Trends and Spurious Regressions

P. C. B. Phillips

NAKE Workshop report by Thijs Knaap, Groningen University.
March 1998.

1 Introduction

Time series analysis has changed profoundly in the last two decades or so. In this quiet field that once had the most clear, all-purpose, standard methodology in economics, the leading scientists no longer agree on fundamental premises. Professor Phillips, one of the original instigators of the methodological mayhem, gave a series of lectures that reflected this current state of affairs. The Box-Jenkins (1976) ARMA-analysis, once the universal tool, is only present at the sideline. A more general framework must capture that which was ignored by ARMA, but interpretation of the results is difficult. Finally, the oldest tool in econometrics, the OLS-estimator, saves the day when all the high-brow methods fail.

In this short, and necessarily incomplete, report all analysis will be asymptotic for two reasons: first, there usually is an infinity of parameters to estimate (think of serial covariance matrices) and secondly, in the words of professor Phillips, 'finite sample analysis is so darn difficult.'

In the next section we look trends. Section 2.1 deals with the information content of regressors. If we are able to quantify this measure we can see which processes may be recognized from the data, and we can explain why some parameters drown in the disturbances. Trend extraction is the subject of Section 2.2. In Section 3 we look at non-stationary processes and the behavior of standard estimators when confronted with them. Because all observations, and not just their covariances, matter, this analysis is carried out in function spaces. In the concluding section, we summarize results and try to make sense of what we found.

2 Trends

2.1 Signal and Noise

In a somewhat mechanistic way, time-series data y_t can be seen as the sum of two components, signal s_t and noise u_t. If both constituent parts are random variables, we may define the signal-to-noise ratio SNR:

$$\text{SNR} \equiv \frac{\text{var}\,(s_t)}{\text{var}\,(u_t)}.$$

This ratio can be computed for different processes, including all autoregressive and moving-average processes.

If we want to explain the data with a linear regression, that is, we write $y_t = X_t\beta + u_t$, we get a consistent estimate for β only if the variance of $\hat{\beta}$ goes to zero. For scalar X_t, there holds that $\text{var}(\beta) = \sigma_u^2 / \sum X_t^2$ so that consistency is guaranteed iff

$$\sum_{t=1}^{T} X_t^2 \to \infty \text{ as } T \to \infty.$$

This is the *excitation condition.* We can link this condition with the SNR for the process, which is

$$\text{SNR} = \frac{\beta^2 \text{var}\,(X_t)}{\text{var}\,(u_t)} = \frac{\beta^2 \frac{1}{T} \sum_{t=1}^{T} X_t^2}{\sigma_u^2}.$$

If the ratio either diverges or converges to a positive constant, we may estimate consistently. Three examples of this rule may be found in Table 1.

$y_t =$	SNR	**Estimate β ?**
$\beta y_{t-1} + u_t$	$\beta^2/(1-\beta^2)$	consistent
$\beta t + u_t$	$\beta^2 T(T+1)(2T+1)/6\sigma_u^2 \to \infty$	consistent
$\beta/t + u_t$	$\beta^2\pi^2/6T \to 0$	not consistent

Table 1: Signal in different processes

If the regressor does not satisfy the excitation condition, the limiting distribution of the estimator is nondegenerate. That is, the attained value of the estimator depends on the disturbances, even when the sample is infinitely big. If the condition is satisfied, the consistency of the estimate implies a degenerate limiting distribution.

From Table 1, it appears that for high powers of t in the regressor we get consistent estimates. This intuition is correct: when we evaluate $y_t = \beta/t^\alpha + u_t$ we find that the SNR is $\mathcal{O}\left(T^{1-2\alpha}\right)$. For $\alpha < 0.5$, the excitation condition is not satisfied.

The theory in this section can be generalized to the case where regressors have more than one dimension. The variance of the OLS estimate of β then is $\sigma_u^2(X'X)^{-1}$, so that the excitation condition becomes

$$\begin{cases} [\text{diag}(X'X)]_i \to \infty & \forall i \\ (X'X) \text{ invertible} & \end{cases}.$$

An example where the second part of this condition causes trouble is the equation $y_t = \beta' x_t + u_t$ with $x_t = \mu \cdot t + v_t$ where μ is a vector. The limit of $T^{-3}X'X$ with $[X]_{ij} = \mu_j \cdot i + v_i$ is the singular matrix $\frac{1}{3}\mu\mu'$. However, it can be shown that applying OLS to the above system *does* render consistent estimates, and even the usual χ^2-tests are valid. The same holds for cases with quadratic trends, more than one trend and in the presence of integrated regressors. The analysis breaks down, however, when there is nonstationarity in the autoregression. That is, a unit root will cause inconsistent estimates.

2.2 Trend Extraction

We know now that a trend must exhibit a certain level of variation in order to be estimable. From this perspective, it would seem that the use of powerful regressors should be advocated. Yet, as the following example shows, more powerful regressors can also be powerfully wrong if the wrong type of process is assumed.

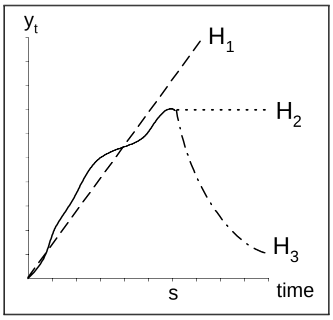

Figure 1: Three trend hypotheses and their predicted paths.

Consider Figure 1. We have data for y_t up until $t = s$; if we want to make a prediction about the values of y_t beyond that, we need to formulate a model for y_t. We consider three options:

1. H_1: $y_t = \beta t + u_t$, a linear trend. The regressor t provides lots of signal $[\mathcal{O}(T^3)]$ and we get narrow error bands around the predicted values.

2. H_2: $y_t = \sum_{i=1}^{t} u_i$, a Martingale process. Our prediction is $\hat{y}_{s+j} = y_s$, the last value attained. Because the process is all noise and no signal the error bands diverge.

3. H_3: $y_t = \theta y_{t-1} + u_t$, an autoregressive process. If the process is stationary, the confidence bands converge to $\sigma_u^2/(1-\theta^2)$, a constant.

Professor Phillips stressed that in a textbook exercise you would usually find the hypothesis spelled out for you, while in real analysis the problem is to find the proper model. As we see above, our estimates may go anywhere depending on the choice of H_i, as may the confidence bands. Even though some regressors are superior in signal, their choice can give misleading estimates.

3 Non-stationarity

3.1 OLS with a unit root

So far we have seen that estimation in time-series models is possible as long as there is enough information in the regressors. We made one reservation, concerning the cases in which a unit root is present. These cases will now be explored further.

Consider the simple AR(1) model $y_t = \theta y_{t-1} + u_t$. If $\theta < 1$, the signal is of $\mathcal{O}(T)$, and there holds

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \Rightarrow \mathcal{N}\left(0, 1 - \theta^2\right),$$

where $\Rightarrow$ is convergence in distribution. As $\theta \to 1$, we see that this distribution becomes degenerate. In fact, we have that $y_t = \sum u_t$ and $\text{var}(y_t) = \sigma_u^2 t$. The signal $\sum \text{var}(y_t)$ now becomes stronger, of order $\mathcal{O}(T^2)$, so that we must rescale and look at

$$T\left(\hat{\theta} - \theta\right) = \frac{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t}{\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2}. \tag{1}$$

We are no longer in a position to invoke simple limit theorems to bind the sums in this fraction. As y_t consists of the sum of all u_s, $s \leq t$, these errors do not average out and we must move our analysis to a space where we can handle all this data. That is why we consider the data as an object in a function space $\mathcal{F}$, where functions are defined on $[0, 1]$ and continuous from the right. With T datapoints we introduce the function $X_T(r) = y_{\lfloor Tr \rfloor}/\sqrt{T}$. Here, $\lfloor \cdot \rfloor$ is the *entier* or integer function. It is clear that $X_T(r)$ is an element of $\mathcal{F}$.

We may invoke the so-called Phillips-Solo device (Phillips and Solo, 1992) to assert that $X_T(r) \Rightarrow B(r)$, standard Brownian motion (BM hereafter). The distribution of many functionals of BM is known, so if we can rewrite formula (1) in terms of BM using the Phillips-Solo device we may be able to derive a distribution. The denominator can be modified as

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(X_T(t/T)\right)^2 \\ &= \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}} X_T^2(r)\,\mathrm{d}r\end{aligned}$$

$$= \int_0^1 X_T^2 (r) \,\mathrm{d}r.$$

We can call upon the continuous mapping theorem (Billingsley 1968, p.30) to state that the last expression converges in distribution to $\int_0^1 B^2 (r)\mathrm{d}r$.

The numerator of (1) can similarly be written as

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t &= \sum_{t=1}^{T} X_T (t/T) \frac{u_t}{\sqrt{T}} \\ &= \int_0^1 X_T (r) \,\mathrm{d}X_T (r) \\ &\Rightarrow \int_0^1 B (r) \,\mathrm{d}B (r) . \end{aligned}$$

where the last two steps are nontrivial. Thus, the distribution of the OLS estimator is (in shorthand notation):

$$T \left(\hat{\theta} - \theta\right) \Rightarrow \frac{\int B\mathrm{d}B}{\int B^2}. \tag{2}$$

This function of BM has a known distribution which may be used for inference.

3.2 Regressors in the presence of a unit root

Above we explored a simple AR(1) model and found a distribution for $\hat{\theta}$. In the usual ARMA-type of analysis, we could add regressors to this model (say, $y_t = X\beta + \theta y_{t-1} + u_t$) and still use the same distribution for $\hat{\theta}$. Matters are not as simple in the presence of a unit root.

First of all, because we are working in the function space $\mathcal{F}$, we have to give meaning to the term regression. If we make the assumption that $\mathcal{F}$ is a Hilbert space where $\int f^2 < \infty$ for all f in $\mathcal{F}$, we can define the inner product

$$< f, h >= \int_0^1 f(r) \cdot h(r) \,\mathrm{d}r \qquad f, h \in \mathcal{F}$$

and the norm $\|f\| \equiv< f, f >^{\frac{1}{2}}$. Regressors are functions in $\mathcal{F}$ that depend on certain parameters. For instance, a constant and a linear time trend could be represented by a function $g(\alpha, \beta, r) = \alpha + \beta r$. Regressing data $X(r)$ on g means we find α and β such that $\|X - g\|$ is minimized.

The problem is that when we prefilter data in which unit root processes might play a role, the filter affects the distribution of our autoregressive coefficient. Suppose we work with the model

$$y_t = \theta y_{t-1} + \beta t + \alpha + u_t$$

and we regress y_t on $X = [1, t]$. After that, we estimate θ. We do not know that in reality, $\theta = 1$. The distribution of the OLS-estimator of θ is

$$T\left(\hat{\theta} - 1\right) = \frac{\frac{1}{T}\mathbf{y}_{-1}^{'} Q_X \mathbf{u}}{\frac{1}{T^2}\mathbf{y}_{-1}^{'} Q_X \mathbf{y}_{-1}}$$

with $\mathbf{y}_{-1}$ and $\mathbf{u}$ the vector of lagged y_t and disturbances, and Q_X the projection matrix $\left(I - X\left(X'X\right)^{-1} X'\right)$. Using similar analysis as in the preceding section, we find that

$$T\left(\hat{\theta} - 1\right) \Rightarrow \frac{\int B_X \mathrm{d}B}{\int B_X^2 \mathrm{d}r} \tag{3}$$

with $B_X = B - \left(\int Bg'\right)\left(\int gg'\right)^{-1} g$ and $g(r) = (1, r)$. That is, B_X is detrended BM and the distribution in (3) is different from the distribution in (2).

Thus we come to the important conclusion that tinkering with the data will affect the distribution of the autoregressive estimator. Every time we want to test for a unit root in practice we need to reconstruct the test, using the regressors of interest. This is why Dickey and Fuller (1979) found different Monte Carlo distributions for their autoregressive estimator when used with a constant or with a constant and a trend.

3.3 The Bayesian alternative

Phillips and Ploberger (1996) present an elegant way out of the above problem of changing distributions. By using a Bayesian asymptotics, that do not rely on the (functional) Central Limit Theorems that break down in classical analysis, they are able to find consistent properties of models with a unit root.

The principle that makes Bayesian analysis work here is that analysis is only carried out after conditioning on the data. Because of that, the (type of) randomness of the data no longer plays any role in in estimators; we are left with functions only of the parameters.

4 Conclusion

In this report on professor Phillips' lectures at the NAKE workshop in Nijmegen we looked at two of the main themes, the treatment of trends in data and integrated processes. Because of the speed and versatility of the lecturer, it is not possible to give a close account of every main or auxiliary result that was derived. However, the main message must be clear from the above.

4.1 The themes

Trends in time series cannot always be extracted. If the variability of the trend is too small, the parameter can not be estimated consistently. With time trends, the borderline case is $t^{-\frac{1}{2}}$. Powerful trends are better regressors, but the researcher must be cautious. The different implications of various trend hypotheses warrant a conservative choice.

Integrated processes open up a new ball park in time series econometrics. The analysis must be carried out in a function space, rather than in $\mathbb{R}^n$. The distribution of various estimators changes when new regressors are brought in, so that confidence intervals must be computed over and over again. Bayesian analysis may be a helpful aid.

4.2 Discussion

Can nonsense regressions be of help in explaining the data? The analysis of the old problem of 'trend correlation' looks very clean today, and without doubt processes can be disected with much greater precision than thirty years ago. However, suppose we find that two variables are cointegrated, that is,

$$\begin{aligned} y_t &= \beta x_t + u_t \\ \Delta x_t &= \nu_t. \end{aligned}$$

We can now explain variation in y_t, estimate β, but have we explained the process very well? The term that drives the model, ν_t, is still modelled as an error term, of which we know the expectation and the variation, but not much more. This should be a clear indication that all is not well in time series analysis today, in spite of the sophisticated models of the 1990s.

A critical look at econometrics much along these lines may be found in Phillips (1996). It is argued there that what is called a 'spurious regression' today may also be viewed as a valid regression, if we take a wider view of the aim of econoemtrics. Phillips argues that the analysis must necessarily be deficient, and that hopes of the estimated model fitting anywhere beyond the sample are, well, hopes. If so-called spurious regressions seem to explain a serious amount of variation then this is, according to the author, an acceptable way of formailizing the data.

5 References

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, revised edition, Holden-Day

Dickey, D.A. and W.A. Fuller (1979), ‘Distribution of the estimators for autoregressive time series with a unit root,’ *Journal of the American Statistical Association* **74**, pp. 427-31

Phillips, P.C.B. (1996), ‘Spurious regression unmasked,’ *Cowles Foundation Discussion Paper #1135*, Yale University, Yale

Phillips, P.C.B. and W. Ploberger (1996), ‘An asymptotic theory of Bayesian inference for time series,’ *Econometrica* **64**, pp. 381-412

Phillips, P.C.B. and V. Solo (1992), ‘Asymptotics for linear processes,’ *Annals of Statistics* **20**, pp. 971-1001