Schwarz Information Criterion (SIC) selects the ARMA(1,1) model over correct AR(1) specification only infrequently. The poorer performance of the AIC confirms the well-know superiority of the former in model selection; see Lutkepohl (1991).

It is clear from this experiment that spurious inference can be severe even in sample sizes that economists would usually consider rather large. This begs the question: how large does T have to be before asymptotic theory does take hold? For $\phi = .01$, the actual size of the *t*-test at the nominal .05 level falls from .46 for T = 1,000, to .36 for T=10,000, to .17 for T=100,000, and .06 for T=1,000,000 (based on 100 trials). Progress towards the asymptotic distribution is slow indeed.

In summary, *ZILC* predicts correctly predicts that weakly identified ARMA parameters are vulnerable to spurious inference, providing a note of caution to practitioners seeking to test for model specification. Although with sufficient sample size asymptotic theory will take hold, *ZILC* implies that spurious inference will occur regardless of sample size if near cancellation in the ARMA model is close enough. Ansley and Newbold (1980) reported that confidence intervals in ARMA models are too narrow when there is near parameter redundancy.

Section 3.2. Non-Linear Regression.

In this section we consider non-linear regression models that take the form of (2.2.5) directly, namely

$$y_i = \gamma \bullet g(\beta, w_i) + \varepsilon_i; \quad i = 1,..,N. \tag{3.2.1}$$

where the explanatory variables w are exogenous and errors ε are i.i.d. normal with mean zero and standard deviation σ_ε. The information measure for β obtained from the asymptotic covariance matrix is given by

$$I_{\hat{\beta}}(\beta, \gamma, \sigma_\varepsilon, W) = \gamma^2 \left[\frac{T \bullet m_{11} \bullet (1 - r_{01}^2)}{\sigma_\varepsilon^2} \right], \tag{3.2.2}$$

where m_{11} denotes the sample first sample moment of the first derivative of g, and r_{01} denotes the sample correlation between the zero and first derivatives. *ZILC* holds in non-linear regression models that have this form since I goes to zero as γ approaches zero.

In the case where f is linear in exogenous variables x and z we have:

$$y_i = \gamma(x_i + \beta z_i) + \varepsilon_i\,. \tag{3.2.3}$$

This model is of more general interest than might first appear since linearization of *f(.)* gives (3.2.3) as the first order approximation to (3.2.1), with x and z being the zero and first order derivatives respectively. Models of this form also arise directly in practice; for example Staiger, Stock, and Watson (1997) estimate Phillips Curve models where the NAIRU is a parameter to be estimated and takes the form:

$$\Delta\pi_t = \gamma(u_{t-1} - \overline{u}) + \varepsilon_t$$

where π is the inflation rate, u is the unemployment rate and $\overline{u}$ is the non-accelerating inflation rate of unemployment, playing the role of β, with z being unity.

The reduced form for (3.2.3) is the linear regression:

$$y_i = \gamma x_i + \lambda z_i + \varepsilon_i \,. \tag{3.2.4}$$

Since β is exactly identified, the least squares estimate of β, which is also maximum likelihood for Normal errors, is the ratio of regression coefficients $\hat{\beta} = \hat{\lambda} / \hat{\gamma}$. The square of the t-ratio for the null hypothesis $\beta = \beta^0$ is given by:

$$t_{\hat{\beta}}^2 = (\frac{\hat{\lambda}}{\hat{\gamma}} - \beta^0)^2 \bullet \left[\hat{\gamma}^2 \frac{T \bullet m_{11}(1 - r_{01}^2)}{s^2} \right] \tag{3.2.5}$$

where s denotes the standard error of the regression.

To see how the sampling distribution might work out in practice, we have simulated data for the case of uncorrelated regressors x and z with unit variances, setting β to zero and σ_ε to unity. Over a range of values for γ and T we obtained the following output from the EViews™ least squares routine:

Table 3.2.1: Sampling Distributions for Non-Linear Regression

		Information I_β		Standard Error of $\hat{\beta}$		Frequency
γ	T	Asymptotic	Median	Asymptotic	Median	$\lvert t_\beta \rvert > 1.96$
1	100	100	108	0.10	0.10	0.049
0.1	100	1	0.88	1.0	1.1	0.001
0.1	1,000	10	9.3	0.32	0.33	0.012
0.1	10,000	100	99	0.10	0.10	0.045
0.01	100	0.01	0.20	10	2.2	0.001
0.01	100,000	10	9.1	0.32	0.33	0.013
0.01	1,000,000	100	98	0.10	0.10	0.053

In the first experiment with γ set to unity and T=100 the model is well-identified, the median of estimated information and standard errors across Monte Carlo trials are close to asymptotic values and the size of the t-test is correct. Reducing the value of γ to 0.1 in the next experiment, the information and standard error are still close to asymptotic values, but the size of the t-test is much too small, reflecting offsetting co-variation in numerator and denominator. The latter phenomenon disappears when T is increased to 10,000. In the final three experiments γ is further reduced to 0.01, and with T = 100 the *ZILCH* effect is apparent, estimated information tends is too large and estimated standard error too small. However, the size of the t-test is very much below the nominal 0.05 level, and this does not get corrected at T= 100,000! The size is correct at T = 1,000,000, indicating how slowly asymptotic theory takes hold in this model.

The reason for the very low frequency of rejection even with very large sample size becomes clear when we note that the t-statistic may be expressed as:

$$t_{\hat{\beta}}^2 = \left(\frac{\hat{\lambda}}{\hat{\gamma}}\right)^2 \bullet \frac{\hat{\gamma}^2 \bullet T}{s^2\left(1+\hat{\beta}^2\right)} = \hat{\lambda}^2 \bullet \frac{T}{s^2} \bullet \frac{1}{\left(1+\hat{\beta}^2\right)} = t_{\hat{\lambda}}^2 \bullet \frac{1}{\left(1+\hat{\beta}^2\right)} \qquad (3.2.4)$$

This is a proper t-statistic – that for classical regression coefficient λ - multiplied by a quantity that is *always* less than one, regardless of sample size. The counterintuitive outcome depends on the lack of correlation between regressors in this example, and is reversed if they are strongly correlated.

To illustrate the importance of correlation between the regressors, we repeated the Monte Carlo experiments for $\gamma = 0.01$ and $T = 100$ for a range of correlations between x and z. Figure 3.2 shows the empirical size for a nominal five percent test, confirming that t-statistics are undersized for low correlations and oversized for high correlations with rejection rates hitting 80 percent in extreme cases.

In summary, *ZILC* applies for a wide range of non-linear regression models, leading to overestimation of information but t-statistics that may be – paradoxically - undersized.

3.3. Instrumental variables estimation with weak instruments.

The phenomenon of spurious inference in *IV* and *GMM* estimation when the instruments are weak is the subject of a large and growing literature. See Phillips (1983), Nelson and Startz (1990a, b), Bound, Jaeger and Baker (1995), Shea (1997), Staiger and Stock (1997), Dufour (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Stock and Wright (2000), Stock, Wright, and Yogo (2002), and Startz, Nelson and Zivot (2002), among others.

To illustrate how *ZILC* applies to *IV* we work with the basic model:

$$\begin{aligned} & y = \beta x + \varepsilon \\ & x = \gamma z + v \\ & V\begin{bmatrix} \varepsilon \\ v \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon\sigma_v \\ \rho\sigma_\varepsilon\sigma_v & \sigma_v^2 \end{bmatrix} \end{aligned} \tag{3.3.1}$$

where y, x, and z are data on N observations, z being fixed and exogenous, ε and v are unobserved *i.i.d.* shocks, and β is the parameter of interest for hypothesis testing.

If ρ is non-zero the least-squares estimate of β is inconsistent, providing the motivation for *IV*. Identification of β requires that γ be non-zero, and z is said to be a 'weak instrument' if γ is close to zero. However, γ is well-identified, being the coefficient in a classical linear regression. The reduced form equation, $y = \lambda z + \omega$; $\lambda = \beta\gamma$, is also a well-identified classical regression. The *IV* estimator of β is:

$$\hat{\beta}_{IV} = \hat{\lambda} / \hat{\gamma} = \beta + \frac{m_{z\varepsilon}}{\gamma + m_{zv}} \tag{3.3.2}$$

where $\hat{\lambda}$ and $\hat{\gamma}$ are least squares estimates and m denotes the sample moment for the indicated variables. The inverse of its asymptotic variance is:

$$I_{\hat{\beta}}(\beta, \gamma, \sigma_\varepsilon, Z) = \gamma^2 \left[\frac{N \bullet m_{zz}}{\sigma_\varepsilon^2} \right]. \tag{3.3.3}$$

Clearly *ZILC* holds, γ_0 being 0, this expression having the form of (2.3) exactly.

The case of strong simultaneity, when ρ is large, has received most attention in the weak instrument literature. In that case, the sampling distribution of $\hat{\beta}_{IV}$ is concentrated around the wrong value because then the numerator and denominator of (3.3.2) are highly correlated. The resulting estimated error variance is biased and the conventional *t*-test rejects the true null hypothesis far too often.

Here we look at what happens when ρ is zero, so there is no simultaneity and least squares estimation of β is optimal, but the researcher does *IV* anyway, using a weak instrument. This case is of interest both because it occurs in practice and because it isolates the role of the weak instrument. In the context of Section 2.3 it corresponds to estimating the ratio of uncorrelated regression coefficients and that analysis predicts that *t*-tests will be undersized. The concentration phenomenon does not apply in this case and the *IV* estimator is median unbiased, though its distribution has fat tails. An estimate of (3.4.3) is given by:

$$I_{\hat{\beta}_{IV}}(\hat{\beta}_{IV}, \hat{\gamma}, \hat{\sigma}_{\varepsilon}, \mathbf{Z}) = \hat{\gamma}^2 \left[\frac{N \bullet m_{zz}}{s^2_{(y-\hat{\beta}_{IV}x)}} \right] \qquad (3.3.4)$$

where the sample variance of the *IV* residuals is the estimated variance of the structural error and m_{zz} is known and fixed. *ZILC* implies that $\hat{\gamma}^2$ will tend to be too large when the instrument is weak, but the residual error variance will also tend to be too large because the *IV* estimate tends to be far from the true value. Nor are the two sources of error in estimating *I* independent.

To see how *ZILC* affects estimation in this case did a sampling experiment in EViews^{TM} using the standard routines and output. The parameter specification and results are given in Table 3.3.1:

	Table 3.3.1		
	IV estimation of β		
	$\beta = 0; \gamma = .01; \rho = 0;$ $\sigma_\varepsilon = \sigma_v = 1; m_{zz} = 1$		
	N= 100; 1,000 Replications		
Std Dev of $\hat{\beta}_{IV}$:			
Asymptotic	10		
In MC sample	43		
MC median Std Error	2.2		
Information *I*:			
Asymptotic	.01		
MC median $\hat{I}$	.21		
Test size at nominal .05 level:			
EViews t-test	0		
Using true *I*.	.03		

We note that the median of $\hat{I}$ (computed as the squared inverse of the estimated standard error) was 0.21, far in excess of the theoretical value of 0.01. Contributing to the bias is the co-variation of the two stochastic components, in particular, when $\hat{\gamma}^2$ is large, s^2 tends to be small, producing more large values of $\hat{I}$. In testing the null hypothesis $\beta =$ 0, its true value, we use alternatively (i) the true *I*, unknown to a real investigator, and (ii) the estimated value from the EViews standard error. At a nominal .05 level the rejection frequency using the true *I* is .03, but using the estimated standard error and *t*-statistic from EViews the corresponding rejection frequency is zero!

To see why *t*-statistics are too small in this case, noting that $(\hat{\beta}_{IV} - \beta) = {m_{z\varepsilon}}/{\hat{\gamma}}$, and using the normalization m_{zz}=1, we have:

$$t_{\hat{\beta}_{IV}}^2 = (\hat{\beta}_{IV} - \beta)^2 \bullet \hat{I} = (\frac{m_{z\varepsilon}}{\hat{\gamma}})^2 \bullet \left[\frac{N \bullet \hat{\gamma}^2}{s_{(y-\hat{\beta}_{IV}x)}^2}\right] = \left[\frac{\sqrt{N} \bullet m_{z\varepsilon}}{s_{(y-\hat{\beta}_{IV}x)}}\right]^2. \qquad (3.3.5)$$

Thus, sampling variation in $\hat{\gamma}$ tends to cancel, since it is a factor of both the numerator and denominator of *t*. What remains is the square of a standard normal variable with variance σ_ε^2 over the residual variance estimator of σ_ε^2. Because $\hat{\beta}_{IV}$ is median unbiased but has a large dispersion, s^2 tends to be too large; the Monte Carlo median of s^2 is 1.44. The net effect is that *t*-statistics are too small and the frequency of rejection far below the nominal level.

4. Summary and Conclusions

This paper introduces the Zero-Information-Limit-Condition (*ZILC*) which holds when the inverse of the asymptotic variance of an estimator, say $\hat{\beta}$, goes to zero as another parameter, say γ, approaches a critical value. We show that a common feature of models in which *ZILC* holds is that the inverse of the asymptotic variance of $\hat{\beta}$ is approximately proportional to γ^2. In practice, the standard errors and *t*-statistics produced by econometric software packages and relied upon in empirical research use estimates of the asymptotic variance. This common functional form suggests that the inverse of the variance, or ‘information,’ tends to be overestimated and standard errors tend to be too small. The most familiar example is the ‘weak instrument’ problem, but others discussed in the paper include ARMA models with near cancellation as well as non-linear regression models.

Models in which the *ZILC* holds turn out to share a common linear approximation in the form of a linear regression in which the parameter of interest is expressed as the ratio of coefficients. Whether the *t*-test is over-sized or over-sized is shown to depend on correlation between the ‘regressors’ in this linear representation. Thus the counterintuitive case of undersized *t*-tests in the face of underestimated standard errors is not difficult to construct, indeed two of our three examples fall into this category. We surmise that other models of importance in applied econometrics are candidates for *ZILC*;

GARCH, distributed lag, and unobserved components models among them. Progress toward valid inference presumably lies in the direction of likelihood based confidence intervals as suggested by Dufour (1997), demonstrated by Zivot, Nelson, and Startz (1998) for the *IV* case, and the local-to-zero asymptotics framework of Staiger and Stock (1997).

References

Anderson, T. W. and H. Rubin (1947): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20, 46-63.

Ansley, C. F. and P. Newbold, (1980), Finite sample properties of estimators for auto-regressive moving average processes, *J. of Econometrics*, 13, 159-184.

Bound, J, Jaeger, D. A. and Baker, R. M. (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *J. of American Statistical Association*, vol. 90, 443-450.

Box, George E. P. and Jenkins, Gwilym M. (1970, revised edition 1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden-Day.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247-254.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33-43.

Dufour, J.-M. (1997): "Some impossibility theorems in econometrics with application to structural and dynamic models," *Econometrica*, 65, 1365-88.

Hansen, Bruce E. (1996): Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis," *Econometrica*, 64, 413-430.

Lutkepohl, Helmut (1991), Introduction to Multiple Time Series Analysis, Berlin: Springer-Verlag.

Nelson, C. R. and Startz, R. (1990a): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967-976.

__________ and ________ (1990b): "The Distribution of the Instrumental Variables Estimator and Its *t*-ratio when the Instrument is a Poor One," *J. of Business*, 63, S125-S140.

Phillips, Peter C. B. (1983), "Exact small sample theory in the simultaneous equation model," in *Handbook of Econometrics*, vol. 1 chapter 8, ed. by Zvi Griliches and Michael D. Intrilligator, Amsterdam: North Holland.

Shea, J., (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *The Review of Economics and Statistics*, Vol. LXXIX, No. 2, 348-352.

Staiger, D. and Stock, J. H. (1997): "Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 65, 557-586.

Staiger, D., Stock, J. H. and Watson, M. W., (1997). "The NAIRU, Unemployment, and Monetary Policy," *J. of Economic Perspectives*, vol. 11 no. 1 (Winter), 33-49.

Stock, J. H. and Wright, J. H. (2000), "GMM with Weak Identification," *Econometrica,* 68, 1055-1096.

Stock, J. H. and Wright, J. H., and Yogo, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments" Journal of Business and Economics Statistics, 20 (4), 518-529.

Wang, J. and Zivot, E., (1998): "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica* , Vol. 66, No. 6, 1389-1404.

Zivot, E., Startz, R., and Nelson, C.R. (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, Vol. 39, No. 4, pp. 1119-1144.

Figure 2.1: The *ZILC* point and upward bias in estimated information.

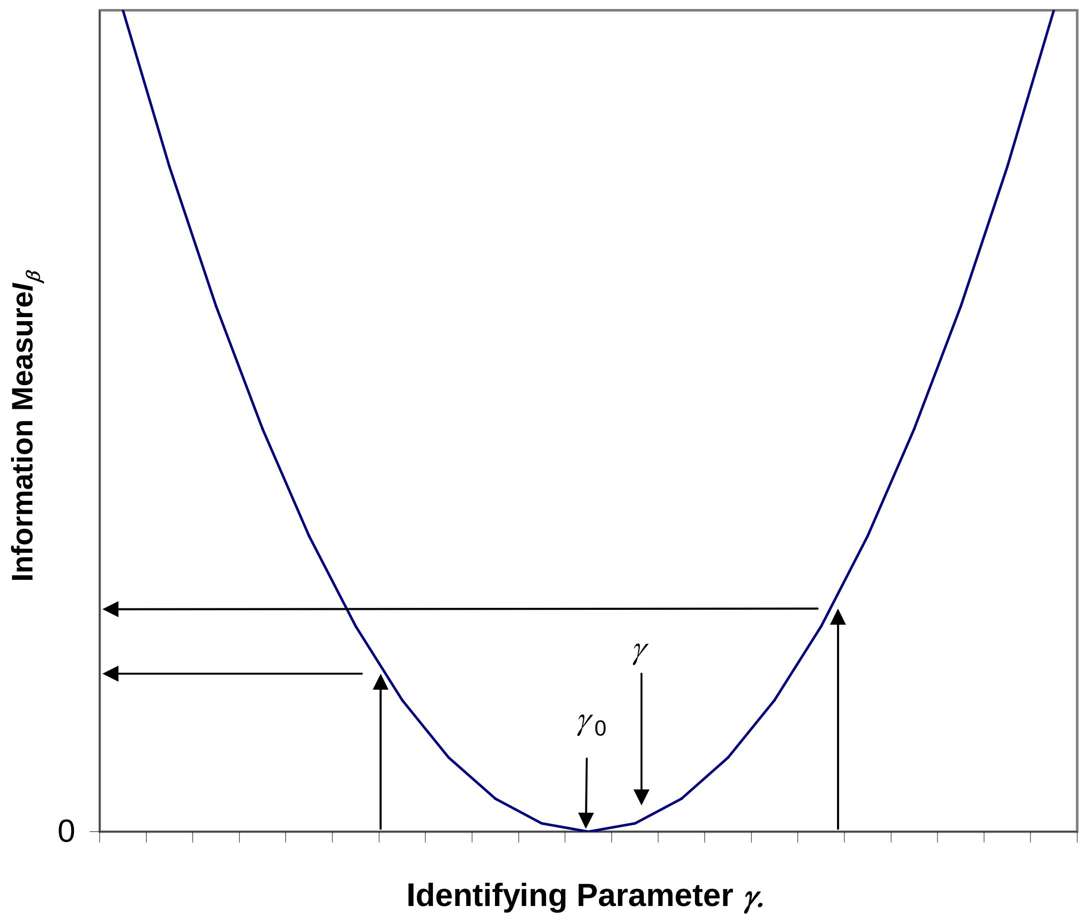
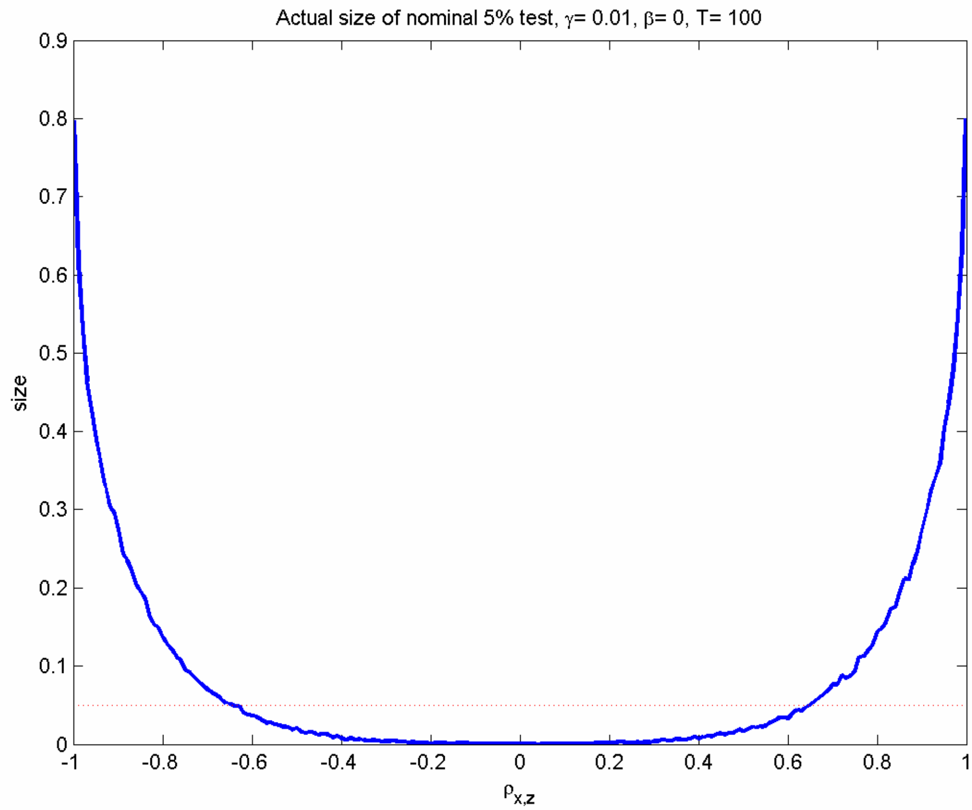


Figure 3.2



Actual size of nominal 5% test, γ= 0.01, β= 0, T= 100
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
size
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
$\rho_{x,z}$