On the Correlations of Trend-Cycle Errors

Wada, Tatsuma

19 December 2011

On the Correlations of Trend-Cycle Errors*

Tatsuma Wada†

Wayne State University

This version: December 19, 2011

Abstract

This note provides explanations for an unexpected result, namely, the estimated parameter of the correlation coefficient of the trend shock and cycle shock in the state-space model is almost always (positive or negative) unity, even when the true variance of the trend shock is zero. It is shown that the set of the true parameter values lies on the restriction that requires the variance-covariance matrix of the errors to be nonsingular, therefore, almost always the likelihood function has its (constrained) global maximum on the boundary where the correlation coefficient implies perfect correlation.

JEL Classification Number: C13, C22.

Keywords: Trend-Cycle Decomposition, Unit-root, Maximum likelihood.

1 Introduction

When the trend-cycle decomposition of economic time series data is implemented through a state-space (or unobserved components, UC) model, the correlation of a shock (error) to the trend and a shock (error) to the cycle is often assumed to be zero. This is due to the fact that the correlation coefficient is generally unidentified, as Watson (1986) demonstrates. A recent influential paper by Morley et al. (2003), however, shows that the correlation of errors can be identified only if identification conditions are satisfied; and that whether or not allowing such a correlation is a key to understanding the substantial differences between "business cycles" estimated by the Beveridge-Nelson decomposition and by the UC model.

Among others, a study by Perron and Wada (2009) finds that US GDP follows a stationary process, when the trend function is allowed to have sudden changes. Although there is no shocks to trend,[1] the maximum likelihood estimator implies perfect correlation. In addition, their simulation shows that the estimated correlation parameter is almost always (positive or negative) unity. This unexpected result is worth exploring. Since, in practice, researchers may estimate the correlation of the errors by simply assuming the data are non-stationary without applying a variety of unit root

*I am grateful to an anonymous referee for useful comments. All remaining errors are mine.

†Department of Economics, Wayne State University, 656 W. Kirby St., Detroit, MI, 48202 (tatsuma.wada@wayne.edu).

[1] Except the shock that causes the trend to change suddenly.

tests; hence, knowing properties of the estimator when the true parameter is not identified, serves the purpose of diagnosing model misspecification. In line with the idea of detecting misspecification via correlated UC models, Morley et al. (2011) propose a likelihood ratio test of stationarity utilizing the fact that the correlation is observed only when trend errors exist.

This note demonstrates that such a perfect correlation, when the true data generating process (DGP) is stationary, is artificially created due to the restriction requiring that the variance-covariance matrix of the errors be positive-semi-definite. Since the true parameters lie on the boundary of this restriction, where the correlation coefficient is positive or negative unity, or one of the variances is zero, the estimated correlation coefficient is almost always 1 or -1, as Table 1 displays,[2] rather than the undefined correlation with zero variance. The rest of this note is organized as follows. Section 2 explains our model. The likelihood function and its properties for a simple model under the restriction are analyzed in Section 3. An extension to the AR(p) error in the cyclical component is presented in Section 4. Section 5 concludes.

2 Model

Our model is given by:

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \tau_{t-1} + u_t \\ \phi(L) c_t &= \varepsilon_t \end{aligned} \tag{1}$$

where y_t is observable variable; τ_t and c_t are the trend and cyclical components, respectively; $\phi(L)$ is the lag-polynomial; the shock to the trend u_t and the shock to the cycle ε_t are drawn from the bivariate Normal distribution:

$$\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} \sim i.i.d.\ N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{u\varepsilon} \\ \sigma_{u\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix} \right).$$

The correlation coefficient is defined as $\rho = \sigma_{u\varepsilon} / (\sigma_u \sigma_\varepsilon)$.

3 A Simple White Noise Case

3.1 The Likelihood Function

Let the data DGP be a white noise process:

$$y_t = \varepsilon_t$$

where ε_t is a zero-mean, normally distributed iid process with a variance of σ_ε^2. Consider the state-space model:

$$\begin{aligned} y_t &= \beta_t + \varepsilon_t \\ \beta_t &= \beta_{t-1} + u_t. \end{aligned}$$

[2] As Table1 reveals, it is still possible to obtain an estimate that is neither 1 nor -1. The relative frequency of finding such estimates is, however, only about 10%. Note also that the results displayed in Table 1 are different from Perron and Wada's (2009) simulation results, since their model includes a change in the slope of the deterministic trend, in addition to the stationary component.

Assuming that ε_t and u_t are Normally distributed iid processes, the vector form of the model is (Tanaka 1996)

$$y = Cu + \varepsilon \sim N\left(0, \sigma_\varepsilon^2 \left(I_T + \delta CC'\right)\right)$$

where

$$u = (u_1, \ldots, u_T)' ; \ \varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)' ; \ \delta = \frac{\sigma_u^2}{\sigma_\varepsilon^2};$$

with the random walk generating matrix

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}.$$

The log-likelihood function (without constant) is

$$L = -\frac{1}{2} \log \left|\sigma_\varepsilon^2 \left(I_T + \delta CC'\right)\right| - \frac{1}{2\sigma_\varepsilon^2} y' \left(I_T + \delta CC'\right)^{-1} y.$$

In order to allow the correlation in errors ε_t and u_t, let us assume

$$u_t = s\varepsilon_t + \sqrt{\frac{\delta - s^2}{\xi}} \eta_t \tag{2}$$

where s is a constant that represents the covariance of u_t and ε_t; shocks ε_t and η_t are independent:

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim N\left(0, \sigma_\varepsilon^2 \begin{bmatrix} \xi & 0 \\ 0 & 1 \end{bmatrix}\right); \ \xi = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}.$$

By (2), u_t and ε_t are now correlated,[3] thereby allowing us to find the likelihood function (see Online Appendix[4] for details). Since the variance of ε_t, σ_ε^2, is non-zero, it is convenient to find the concentrated likelihood function (with respect to σ_ε^2):

$$L(\delta, s) = -\frac{1}{2} \log |D| - \frac{T}{2} \log y' D^{-1} y,$$

[3] Alternatively, we can write the error processes as

$$\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{\delta - s^2}{\xi}} & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix},$$

so that shocks u_t and ε_t are contemporaneously correlated:

$$\begin{bmatrix} u_t \\ \varepsilon_t \end{bmatrix} \sim N\left(0, \sigma_\varepsilon^2 \begin{bmatrix} \delta & s \\ s & 1 \end{bmatrix}\right); \ s = \frac{\sigma_{u\varepsilon}}{\sigma_\varepsilon^2}.$$

[4] http://www.clas.wayne.edu/multimedia/usercontent/File/Economics/wada/tech_appendix_rho.pdf

where $D = I + \delta CC' + s(C + C')$. Using the concentrated likelihood function, we obtain the following three lemmas that clarify the properties of the likelihood function.

Lemma 1 *The first order conditions are given by:*

$$\frac{\partial L(\delta, s)}{\partial s} = -\frac{1}{2} trace\left\{D^{-1}\left(C + C'\right)\right\} + \frac{T}{2}\frac{y'D^{-1}(C + C')D^{-1}y}{y'D^{-1}y} = 0$$

and

$$\frac{\partial L(\delta, s)}{\partial \delta} = -\frac{1}{2} trace\left\{D^{-1}CC'\right\} + \frac{T}{2}\frac{y'D^{-1}CC'D^{-1}y}{y'D^{-1}y} = 0$$

Lemma 2 *At $\delta = s = 0$,*

$$\frac{\partial L(\delta, s)}{\partial \delta} = -\frac{T}{2}\left(\frac{(T+1)}{2} - \frac{y'CC'y}{y'y}\right)$$

and for $T \to \infty$, we have

$$T^{-2}\frac{\partial L(\delta, s)}{\partial \delta} \Rightarrow \frac{1}{2}\left[\int_0^1 [W(r)]^2 dr - \frac{1}{2}\right]$$

where "$\Rightarrow$" represents weak convergence in distribution; and $W(r)$ is the standard Wiener process on $C[0,1]$.

Lemma 3 *At $\delta = s = 0$,*

$$\frac{\partial L(\delta, s)}{\partial s} = -\frac{T}{2}\left(2 - \frac{y'(C + C')y}{y'y}\right)$$

and for $T \to \infty$, we have

$$T^{-1}\frac{\partial L(\delta, s)}{\partial s} \Rightarrow \int_0^1 W(r)\, dW(r) = \frac{1}{2}\left[W(1)^2 - 1\right].$$

3.2 The Restriction

We impose the restriction that the variance-covariance matrix of the errors is positive-semi-definite. In our case, this restriction is $\delta \geq s^2$. Figure 1 displays such a restriction on the parameter space: The horizontal axis represents s, the covariance parameter, while the vertical axis represents δ, the ratio of the variance of the trend errors to the variance of the cycle error. As one can see, the boundary is a parabola: along the quadratic curve, the correlation coefficient is -1 (when corresponding s is negative) or 1 (when corresponding s is positive), except for the point where $\delta = s = 0$, at which the correlation is undefined and at which the set of the true parameter values is located. It is obvious from Figure 1 that the true parameter values are, in fact, close by the restriction that implies a perfect correlation.

As is well known, the gradient at the true parameter values is not necessarily zero, since the gradient itself is composed of random variables. Because $E\int_0^1 [W(r)]^2 dr = 1/2$ and $E\left[W(1)^2\right] = 1$,[5] the gradient becomes zero only as the expected value, asymptotically. Still, the fact that the probability of $\chi^2(1)$ being less than unity is 0.683, implies $T^{-1}\partial L(\delta, s)/\partial s$ is often negative at the

[5] $E\int_0^1 [W(r)]^2 dr = \int_0^1 E[W(r)]^2 dr = E\int_0^1 r dr = \frac{1}{2}$; and $E[W(1)]^2 = E\left[\chi^2(1)\right] = 1$.

true parameter values. Hence, the likelihood function has its global maximum at somewhere other than the true parameter values; many times it can be found on the boundary of the parameter space with negative s. As an example, see Figure 2: The likelihood function does not have its global maximum at $\delta = s = 0$. It is important to keep in mind that increasing the sample size does not prevent one from finding the artificial perfect correlation.

4 An AR(p) Case

Our framework can be extended to an AR(p) case. To do so, first we modify our model:

$$y = Cu + \Phi^{-1}\varepsilon$$

where

$$\Phi^{-1} = \begin{bmatrix} 1 & 0 & & & 0 \\ \phi & 1 & 0 & & \\ \phi^2 & \phi & 1 & \ddots & \\ & & & \ddots & 0 \\ \phi^{T-1} & & \phi^2 & \phi & 1 \end{bmatrix}$$

for an AR(1) case. The log-likelihood function (concentrated with respect to σ_ε^2) is then

$$L(\delta, s, \Phi) = -\frac{1}{2}\log|A| - \frac{T}{2}\log y'\Phi' A^{-1}\Phi y \tag{3}$$

where $A = I + sC'\Phi' + s\Phi C + \delta\Phi CC'\Phi'$. Similar to the previous section, our strategy here is that we concentrate the likelihood function with respect to Φ, so that (3) is a function of (δ, s):

$$L(\delta, s, \Phi(\delta, s)) = -\frac{1}{2}\log|A(\delta, s)| - \frac{T}{2}\log y'\Phi(\delta, s)' A(\delta, s)^{-1}\Phi(\delta, s)\, y. \tag{4}$$

Note that $\Phi(\delta, s)$ is obtained by solving $\partial L(\delta, s, \Phi)/\partial\phi = 0$ with respect to ϕ.

Assuming DGP to be $y = \Phi^{-1}\varepsilon$, Lemmas 1-3 are altered to:

Lemma 4 *The first order conditions are given by:*

$$\frac{\partial L(\delta, s, \Phi(\delta, s))}{\partial s} = -\frac{1}{2} trace\left\{A^{-1}\left(C'\Phi' + \Phi C\right)\right\} + \frac{T}{2}\frac{y'\Phi' A^{-1}(C'\Phi' + \Phi C)A^{-1}\Phi y}{y'\Phi' A^{-1}\Phi y} = 0$$

and

$$\frac{\partial L(\delta, s, \Phi(\delta, s))}{\partial \delta} = -\frac{1}{2} trace\left\{A^{-1}\Phi CC'\Phi'\right\} + \frac{T}{2}\frac{y'\Phi' A^{-1}\Phi CC'\Phi' A^{-1}\Phi y}{y'\Phi' A^{-1}\Phi y} = 0$$

Lemma 5 *At $\delta = s = 0$ and for $T \to \infty$, we have*

$$T^{-2}\frac{\partial L(\delta, s, \Phi(\delta, s))}{\partial \delta} \Rightarrow \frac{(1-\phi)^2}{2}\left[\int_0^1 [W(r)]^2\, dr - \frac{1}{2}\right]$$

where "$\Rightarrow$" represents weak convergence in distribution; and $W(r)$ is the standard Wiener process on $C[0,1]$.

Lemma 6 *At $\delta = s = 0$ and for $T \to \infty$, we have*

$$T^{-1}\frac{\partial L\left(\delta, s, \Phi\left(\delta, s\right)\right)}{\partial s} \Rightarrow \left(1-\phi\right)\int_0^1 W\left(r\right)dW\left(r\right) = \frac{\left(1-\phi\right)}{2}\left\{\left[W\left(1\right)\right]^2 - 1\right\}.$$

Clearly, $\phi = 0$ is the previous case. Also, it is not difficult to extend the Lemmas noted above to an AR(2) model, which is argued in Morley et al. (2003).[6] For a finite sample, Figure 3 is computed as follows: first, $\widehat{\phi}$ is obtained as a function of (δ, s), i.e., by solving $\partial L\left(\delta, s, \Phi\right)/\partial\phi = 0$ with respect to ϕ, using the MATLAB function "`fzero`." Then, given $\widehat{\phi}$, the likelihood function is computed by (4).

5 Conclusion

When data are generated by a stationary process, the correlation in the error of a stochastic trend and errors of cycles in the state-space model is undefined because there is no stochastic trend. If one allows for such a correlation and estimate the parameters, the correlation parameter will be unidentified, and unexpectedly, the estimated parameters will almost always be 1 or -1. It is shown that the following two facts explain such a result: (i) We impose the restriction that requires the variance covariance matrix of the errors to be positive semi-definite (in other words, either the variance of the trend error is zero or the correlation is perfect). The set of the true parameter values lies on the boundary in the parameter space and in the neighborhood of parameters that imply perfect correlation. (ii) The likelihood function has its (constrained) global maximum at the true parameters only on average. Almost always its global maximum is on the boundary.

However, caution is necessary. Unlike Morley et al. (2011), whose proposed likelihood ratio test of stationarity compares the likelihood value with the restriction of $\delta = s = 0$ to the likelihood value without the restriction, our diagnosis principle does not fit the framework of a rigorous test of stationarity. This is because an estimated perfect correlation does not necessarily mean that the true DGP is stationary; it might as well be that the true DGP is non-stationary with a perfect correlation.

References

[1] Fuller, W.A. (1996) "*Introduction to Statistical Time Series*" John Wiley & Sons, Inc.

[2] Morley, J.C., C.R. Nelson and E. Zivot (2003) "Why are Beveridge-Nelson and Unobserved-Component Decompositions of GDP So Different?" *Review of Economics and Statistics* 85, 235-243.

[3] Morley, J.C., I. Panovska and T.M. Sinclair (2011) "A Likelihood Ratio Test of Stationarity Based on a Correlated Unobserved Components Model," RPF Working Paper 2008-11, George Washington University.

[4] Perron, P. and T. Wada (2009) "Let's Take a Break: Trends and Cycles in US Real GDP," *Journal of Monetary Economics* 56 (6) 749-765

[6] In an AR(2) case, for example, the coefficients of the limit distributions in Lemmas 5 and 6 need to be replaced by $(1-\phi_1-\phi_2)^2$ and $(1-\phi_1-\phi_2)$, respectively. See Online Appendix.

[5] Tanaka, K. (1996) "*Time Series Analysis*" John Wiley & Sons, Inc.

[6] Watson, M. (1986) "Univariate Detrending Methods with Stochastic Trend," *Journal of Monetary Economics* 18, 49-75.

6 Appendix

- Proof of Lemma2

$$\left.\frac{\partial L(\delta,s)}{\partial \delta}\right|_{\delta=s=0} = -\frac{T}{2}\left(\frac{(T+1)}{2}-\frac{y'CC'y}{y'y}\right) = -\frac{T}{2}\left(\frac{(T+1)}{2}-\frac{\varepsilon'CC'\varepsilon}{\varepsilon'\varepsilon}\right).$$

Since

$$T^{-1}\varepsilon'\varepsilon \underset{p}{\rightarrow} \sigma_\varepsilon^2$$

$$T^{-2}\varepsilon'CC'\varepsilon = T^{-2}\sum_{t=1}^{T} S_t^2 \Rightarrow \sigma_\varepsilon^2 \int [W(r)]^2\, dr$$

where

$$S_t = S_{t-1}+\varepsilon_t \quad \text{with } S_0 = 0;$$

and "$\underset{p}{\rightarrow}$" denotes convergence in probability. We have then

$$T^{-2}\left.\frac{\partial L(\delta,s)}{\partial \delta}\right|_{\delta=s=0} \Rightarrow \frac{1}{2}\left[\int_0^1 [W(r)]^2\, dr - \frac{1}{2}\right]$$

- Proof of Lemma 3

$$\left.\frac{\partial L(\delta,s)}{\partial s}\right|_{\delta=s=0} = -\frac{1}{2}trace\left(C+C'\right)+\frac{T}{2}\frac{y'\left(C+C'\right)y}{y'y} = -T+\frac{T}{2}\frac{\varepsilon'\left(C+C'\right)\varepsilon}{\varepsilon'\varepsilon} = -T+T\frac{\varepsilon'C\varepsilon}{\varepsilon'\varepsilon}$$

Note further that

$$T^{-1}\varepsilon'\varepsilon \underset{p}{\rightarrow} \sigma_\varepsilon^2$$

$$T^{-1}\varepsilon'C\varepsilon = T^{-1}\sum_{t=1}^{T} S_t\varepsilon_t = T^{-1}\sum_{t=1}^{T} S_{t-1}\varepsilon_t + T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \Rightarrow \frac{1}{2}\sigma_\varepsilon^2\left\{[W(1)]^2-1\right\}+\sigma_\varepsilon^2.$$

Hence,

$$T^{-1}\left.\frac{\partial L(\delta,s)}{\partial s}\right|_{\delta=s=0} \Rightarrow -1+\frac{1}{2}\left\{[W(1)]^2-1\right\}+1$$
$$= \frac{1}{2}\left\{[W(1)]^2-1\right\}.$$

- Proof of Lemma 4 (Sketch)

  The derivations are similar to those given in Lemma 1. Note that the Envelope theorem:

$$\frac{\partial L(\delta,s,\Phi(\delta,s))}{\partial\Phi(\delta,s)}\frac{\Phi(\delta,s)}{\partial\delta}=0;\quad \frac{\partial L(\delta,s,\Phi(\delta,s))}{\partial\Phi(\delta,s)}\frac{\Phi(\delta,s)}{\partial s}=0,$$

  is used to obtain the results.

- Proof of Lemma 5

  At $\delta = s = 0$, $A = I$ and $\partial A/\partial\phi = 0$ holds. Then, we can prove that the estimator for ϕ is a consistent estimator. To show this, noting that $\Phi = (1-\phi)I + \phi C^{-1}$, the first order condition for ϕ,

$$0=\frac{\partial L(\delta,s,\Phi)}{\partial\phi}=-\frac{1}{2}trace\left(A^{-1}\frac{\partial A}{\partial\phi}\right)-\frac{T}{2}\frac{-2trace\left(y'\frac{\partial\Phi'}{\partial\phi}A^{-1}\Phi y\right)-trace\left(\frac{\partial A}{\partial\phi}A^{-1}\Phi yy'\Phi' A^{-1}\right)}{y'\Phi' A^{-1}\Phi y},$$

  is simplified at $\delta = s = 0$ to

$$0=y'\frac{\partial\Phi'}{\partial\phi}\Phi y=y'\left(C^{-1}-I\right)'\Phi y,$$

  and hence,[7]

$$\widehat{\phi}=\frac{\sum_{t=2}^{T}y_{t-1}y_t}{\sum_{t=2}^{T}y_{t-1}^2}\underset{p}{\rightarrow}\phi.$$

  Since the least square estimator is consistent under the DGP process considered here (see Fuller 1996, for example), our estimator for the autoregressive parameters is consistent.

[7] The first order condition is now

$$\begin{aligned} y'\left(C^{-1}-I\right)'\Phi y &= \left(\begin{bmatrix} 0 & & & 0 \\ -1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & -1 & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}\right)'\left(\begin{bmatrix} 1 & & & 0 \\ -\phi & 1 & & \\ & \ddots & \ddots & \\ 0 & & -\phi & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}\right) \\ &= -\sum_{t=2}^{T}(y_t-\phi y_{t-1})\,y_{t-1} \end{aligned}$$

Next, we note that

$$\begin{aligned}\Phi CC'\Phi' &= \widehat{\phi}^2 I + \widehat{\phi}\left(1-\widehat{\phi}\right)\left(C+C'\right) + \left(1-\widehat{\phi}\right)^2 CC'; \\ trace\left(\Phi CC'\Phi'\right) &= \widehat{\phi}^2 T + \widehat{\phi}\left(1-\widehat{\phi}\right)2T + \left(1-\widehat{\phi}\right)^2 \frac{T\left(T+1\right)}{2}.\end{aligned}$$

Since $\widehat{\phi} \underset{p}{\rightarrow} \phi$, and the fact that DGP is $y = \Phi^{-1}\varepsilon$, the same argument as Lemma 2 leads to the limit distribution.

- Proof of Lemma 6

  The limit distribution is obtained since

$$\Phi C + C'\Phi' = \left(1-\widehat{\phi}\right)\left(C+C'\right) + 2\widehat{\phi}I; \quad trace\left(\Phi C + C'\Phi'\right) = 2T.$$

Table 1: Frequency Distributions of the Estimated Correlation ρ

Estimated ρ	Data Generating Process AR(0)	AR(1)	AR(2)
$\rho < -0.99$	710	654	711
$-0.99 \leq \rho \leq 0.99$	85	116	102
$0.99 < \rho$	205	230	187
$\lvert\rho\rvert > 0.99$	915 (91.5%)	884 (88.4%)	898 (89.8%)

Notes: 1) The estimated frequencies are are computed from the sample size $T = 500$ with 1,000 replications. 2) For the AR(1) model, DGP is $y_t = 0.9\, y_{t-1} + \epsilon_t$, where $\epsilon_t \sim i.i.d.N(0,1)$ and $y_0 = 0$. 3) For the AR(2) model, DGP is $y_t = 1.28\, y_{t-1} - 0.38\, y_{t-2} + \epsilon_t$, where $\epsilon_t \sim i.i.d.N(0,1)$ and $y_0 = y_{-1} = 0$. For more details, see Online Appendix.

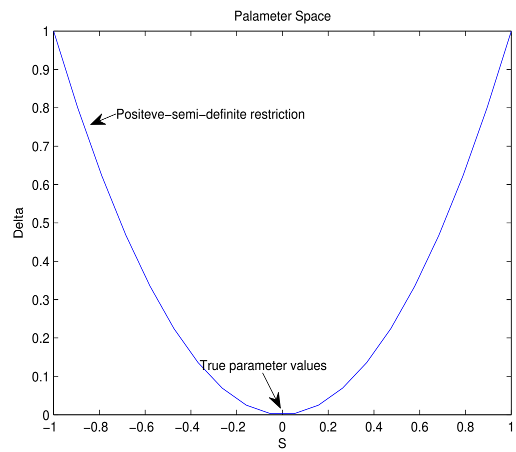

Figure 1: Parameter space and the restriction

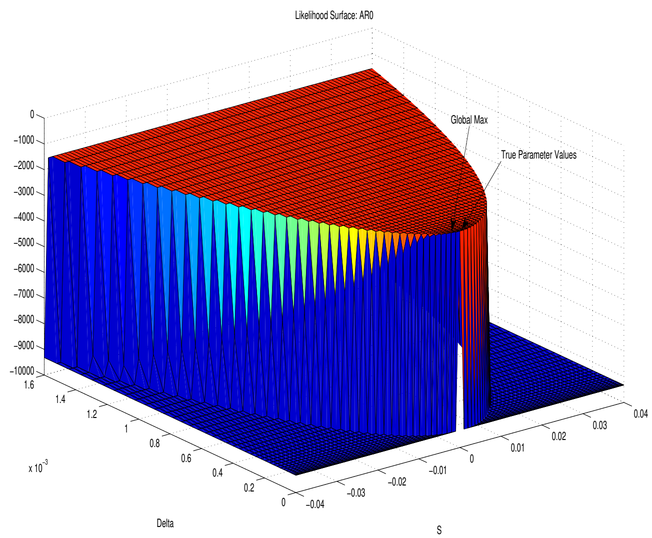

Figure 2: Typical Likelihood Surface: AR(0) with T=500.

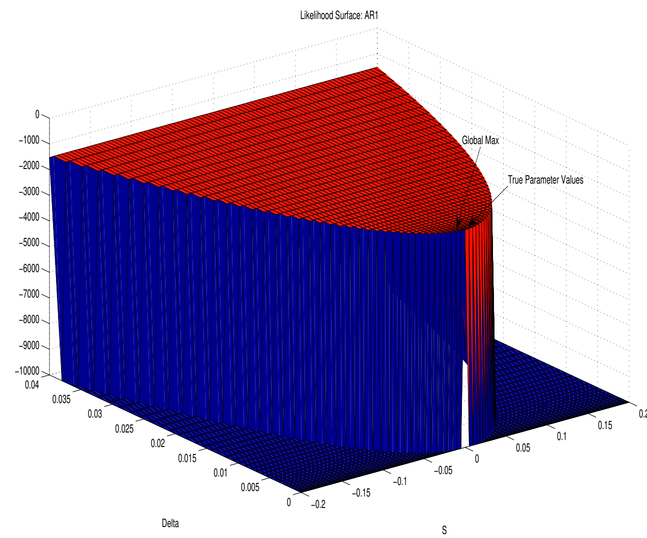

Figure 3: Typical Likelihood Surface: AR(1) with T=500.