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 October 24, 2003

Copyright © 2002, 2003 by Charles R. Nelson and Richard Startz

Not to be quoted without authors’ permission; comments welcomed.

Abstract

The fact that weak instruments lead to spurious inference is now widely recognized. In this paper we ask whether spurious inference occurs more generally in weakly identified models. To distinguish between models where spurious inference will occur from those where it does not, we introduce the Zero-Information-Limit-Condition (ZILC). When ZILC holds, the information or precision of parameter estimates is overestimated. Further, the numerator and denominator of the t-statistic will under certain circumstances be functionally related, not independent. We discuss how ZILC applies to models encountered in practice and show that spurious inference does occur when ZILC holds.

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 Yu-chin Chen, Kundan Kishor, Eric Zivot, and participants at the conference on “State Space Models, Regime Switching, and Identification” at Washington University in St. Louis, May 10-11, 2002 for helpful suggestions, 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.

The “weak instrument” problem in IV and GMM is now widely recognized; in these weakly identified models, point estimates, standard errors, confidence intervals, and t-tests are highly unreliable. In particular, in Nelson and Startz (1990a, b) we showed that when an explanatory variable is endogenous, but the instrumental variable only weakly correlated with it, the true null hypothesis regarding its coefficient is rejected far too often. Stock, Wright, and Yogo (2002) present a recent review of this literature.

In this paper we want to consider weakly identified models more broadly and ask whether the spurious inference problem is a general phenomenon. By ‘spurious inference’ we mean that the test has the wrong size in finite samples, even though asymptotically the size is correct. If spurious inference is not unique to IV and GMM, under what conditions does it occur? Clearly, weak identification does not always imply spurious inference. For example, in classical regression the coefficients are weakly identified (the data contain little information about them) if the fixed regressors are nearly collinear. But as long as they are not perfect collinearity, t-statistics are exactly t-distributed and test size is correct. In ARMA models with near cancellation the coefficients are weakly identified and, as we shall see, standard errors are much too small and the actual size of t-statistics is much too large; an example of spurious inference that is familiar to practitioners of time-series analysts.

To distinguish between models where spurious inference will occur when identification is weak from those where it does not we introduce the Zero-Information-Limit-Condition (ZILC) in section 2. We show that when ZILC holds, the information or precision of parameter estimates will be overestimated. Further, it becomes clear that the numerator and denominator of the t-statistic will under certain circumstances be functionally related rather than independent. In section 3 we discuss how ZILC applies to models encountered in practice and show that spurious inference does occur when ZILC holds. Section 4 concludes.
2. The Zero-Information-Limit Condition (ZILC).

Consider a model with scalar parameters $\beta$ and $\gamma$ as well as a vector of parameters $\sigma$, where $\beta$ is the parameter of interest for hypothesis testing. The variance of an estimate of $\beta$, say $\hat{\beta}$, will be a function of the parameters and a data matrix $W$. To emphasize the functional dependence of the variance on the parameters and data we will use the notation $V_\hat{\beta}(\beta, \gamma, \sigma, W)$. A natural measure of information associated with $\hat{\beta}$ is the inverse of the variance, sometimes called the precision of the estimate. Thus we define:

$$I_\hat{\beta}(\beta, \gamma, \sigma, W) = [V_\hat{\beta}(\beta, \gamma, \sigma, W)]^{-1}.$$ (2.1)

The abbreviation $I$ is used when the meaning is clear.

We will say that the Zero-Information-Limit Condition (ZILC) holds when there is a value of $\gamma$, say $\gamma_0$, such that

$$\lim_{\gamma \to \gamma_0} I_\hat{\beta}(\beta, \gamma, \sigma, W) = 0.$$ (2.2)

When ZILC holds, identification of $\beta$ is conditional on $\gamma$, and identification becomes weak as the identifying parameter $\gamma$ approaches $\gamma_0$.

The non-negativity of $I$ implies the function must be concave from above in the neighborhood of $\gamma_0$ as illustrated in Figure 1. Assuming the function $I$ is continuous and twice differentiable in $\gamma$, we expand the function around the ZILC point $\gamma_0$ noting that the zero and first order derivatives are zero at that point. One then obtains the second order approximation

$$I_\hat{\beta}(\beta, \gamma, \sigma, W) \approx (\gamma - \gamma_0)^2 \cdot I''(\beta, \gamma_0, \sigma, W)/2.$$ (2.3)
The results in this paper flow from the fact that the information measure is approximately proportional to the squared deviation of \( \gamma \) from the ZILC point because test statistics often involve an estimate of \( I \). The t-statistic for testing the null hypothesis \( \beta = \beta^0 \) has the form

\[
t_{\hat{\beta}} = (\hat{\beta} - \beta^0) \sqrt{\hat{I}_{\hat{\beta}}},
\]

where \( \hat{I} \) is an estimate of the information measure based on parameter estimates \( \hat{\beta}, \hat{\gamma}, \) and \( \hat{\sigma} \) and data \( W \). Any general properties of the joint sampling distribution of \( \hat{\beta} \) and \( \hat{I} \) are therefore relevant, and the role played by \( \hat{\gamma} \) in (2.3) suggests two.

One implication comes from the fact that there is an upward bias in estimating the factor of proportionality \( (\gamma - \gamma_0)^2 \). Note that the value of \( \gamma_0 \) is known, being implied in general by the model specification, while \( \gamma \) is unknown and must be replaced by an estimate, \( \hat{\gamma} \). If the latter is unbiased, then the expectation of the estimate of the proportionality term is

\[
E(\hat{\gamma} - \gamma_0)^2 = (\gamma - \gamma_0)^2 + V_{\hat{\gamma}}
\]

The upward bias resulting from the unavoidable necessity of estimating \( \gamma \) suggests, though does not by itself prove, that

\[
E[I_{\hat{\beta}}(\hat{\beta}, \hat{\gamma}, \hat{\sigma}, W)] \geq I_{\hat{\beta}}(\beta, \gamma, \sigma, W)
\]

Further, if \( \gamma \) is itself well-identified (in the sense that \( V_{\hat{\gamma}} \) does not depend on \( \gamma \), for example, when it is a classical regression coefficient), then bias in estimating \( (\gamma - \gamma_0)^2 \) becomes large relative to its true value when \( \gamma \) is close to \( \gamma_0 \). It is this relative or multiplicative bias that matters for the size of the t-statistic, since the estimate \( \sqrt{\hat{I}_{\hat{\beta}}} \) is multiplied times \( \hat{\beta} \).
Note however that asymptotic theory is still valid in models where ZILC holds. As sample size grows, $V\hat{\gamma}$ shrinks and becomes small relative to $(\gamma-\gamma_0)^2$ as the consistency of ML takes hold. 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 be more serious the closer the identifying parameter is to the ZILC point, a conjecture that we refer to as ZILCH.

The second implication following from the proportionality of $\hat{I}$ to $\hat{\gamma}^2$ in (2.3) is the potential interaction between estimates of $\beta$ and $\gamma$. Unlike the situation in classical linear regression where the components of the t-statistic are independent, there is no reason in general for that to be true when ZILC holds. For example, if large sampling errors in $\hat{\beta}$ tend to be associated with small deviations in $\hat{\gamma}$ from the ZILC point, then the dispersion of the t-statistic may be attenuated, reducing the frequency of rejections relative to the nominal level.

Both types of spurious inference occur in the examples that follow.

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

3.1. Instrumental variables estimation with weak instruments.


To illustrate how ZILC applies to IV we work with the basic model:
\begin{align*}
  y &= \beta x + \varepsilon \\
  x &= \gamma z + \nu \\
  \nu \begin{bmatrix}
      \varepsilon \\
      \nu
    \end{bmatrix} &= 
    \begin{bmatrix}
      \sigma_\varepsilon^2 & \rho \sigma_\varepsilon \sigma_\nu \\
      \rho \sigma_\varepsilon \sigma_\nu & \sigma_\nu^2
    \end{bmatrix}
\end{align*}

(3.1.1)

where \( y, x, \) and \( z \) are data on \( T \) observations, \( z \) being fixed and exogenous, \( \varepsilon \) and \( \nu \) are unobserved \( i.i.d. \) disturbances with mean zero, and \( \beta \) is the parameter of interest for hypothesis testing. If \( \rho \) is non-zero then \( x \) is correlated with \( \varepsilon \), reflecting feedback or simultaneity, and the least-squares estimate of \( \beta \) is inconsistent, providing the motivation for IV.

Identification of \( \beta \) in general requires that \( \gamma \) be non-zero, and \( z \) is said to be a weak instrument if \( \gamma \) is close to zero. However, \( \gamma \) is well-identified, being the coefficient in a classical linear regression. The reduced form equation for \( y, y = \theta z + \omega \), is also a well-identified classical regression. The IV estimate of \( \beta \) may be written:

\[
\hat{\beta}_{IV} = \hat{\theta} / \hat{\gamma} = \beta + \frac{m_{z\varepsilon}}{\gamma + m_{z\nu}}
\]

(3.1.2)

where hats indicate least squares estimates and we use the standard notation for sample second moments.

Under normality the IV estimator is also maximum likelihood in this model, and the inverse of the asymptotic variance is:

\[
I_{\hat{\beta}}(\beta, \gamma, \sigma_\varepsilon, Z) = \gamma^2 \left[ \frac{T \cdot m_{zz}}{\sigma_\varepsilon^2} \right].
\]

(3.1.3)

Clearly ZILC holds in the IV model, with \( \gamma_0 \) being 0, since this expression, which has the form of (2.3) exactly, goes to zero as \( \gamma \) goes to zero.

The case of strong simultaneity, when \( \rho \) 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, essentially because then the numerator and denominator of (3.1.2) are highly correlated. The resulting estimated error variance is biased and the conventional t-test rejects the true null hypothesis far too often. Thus, an investigator relying on nominal significance levels will tend to find a relationship between \( y \) and \( x \) though none exists when there is feedback from \( y \) to \( x \) and the instrument \( z \) is weak.

Here we look at what happens when \( \rho \) is zero, so there is no simultaneity and least squares estimation of \( \beta \) is optimal, but one does IV anyway, using a weak instrument. This case is of interest here not only because it presumably occurs in practice but because it isolates the role of the weak instrument. 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 (3.1.3) is given by:

\[
I_{\hat{\beta}_{IV}}(\hat{\beta}_{IV}, \hat{\gamma}, \hat{\sigma}_e, Z) = \hat{\gamma}^2 \frac{T \cdot m_{zx}}{s^2(\bar{y} - \hat{\beta}_{IV} x)}
\]

(3.1.4)

where the sample variance of the IV residuals is the estimated variance of the structural error and \( m_{zx} \) is known and fixed. The numerator will tend to be too large when the instrument is weak, but the residual error variance in the denominator 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 ZILCH works in this case we have done a sampling experiment with the following parameter specification:

\[ \beta = 0; \]
\[ \gamma = .01; \]
\[ \sigma_e = \sigma_y = 1; \]
\[ m_{zx} = 1; \]
\[ \rho = 0; \]

Implying that: \( I_{\hat{\beta}_{IV}} = .01. \)
Estimation was done in EViews\textsuperscript{TM} using the standard routines and output. In a sample of 1,000 runs we found 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 \( \hat{I} \). As predicted by ZILCH, estimated information is biased upward.

In testing the null hypothesis \( \beta = 0 \), its true value, we use alternatively (1) the true \( I \), unknown to a real investigator, and (2) the estimated value (specifically from EViews’ standard error). At a nominal 5% significance level the rejection frequency using the true \( I \) is .03. In contrast, using the estimated standard error and t-statistic from EViews, we have a corresponding rejection frequency of zero! How can it be that we greatly overestimate the information in the data but get t-statistics that are too small?

It becomes clear what is going on from the square of the t-ratio. Noting that
\[
\gamma \beta \epsilon = -z \text{IV} \text{m} = -z, \quad \text{and using the normalization } m_{z\epsilon}=1, \text{ we have:}
\]

\[
t^2_{\hat{\beta}_{IV}} = (\hat{\beta}_{IV} - \beta)^2 \cdot \hat{I} = \left( \frac{m_{z\epsilon}}{\hat{\gamma}} \right)^2 \left[ \frac{T \cdot \hat{\gamma}^2}{s^2 (y - \hat{\beta}_{IV} x)} \right] = \left[ \frac{\sqrt{T} \cdot m_{z\epsilon}}{s (y - \hat{\beta}_{IV} x)} \right]^2.
\]

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 \( \sigma_{\epsilon}^2 \) over the residual variance estimate of \( \sigma_{\epsilon}^2 \). 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, as observed in the Monte Carlo experiment.
Section 3.2. Non-Linear Regression

In this section we consider non-linear regression models of the form

\[ y_i = \gamma \cdot f(w_i, \beta) + \varepsilon_i; \quad i = 1, \ldots, N. \]  

(3.2.1)

where the explanatory variables \( w \) are exogenous and errors \( \varepsilon \) are i.i.d. normal with mean zero and standard deviation \( \sigma_\varepsilon \). The information measure for \( \beta \) obtained from the asymptotic covariance matrix is readily shown to be

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

(3.2.2)

where \( m_{11} \) denotes the sample first moment of the first derivative of \( f \) and \( r_{01} \) denotes the sample correlation between the zero and first derivatives. ZILC holds in non-linear regression models that have the form (3.2.1), since \( I \) goes to zero as \( \gamma \) approaches zero.

In the case where \( f \) is linear in exogenous variables \( x \) and \( z \) we have a model of the form:

\[ y_i = \gamma(x_i + \beta z_i) + \varepsilon_i. \]  

(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. In its simplest form the model takes the form

\[ \Delta \pi_t = \gamma(u_{t-1} - \bar{u}) + \varepsilon_t \]
where $\pi$ is the inflation rate, $u$ is the unemployment rate and $\bar{u}$ is the non-accelerating inflation rate of unemployment, playing the role of $\beta$ in this example, with $z$ being unity.

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

$$y_i = \gamma x_i + \theta z_i + \varepsilon_i.$$  \hspace{1cm} (3.2.4)

Since $\beta$ is exactly identified, the least squares estimate of $\beta$, which is also maximum likelihood for Normal errors, is simply:

$$\hat{\beta} = \hat{\theta} / \hat{\gamma}.$$ \hspace{1cm} (3.2.5)

The square of the t-ratio for the null hypothesis $\beta = \beta^0$ may be written:

$$t_{\beta}^2 = (\hat{\theta} / \hat{\gamma} - \beta^0)^2 \cdot \left[ \hat{\gamma}^2 \frac{N \cdot m_1 (1-r_{01}^2)}{s^2} \right]$$ \hspace{1cm} (3.2.3)

where $s$ denotes the standard error of the regression. As in the IV model, $\hat{\gamma}$ appears in both the denominator and the numerator of the t-ratio, so sampling variation will tend to cancel.

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 $\beta$ to zero and $\sigma_\varepsilon$ to unity. Over a range of values for $\gamma$ and $N$ we obtained the following output from the EViews™ least squares routine:
Table 3.2.1: Sampling Distributions for Non-Linear Regression

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$N$</th>
<th>Information $I_\beta$</th>
<th>Standard Error of $\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Asymptotic Median</td>
<td>Asymptotic Median</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100 108</td>
<td>0.10 0.10</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>1 0.88</td>
<td>1.0 1.1</td>
</tr>
<tr>
<td>0.1</td>
<td>1000</td>
<td>10 9.3</td>
<td>0.32 0.33</td>
</tr>
<tr>
<td>0.1</td>
<td>10000</td>
<td>100 99</td>
<td>0.10 0.10</td>
</tr>
<tr>
<td>0.01</td>
<td>100</td>
<td>0.01 0.20</td>
<td>10 2.2</td>
</tr>
<tr>
<td>0.01</td>
<td>100000</td>
<td>10 9.1</td>
<td>0.32 0.33</td>
</tr>
<tr>
<td>0.01</td>
<td>1000000</td>
<td>100 98</td>
<td>0.10 0.10</td>
</tr>
</tbody>
</table>

In the first experiment with $\gamma$ set to unity and $N=100$ 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 (at a nominal .05 level) is correct. Reducing the value of $\gamma$ 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 $N$ is increased to 10,000. In the final three experiments $\gamma$ is further reduced to 0.01, and with $N=100$ 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 $N=100,000$! The size is correct at $N=1,000,000$, 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:

$$t^2_{\hat{\beta}} = \left(\frac{\hat{\theta}}{\hat{\gamma}}\right)^2 \cdot \frac{\hat{\gamma}^2 N}{s^2 (1 + \hat{\beta}^2)} = \hat{\theta}^2 \cdot \frac{N}{s^2} \cdot \frac{1}{(1 + \hat{\beta}^2)} = t^2_{\hat{\theta}} \cdot \frac{1}{(1 + \hat{\beta}^2)} \quad (3.2.4)$$
This is a proper t-statistic – that for classical regression coefficient $\theta$- multiplied by a quantity that is always less than one, thus reducing its dispersion regardless of sample size.

Section 3.3. The ARMA (1,1) model with near cancellation.

The ARMA(1,1) model may be written:

$$(1 - \phi L)y_t = (1 - \theta L)e_t; t = 1, \ldots, T; e_t \sim \text{i.i.d. } N(0, \sigma^2_e), \quad |\phi|1; |\theta|1. \quad (3.3.1)$$

In the case $\phi = \theta$, the model reduces to white noise, the AR and MA polynomials in lag operator $L$ canceling exactly. Since any common value for both $\phi$ and $\theta$ in the interval (-1,1) would correspond to the same distribution of the data, the model is identified only if $\phi \neq \theta$. This is also clear from the asymptotic covariance matrix:

$$V_{\hat{\phi}, \hat{\theta}}(\phi, \theta) = T^{-1} \frac{(1 - \phi \theta)}{(\phi - \theta)^2} \begin{bmatrix}
(1 - \phi^2)(1 - \phi \theta) & (1 - \phi^2)(1 - \theta^2) \\
(1 - \phi^2)(1 - \theta^2) & (1 - \theta^2)(1 - \phi \theta)
\end{bmatrix} \quad (3.3.2)$$

which is infinite if $\phi = \theta$. However, as long as this equivalence is not exact, the ARMA(1,1) model is identified and asymptotic theory holds.

What happens, though, when the difference between $\phi$ and $\theta$ is small? In a Monte Carlo experiment where the true parameter values were $\phi = .01$ and $\theta = 0$ with $T=1,000$, the t-test for testing the (true) null hypotheses $\theta = 0$ at a nominal .05 level had an actual size of .46. This is the more surprising in light of the constraint on estimates to fall into the interval of invertibility while asymptotic theory does not correspondingly limit the standard error, factors which might have suggested the test would be undersized. Instead, there is a tendency to over-estimate the MA order of the model.

Testing for MA order comes up frequently in practice and can have important implications for model specification, for example in the trend-plus-cycle decomposition.
of real GDP. A random walk trend plus AR(p) cycle implies a ARIMA(p,1,p) reduced form; see Morley, Nelson, and Zivot, 2003. In the case p=1, rejection of ARMA(1,0) in favor of ARMA(1,1) would provide evidence in support of the decomposition.

Also, it is clear from the value of 1,000 for \( T \) that spurious inference is not limited to sample sizes that economists usually think of as small. Although with sufficient sample length asymptotic theory does take hold, it turns out that spurious inference emerges regardless of sample size if the absolute difference between the two parameters is small enough.

To see why spurious inference occurs when there is near cancellation, we re-parameterize the model in terms of \( \theta \) and the difference \( \gamma = \phi - \theta \) as follows:

\[
(1 - (\theta + \gamma)L)y_t = (1 - \theta L)\epsilon_t. \tag{3.3.3}
\]

The covariance matrix \( V_{\hat{\theta}, \hat{\gamma}}(\theta, \delta) \) for this re-parameterized model is easily obtained and inverting the element for \( \theta \) gives information measure:

\[
I_{\theta}(\theta, \gamma) = \frac{\gamma^2 T}{(1 - \theta^2 + \theta \gamma)^2 (1 - \theta^2)}. \tag{3.3.4}
\]

Since this function approaches zero as \( \gamma \) as goes to zero, ZILC holds for the ARMA(1,1) model, \( \gamma \) being the identifying parameter with \( \gamma_0 = 0 \).

We note that estimated information \( I_{\theta}(\hat{\theta}, \hat{\gamma}) \) will depends primarily on the estimate of \( \gamma^2 \) since (3.3.4) is relatively insensitive to \( \theta \). Under the null hypothesis \( \theta = 0 \), it reduces to simply \( \gamma^2 T \). The bias in estimating \( \gamma^2 \) is the variance of the estimate of \( \gamma \), and that is given by:

\[
V_{\gamma}(\theta, \gamma) = T^{-1} [1 - (\theta^2 - \theta \gamma)^2] \simeq T^{-1} (1 - \theta^4), \tag{3.3.5}
\]
the approximation holding for small $\gamma$. Note that $ZILC$ does not hold for $\gamma$ since (3.3.5) is non-zero for all values of the parameters within stationarity and invertibility bounds. Under the null hypothesis $\theta = 0$ it is simply $T^{-1}$ and it is insensitive to $\theta$ across a wide range since remaining terms are of fourth order. Thus, $\gamma$ is well identified.

The upward bias in estimating the information measure for $\theta$ will be large if $\gamma^2$ is small relative to $T^{-1}$. For the example above, $T = 1,000$, $\theta = 0$, and $\gamma = .01$, we have:

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

Thus, we can expect $\hat{\theta}$ to be an order of magnitude too large.

**Sampling Experiments.**

To calibrate the importance of the spurious inference in the ARMA(1,1) model we have done a series of Monte Carlo (MC) experiments where the data generating process (DGP) is AR(1) and we fit both AR(1), which is correctly specified and well identified, and the ARMA(1,1) model which is also correctly specified but weakly identified. For the latter, the true value of $\theta$ is zero so $\gamma(=\phi)$ is both the identifying parameter and the AR coefficient, which we set at values of .01, .05, .10 and .20 in successive experiments. Series length $T$ is 1,000, and the number of replications is 1,000.

Estimation is done within EViews™. After each series is generated we estimated both the AR(1) model and the ARMA(1,1) model, recording coefficient estimates and standard errors. Individual replications were discarded from the MC sample if either stationarity or invertibility was violated by point estimates; this occurred in two replications with $\gamma$ values of .01 and .05 and not at all for larger values.

Standard errors computed by EViews™ come out of the general non-linear estimation algorithm in that software package and will differ in general from the values that would be obtained by plugging the coefficient estimates into the asymptotic formula. However, we find little difference in the resulting t-tests.

**Discussion of results for estimation of $\phi$ in AR(1) model.**
Table 3.3.1 presents results for estimating the AR(1) model. As is well known, bias in the coefficient estimate is small when the actual value is small, and negligible in this context. It is useful as a benchmark to note that the asymptotic standard error of \( \hat{\phi} \) is roughly \( T^{-0.5} \) or .032. From the first panel we see that this corresponds well to the standard deviation in the MC sample as well as to the median of estimated standard errors in the MC sample. The inverse of the squared standard error is an estimate of the information measure, which we denote \( \hat{I}_\phi \); to indicate that it is distinct from an estimate obtained by evaluating the asymptotic \( I \) function at the coefficient estimates. The median in the MC sample is close to the asymptotic value of roughly 1,000. The last panel shows that the t-test based on the estimated standard error has nearly correct size and considerable power against the alternative of zero. Thus, for the AR(1) model, which is unconditionally identified, asymptotic theory is very accurate in a series of length 1,000.

Discussion of results for estimation of \( \theta \) in ARMA(1,1) model.

In contrast to \( \phi \) which is well identified in the AR(1) model, identification of both \( \phi \) and \( \theta \) in the ARMA(1,1) model is conditional on the absolute difference between them. We focus here on inference for \( \theta \), although results for \( \phi \) are qualitatively the same. A test of the true null hypothesis \( \theta = 0 \) is also a test that the MA order is zero.

The first panel of Table 3.3.2 compares the (true) asymptotic standard deviation of estimated \( \theta \) with the sample standard deviation across the MC sample and with the median of the estimated standard error across the MC sample. Note that the actual standard deviation cannot be larger than one because coefficient estimates are bounded 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 3.3.2 under .01. Nevertheless, the downward bias in estimated standard error is so strong that even in this case the median standard error is well below the actual standard deviation; 0.359 versus .659 respectively.

Correspondingly, the estimated information measure \( \hat{I}_\phi \), the inverse of the variance estimate from EViews, is seen to be substantially overestimated. In shown in Table 2, the upward bias is much greater than can be attributed to upward bias in
estimating the factor $\delta^2T$ alone. However, as larger values of $\phi(=\delta)$ are considered, this effect diminishes until, for $\phi=.2$, there is little difference between asymptotic and the median of estimated information across the MC sample.

The third panel of Table 3.3.2 reports the results for versions of the Wald t-test using alternative standard errors as well as a conditional t-test and the likelihood ratio test, all at a nominal 0.05 significance level. The first t-test uses the standard error provided by EViews and the actual size is greatly in excess of its nominal .05 value. The second t-test uses a standard error computed using the asymptotic formula in which coefficient values are replaced by the point estimates. Again, the size is excessive. The third t-test also uses the asymptotic formula but computes it under the null hypothesis, using the AR(1) estimate for $\phi$. It rejects far too infrequently in the weakly identified case we are interested in.

The identifying role of $\gamma$ suggests that a test that conditions on its significance of might improve the size performance of the t-test for $\theta$. Here, if it exceeds a critical value designed to give correct size in the experiment for $\phi = .01$, then the customary t-test is used, otherwise the null $\theta = 0$ is accepted. Unfortunately, the size is somewhat excessive for the other values of $\phi$ considered here.

Finally, 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).

Discussion of inference for $\gamma(=\phi)$ in the ARMA(1,1) model.

We saw above that the identifying parameter $\gamma$ is well identified with asymptotic standard error of about $T^{-0.5}$. What we see in Table 3.3.3 is that both the MC sample standard deviation and MC median of estimated standard errors are very close to that number, though standard errors are a bit low for the smallest value of $\gamma$, .01. This is also reflected in the moderate upward bias in the estimated information measure and the
moderate excess size of the t-test. Better size results are obtained by using the asymptotic standard deviation, $T^{0.5}$, in place of the estimated standard error from EViews, rejections being very close to .05 for all values of $\delta$. Power against the null that $\gamma$ is zero is very comparable to that of the t-test in the AR(1) model.

**Section 3.4. Stochastic Regressors with Near Multicollinearity.**

Our fourth example of weakly identified models is the regression model with highly correlated stochastic regressors. The data generating mechanism may be specified as follows. The two regressors, $x_1$ and $x_2$, are jointly distributed with covariance matrix given by:

$$
\text{Cov}(x_{1,i}, x_{2,j}) = \begin{bmatrix} \sigma_1^2 & \gamma \sigma_1 \sigma_2 \\ \gamma \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}.
$$

An error term, $e_i$, is independent of the regressors with standard deviation $\sigma_e$. The $i$th observation on the dependent variable, $y_i$, is generated by

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + e_i,$$

where the coefficients are to be estimated by least squares. To check whether ZILC applies in this model we note that

$$I(\beta, \gamma, \sigma_e, W) = \left( \sigma_e^2 \left[ XX^\prime \right]^{-1} \right)^{-1} = I(\sigma_e, X),$$

where $X$ denotes the N by 2 matrix of observations drawn from the distribution of the regressors.

Although the model is not identified if $\gamma = 1$, in which case $I$ is identically zero, the ZILC (2.2) does not hold because $I$ is a function of the actual realization of $X$, not of the parameters of its distribution which includes $\gamma$. In particular, $I$ will depend on the sample
correlation between the two regressors in the realization of X, and that will be zero with probability zero. In the case of Normal errors the t-statistic will have a t-distributed and will reflect appropriately the lack of information arising from correlation between the regressors.

4. Summary and Conclusions

Weak identification leads to spurious inference in some models but not in all. In this paper we introduce the Zero-Information-Limit-Condition (ZILC) to distinguish the cases where spurious inference occurs from those where test statistics are reliable. The key result is that when ZILC holds the information or precision of a parameter estimate is approximately proportional to the square of the sampling error in an identifying parameter. That functional form implies that information tends to be overestimated, that is, standard errors are too small. The effect of this phenomenon on test size turns out to depend on how, or whether, the information is functionally related to the estimate of the parameter of interest. We consider four seemingly unrelated models from practice where weak identification may occur but where ZILC applies differently.
References


_________ 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.


Table 3.3.1: Inference for $\phi$ in AR(1).
DGP is AR(1) with coefficient $\phi$; $T=1,000$.
Monte Carlo sample of 1,000 replications;
estimates from EViews.

<table>
<thead>
<tr>
<th>True value of $\phi$ in DGP:</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Dev. of $\hat{\phi}$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic (true)</td>
<td>0.032</td>
<td>0.032</td>
<td>0.031</td>
<td>0.031</td>
</tr>
<tr>
<td>In Monte Carlo sample</td>
<td>0.033</td>
<td>0.032</td>
<td>0.032</td>
<td>0.031</td>
</tr>
<tr>
<td>MC median Std Error</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.031</td>
</tr>
</tbody>
</table>

| Information $I_{\hat{\phi}}$: |      |      |      |      |
| Asymptotic (true)             | 1000 | 1003 | 1010 | 1042 |
| MC median of est. $I_{\hat{\phi}}$ | 998   | 1000 | 1007 | 1038 |

| Rejections at .05 level:      |      |      |      |      |
| t($\phi$=true): size          | 0.055| 0.057| 0.054| 0.054|
| t($\phi$=0): power            | 0.067| 0.349| 0.876| 1.000|
Table 3.3.2: Inference for $\theta$ in ARMA(1,1) Model.

<table>
<thead>
<tr>
<th>True value of $\phi (= \gamma)$ in AR(1) DGP:</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std Dev. of $\hat{\theta}$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic (true)</td>
<td>3.162</td>
<td>0.632</td>
<td>0.316</td>
<td>0.158</td>
</tr>
<tr>
<td>In MC sample</td>
<td>0.659</td>
<td>0.556</td>
<td>0.381</td>
<td>0.180</td>
</tr>
<tr>
<td>MC median Std Error</td>
<td>0.359</td>
<td>0.333</td>
<td>0.256</td>
<td>0.152</td>
</tr>
<tr>
<td>Information measure $I_{\hat{\phi}}$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic (true)</td>
<td>0.10</td>
<td>2.50</td>
<td>10.00</td>
<td>40.00</td>
</tr>
<tr>
<td>MC median of est. $I$.</td>
<td>7.77</td>
<td>9.01</td>
<td>15.29</td>
<td>43.41</td>
</tr>
<tr>
<td>$E[\hat{\phi}^2 T]$</td>
<td>1.1</td>
<td>3.5</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>Tests of null hypothesis $\theta=0$ at nominal .05 level, rejections:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EViews t-test.</td>
<td>0.457</td>
<td>0.358</td>
<td>0.245</td>
<td>0.118</td>
</tr>
<tr>
<td>Asy SE using $\hat{\phi}, \hat{\theta}$.</td>
<td>0.371</td>
<td>0.341</td>
<td>0.229</td>
<td>0.116</td>
</tr>
<tr>
<td>Asy SE using AR(1) $\hat{\phi}$.</td>
<td>0.000</td>
<td>0.016</td>
<td>0.050</td>
<td>0.074</td>
</tr>
<tr>
<td>Use t if reject $\gamma = 0$.</td>
<td>0.050</td>
<td>0.121</td>
<td>0.211</td>
<td>0.118*</td>
</tr>
<tr>
<td>Likelihood ratio test.</td>
<td>0.178</td>
<td>0.142</td>
<td>0.103</td>
<td>0.073</td>
</tr>
</tbody>
</table>

ARMA(1,1) selected over AR(1), frequency:
<table>
<thead>
<tr>
<th>AIC</th>
<th>0.380</th>
<th>0.337</th>
<th>0.251</th>
<th>0.180</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIC</td>
<td>0.040</td>
<td>0.026</td>
<td>0.019</td>
<td>0.014</td>
</tr>
</tbody>
</table>

*Uses t-test on $\theta$ if t-statistic on $\delta$ exceed critical value, otherwise accept null. Critical value for first step was chosen to produce actual size of .05 when $\phi=.01$. Std error for $\gamma$ used here is inverse root T, but similar results hold using reported SE.
Table 3.3.3: Inference for identifying parameter $\gamma$ in ARMA(1,1).

DGP is AR(1) with coefficient $\phi$; $T=1,000$.
Monte Carlo sample is 1,000 runs; estimates from EViews.

<table>
<thead>
<tr>
<th>True value of $\phi$ in AR(1) DGP:</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std Dev of $\hat{\gamma} = \hat{\phi}$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic (true)</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>In MC sample</td>
<td>0.034</td>
<td>0.033</td>
<td>0.034</td>
<td>0.033</td>
</tr>
<tr>
<td>MC median Std Error</td>
<td>0.028</td>
<td>0.031</td>
<td>0.032</td>
<td>0.032</td>
</tr>
</tbody>
</table>

| Estimated $I_{\hat{\phi}}$:       |      |      |      |      |
| Asymptotic                         | 1000 | 1003 | 1010 | 1042 |
| MC median of estimated $I$         | 1321 | 1073 | 1002 | 997  |

Tests on $\gamma$ at nominal .05 level, rejections:

| $t(\gamma = true; se=est.)$: size | 0.135| 0.139| 0.081| 0.055|
| $t(\gamma = true; se=.001)$: size | 0.056| 0.066| 0.056| 0.054|
| $t(\gamma = 0; se=.001)$: power    | 0.067| 0.337| 0.851| 1.000|
Figure 1: The ZIL and estimated information.