Comment: Unit Root and Structural Changes in Tropical Sea Levels

Ngai Hang Chan and Wilfredo Palma

1. INTRODUCTION

This NRC report provides an informative account for researchers who are interested in the interaction between statistics and oceanography. The committee should be complimented for accomplishing such a difficult mission. While the report offers a wide spectrum of interesting activities in these two fields, this discussion is limited to a specific component, namely, the data smoothing aspect of oceanographic data. A main purpose of this note is to illustrate how time series techniques that are developed mostly for econometric problems may find applications in physical oceanography.

Specifically, the wind-driven numerical model of the equatorial sea level of the South Pacific given in Miller and Cane (1989) is considered. In Miller and Cane (1989), a Kalman filter approach is used to estimate and to forecast the sea level based on a linear wind-driven numerical model. This model is useful since it provides not only a systematic treatment of missing observations (temporal gaps), but also allows extrapolation to points where direct measurements are difficult (spatial gaps).

In this discussion, the so-called unit root and trend break asymptotics are employed to analyze the data given in Miller and Cane (1989). These techniques have been developed mainly by economists and statisticians to handle nonstationary data; see, for example, Nelson and Plosser (1982), Chan and Wei (1987), Phillips (1987) and Perron (1989), among others. A fundamental idea in this approach lies in incorporating the knowledge of an external event into the analysis of the underlying nonstationary pattern of the data. In this context, one may consider the El Niño phenomenon as an intervention to the sea level. The series can then be divided into two parts. The first part consists of the data obtained prior to the El Niño effect while the second part consists of the post–El Niño data. Although a random walk model is first found to be adequate, further trend break analysis reveals that the unit root nonstationarity may be an artifact attributed to the El Niño effect. This seems to be the first analysis applying trend break techniques to a noneconometric data set. In contrast with the computationally intensive Kalman filter approach, a simple stationary first-order autoregressive series is obtained which gives an efficient model for estimation and prediction. It is hoped that this discussion provides an illustration on the statistical analysis of some of the issues raised in the NRC report.

This note is organized as follows. A description of the data set is given in Section 2, the unit root and trend break asymptotics are reviewed in Section 3 and their applications to analyzing the sea level data are given in Section 4. Concluding remarks are listed in Section 5.

2. THE DATA

The sea level data in Miller and Cane (1989) contains observations of the sea level heights for several islands in the tropical Pacific. It consists of monthly means with tides removed. That is, each data point is the average of approximately 720 (30 × 24) hourly observations for each month with the effect of tides removed. The following islands are of interest: Rabaul, Nauru, Jarvis, Christmas, Santa Cruz, Callao, Kapingamarangi, Tarawa, Canton and Fanning. These islands, shown in Figure 1, are chosen because they provide the longest overlapping time series. For Rabaul, Nauru, Christmas and Tarawa, the observations were taken from 1974 to 1983; for Jarvis from 1977 to 1984; for Santa Cruz from 1978 to 1983; for Callao from 1942 to 1984; for Kapingamarangi from 1978 to 1983; and for Fanning from 1972 to 1983. In this way, the overlapping period is 1978–1983.

The measurements have been filtered in the following way. For a particular island, let Z_{ij} denote the observation at the jth hour of the ith month with tide effects removed $i = 1, \ldots, 72\,(12 \times 6)$, $j = 1, \ldots, 720\,(30 \times 24)$. Define Z_i to be mean of the ith month:

$$Z_i = \frac{\sum_{j=1}^{720} Z_{i,j}}{720}.$$

Ngai Hang Chan is Associate Professor and Wilfredo Palma is Ph.D. Candidate of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

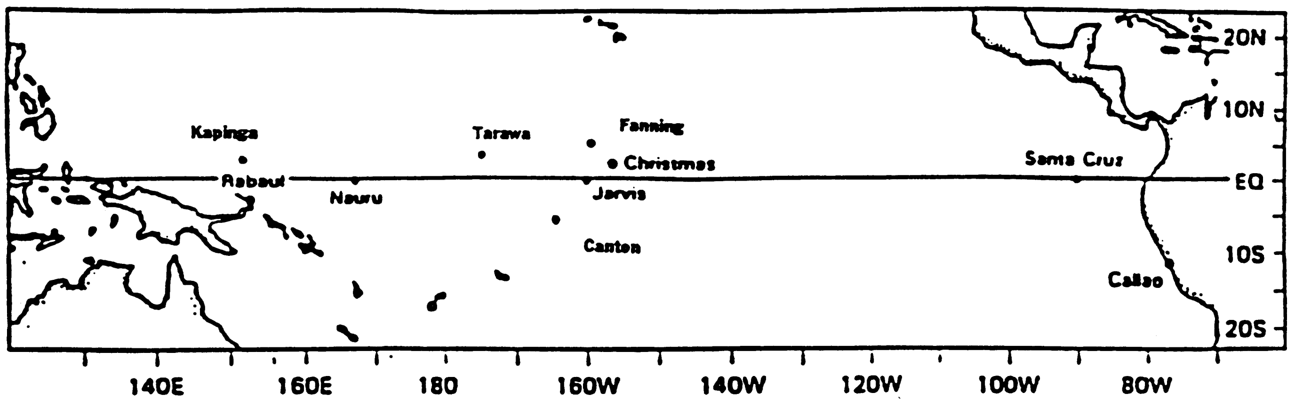

FIG. 1. *Locations of islands.*

There are six years of data $Z_1, \ldots, Z_{72}$, where Z_1 denotes January 1978 and Z_{72} denotes December 1983. Define the *monthly mean* for the ith month as the yearly average, for $i = 1, \ldots, 12$, that is,

$$\overline{Z}_i = \frac{\sum_{k=1}^{6} Z_{i+12(k-1)}}{6}.$$

Finally, define the *monthly anomaly* Y_i, for $i = 1, \ldots, 71$, as

$$Y_i = Z_i - \overline{Z}_{i(\mathrm{mod}\ 12)}.$$

For example, $\overline{Z}_3 = (Z_3 + Z_{15} + Z_{27} + Z_{39} + Z_{51} + Z_{63})/6$ and $Y_{14} = Z_{14} - \overline{Z}_2$. After this filtering process, the data Y_t consists of *deviations* from *monthly means*, expressed in centimeters.

Miller and Cane (1989) develop a Kalman filter model for the sea level data. The data and their predictions are shown in Figure 2. Together with the data, the Kalman filter is used to forecast missing observations. The series $\{Y_t\}$ is referred as the raw data and the Kalman filtered series is referred as the Kalman filtered data herein.

3. UNIT ROOT AND TREND BREAK

In the simplest form, a unit root problem concerns the model

$$y_t = \beta y_{t-1} + \varepsilon_t, \tag{3.1}$$

where $|\beta| \le 1$, $y_0 = 0$ and $\varepsilon_t \sim \mathrm{iid}(0, \sigma^2)$. The main interest lies in testing H: $\beta = 1$, which bears important economic consequences. For a recent survey of empirical studies about this problem, see Stock and Watson (1988). One can estimate the unknown parameter β by its least squares estimate $\widehat{\beta} = \sum_{t=1}^{n} y_t y_{t-1} / \sum_{t=1}^{n} y_{t-1}^2$. Under H: $\beta = 1$, it can

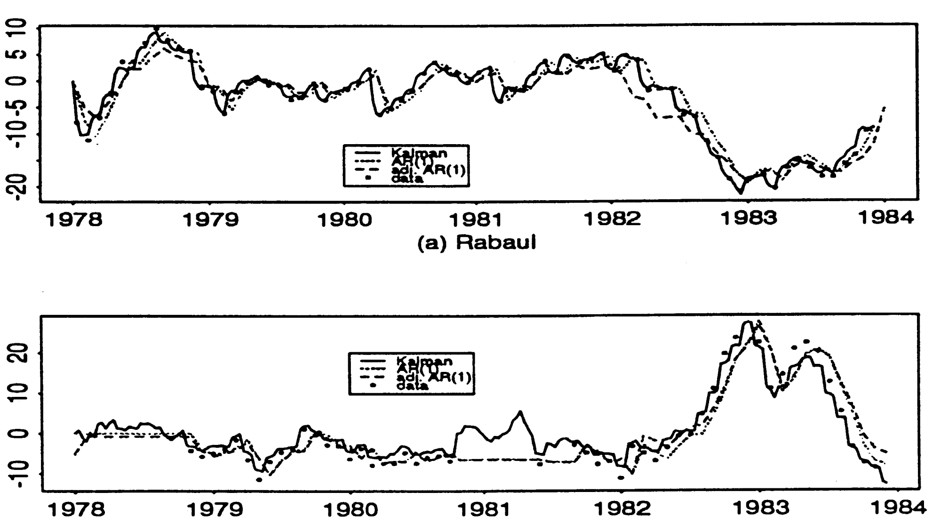

FIG. 2. *Sea height levels* (1978:1–1983:12): *heavy line, Kalman filter; dotted line,* AR(1) *model with* $\beta = 0.93$; *broken line,* AR(1) *model with* $\beta = 0.62$ *for* (a) *and* $\beta = 0.83$ *for* (b), *adjusted by the trend break; dots, raw data.*

be shown that

$$
\begin{aligned}
\tau_n &:= \left(\sum_{t=1}^{n} y_{t-1}^2\right)^{1/2} (\widehat{\beta} - \beta) \\
&\to_{\mathcal{L}} \frac{\int_0^1 W(t)\,dW(t)}{\left(\int_0^1 W(t)^2\,dt\right)^{1/2}} \qquad (3.2)\\
&:= \mathcal{L}(0),
\end{aligned}
$$

where $W(t)$ is a standard Brownian motion and $\mathcal{L}$ designates convergence in distribution as the sample size n tends to infinity. A unified treatment on the transition of $|\beta| < 1$ to β equals 1 for the distribution of τ_n can be found in Chan and Wei (1987). It suffices to point out that instead of classical asymptotic normality, nonstandard limiting results involving stochastic integral of Brownian motion such as $\mathcal{L}(0)$ are rules rather than exceptions in unit root problems. $\mathcal{L}(0)$ is also known as the Dickey–Fuller statistic, and its limiting percentiles can be found in Fuller (1976). While model (3.1) can be used to capture random change in the trend of the data, it has been argued that a deterministic change in the mean of y_t can also produce an artifact of $\beta = 1$. In particular, Perron (1990) proposes to consider the following model

$$y_t = \gamma D(T_B)_t + \beta y_{t-1} + \varepsilon_t, \qquad (3.3)$$

where $D(T_B)_t = 1$ if $t = T_B + 1$ and zero otherwise, and $\varepsilon_t \sim \text{iid}(0, \sigma^2)$. Under H: $\beta = 1$, (3.3) states that the mean of y_t is zero (y_0 is assumed to be zero) for $t = 1, \dots, T_B$ and γ afterward, whence allowing a one-time change in the mean of the series. Under the alternative hypothesis, the series follows

$$y_t = \mu + \gamma DU_t + \varepsilon_t, \qquad (3.4)$$

where $DU_t = 0$ if $t \le T_B$ and 1 otherwise. Under (3.4), there is also a shift in the mean of y_t but there is no unit root. The break point time T_B is assumed to be known and is expressed as a fraction of the sample size, that is, $T_B = \lambda n$, $\lambda \in [0, 1]$. Perron observes that if a trend break is not accounted for, one could hardly reject the unit root hypothesis. To circumvent this difficulty, he proposes to test for $\beta = 1$ by adjusting for the trend break effect presented in (3.3) and (3.4). Specifically, let $\{\widetilde{y}_t\}$ be the residuals of y_t regressing on μ and the indicator DU_t. Let $\widetilde{\beta}$ be the LSE of β from

$$\widetilde{y}_t = \widetilde{\beta}\widetilde{y}_{t-1} + \widetilde{e}_t. \qquad (3.5)$$

Suppose the model is generated by (3.3) with $\beta = 1$. Then as $n \to \infty$,

$$
\begin{aligned}
\widetilde{\tau}_n &:= \left(\sum_{t=1}^{n} \widetilde{y}_{t=1}^2\right)^{1/2} (\widetilde{\beta} - 1) \qquad (3.6)\\
&\to_{\mathcal{L}} H\left[\lambda(1-\lambda)K\right]^{-1/2},
\end{aligned}
$$

where H and K are stochastic integrals of $W(t)$; see Theorem 2 of Perron (1989). In particular, $\widetilde{\tau}_n$ of (3.6) can be used to distinguish (3.4) from (3.3). Owing to the abrupt change of the sea level shown in Figure 2, Perron's procedure seems to be an adequate device to capture a trend break.

4. SEA LEVEL DATA ANALYSIS

From Figure 2, a trend break is noted in the sea levels starting in 1982. At Rabaul (Figure 2a), the sea level drops after the first month of 1982 and rises at the beginning of 1983. At Santa Cruz (Figure 2b), the sea level rises at the beginning of 1982 and drops by the end of that year. Thus, there are different nonstationary patterns in the data. A probable cause of the trend break, as noted by Cane (1984), is the El Niño effect. This is a global weather disturbance phenomenon which warms the Pacific Ocean periodically. For the data set considered here, this effect appears in 1982 with declining strength after 1983. Figure 3 shows the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sea level at Rabaul. The ACF of the series decays very slowly, indicating possible nonstationary patterns in the data. The PACF plot shows a strong first-order component, suggesting a first-order autoregressive process AR(1). To account for these two facts, a first approach is to consider the difference of the series, resulting in a nonstationary random walk model

$$y_t = \beta y_{t-1} + \varepsilon_t, \qquad (4.1)$$

where $\beta = 1$ and $\{\varepsilon_t\}$ is a white noise sequence with variance σ^2. Table 1 shows that the variance of the differenced series, $\widehat{\sigma}^2$, is smaller than the variance of the original series. Further analysis of the residuals of the model (4.1) indicates a good fit of the data. Despite this fact, a random walk model for the sea level would possess an increasing variance, which seems to be incompatible with a stationary physical model of the ocean.

In order to assess the adequacy of the random walk model, the Dickey–Fuller statistic is applied to test for H: $\beta = 1$ in (4.1). Table 2 presents the results of the Dickey–Fuller test. The *long series* is the time series from January 1978 to December 1983 and the *short series* is the time series from January 1978 to

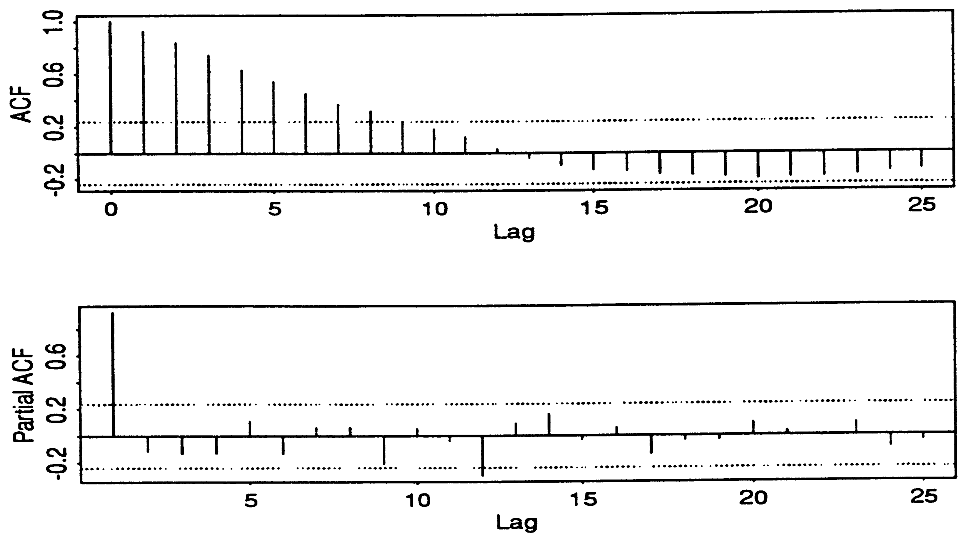

FIG. 3. *Autocorrelations and partial autocorrelations for Rabaul.*

TABLE 1
Variances of the original and the differenced series

Series	Rabaul	Nauru	Jarvis	Chris.	S. Cruz	Callao	Kapin	Tar.	Canton	Fan.
Orig.	61.6	101.4	47.6	78.5	81.3	61.7	26.5	46.3	36.9	50.6
Diff.	8.44	34.9	13.5	18.7	15.9	20.9	21.4	30.9	25.3	11.1

TABLE 2
Full-sample and truncated-sample Dickey–Fuller test

		Data		Kalman data	
Island	Series	$\widehat{\beta}$	τ_n	$\widehat{\beta}$	τ_n
Rabaul	Long	0.93	−1.61	0.99	−0.995
	Short	0.72	−2.92	0.93	−2.18
Nauru	Long	0.89	−1.59	0.98	−1.51
	Short	0.65	−2.53	0.94	−1.88
Jarvis	Long	0.90	−1.55	0.97	−1.58
	Short	0.72	−2.27	0.92	−2.29
Christmas	Long	0.91	−1.40	0.98	−1.27
	Short	0.69	−2.54	0.92	−2.40
Santa Cruz	Long	0.93	−1.27	0.98	−1.40
	Short	0.45	−3.28	0.93	−2.29
Callao	Long	0.83	−2.48	0.96	−2.03
	Short	0.38	−4.81	0.82	−3.83
Kapinga	Long	0.66	−3.07	0.996	−0.43
	Short	0.19	−4.94	0.99	−0.79
Tarawa	Long	0.92	−1.03	0.98	−1.34
	Short	0.84	−1.31	0.96	−1.56
Canton	Long	0.80	−2.29	0.98	−1.55
	Short	0.61	−3.03	0.97	−1.44
Fanning	Long	0.94	−0.901	0.97	−0.163
	Short	0.78	−2.24	0.93	−2.27

TABLE 3
Full-sample Perron test with changing mean

	Data		Kalman data	
Island	$\widehat{\beta}$	$\tilde{\tau}_n$	$\widehat{\beta}$	$\tilde{\tau}_n$
Rabaul	0.620	−3.95	0.947	−2.38
Nauru	0.695	−3.68	0.965	−1.94
Jarvis	0.885	−1.712	0.973	−1.59
Christmas	0.903	−1.55	0.981	−1.18
Santa Cruz	0.883	−1.56	0.973	−1.23
Callao	0.686	−3.22	0.941	−2.25
Kapinga	0.273	−5.23	0.985	−1.23
Tarawa	0.801	−1.82	0.972	−1.71
Canton	0.653	−3.87	0.975	−1.60
Fanning	0.909	−1.39	0.969	−1.65

July 1981. As shown in Table 2, for the long series of the raw data, the hypothesis of unit root cannot be rejected at 5% significance level for most of the islands, excepting Callao, Kapingamarangi and Canton. For the short series, the unit root hypothesis is rejected at 5% significance level in all islands except Tarawa. A similar conclusion is drawn from the tests for the Kalman filtered data. Since the Kalman filtered data can be used to interpolate for missing observations, its consistency with the raw data set provides further arguments against a random walk model for the long series.

The preceding analysis suggests that the El Niño phenomenon affects the mean of the series in 1982–83. This structural change causes the unit root behavior for several islands. To further study the effect of the El Niño, the Perron procedure of Section 3 is employed. Table 3 presents the results of the Perron tests applied to the sea level data. Percentiles of the Perron statistic $\tilde{\tau}$ are reported in Perron (1990). Under the null hypothesis, the time series has a unit root and suffers from a one-time change in the mean at time T_B. Under the alternative hypothesis, the time series has a one-time change in the mean but there is no unit root. As seen in Table 3, for the raw data, the null hypothesis of a structural change and a unit root is rejected at the 5% significance level for Rabaul, Nauru, Kapingamarangi and Canton. The sea levels of these islands exhibit similar behavior during 1982–83. They drop at the beginning of 1982 and rise in 1983. This observation indicates that the unit root phenomenon found in these islands by means of the Dickey–Fuller test alone may due to an artifact of the changing mean caused by the El Niño effect. For the Kalman filtered data, the null hypothesis of a changing mean and a unit root cannot be rejected at the 5% significance level for any islands by the Perron test. This indicates that the unit root structure of the Kalman filtered data may come from a source that is more complex than the one-time changing mean approach of Perron.

5. CONCLUDING REMARKS

Even though a random walk model seems to fit the sea level data well, the Dickey–Fuller test on the short series reveals that the unit root behavior of the long series may be attributed to the trend break of 1982. Further analysis of the raw data using Perron's tests shows that, for Rabaul and Nauru, the unit root behavior is an artifact of the El Niño effect. For these islands, after removing the changing mean effect, a stationary autoregressive model is found.

Figure 2 shows the forecasts of the unit root model and a stationary AR(1) model with the trend break adjusted. The unit root model has an estimated parameter $\beta = 0.93$ for Rabaul and Santa Cruz. For the trend break model with the trend $\mu + \gamma DU_t$ adjusted, a stationary AR(1) model with $\beta = 0.62$ is found for Rabaul and $\beta = 0.83$ is found for Santa Cruz. For simplicity, an autoregressive parameter $\beta = 0.62$ was used for both islands in Figure 2.

The Dickey–Fuller test results are similar when the Kalman filtered data is considered. The Perron test indicates, however, that the unit root cannot be explained by a changing mean structure of the filtered series. The filtered data seems to possess both a unit root and changing means. Further analysis of the Kalman model is currently being undertaken.

ACKNOWLEDGMENT

Research supported in part by the Office of Naval Research Grant N00014-89-J-1851.