Supporting Text

Autocorrelation and Partial Autocorrelation Functions. The lag-ℓ autocorrelation r_ℓ of a time series $\{y_t\}_{t=1}^n$ is

$$r_\ell = \frac{\sum_{t=\ell+1}^n (y_t - \overline{y})(y_{t-\ell} - \overline{y})}{\sum_{t=1}^n (y_t - \overline{y})^2} \qquad 0 \leq \ell \leq n-1$$

where $\overline{y} = (\sum_{t=1}^n y_t)/n$. The 5% significance level, $\sigma(k)$, of the kth autocorrelation in $\{r_\ell\}_{\ell=1}^{20}$ is computed as v

$$\sigma(k) = 2 \cdot \left[\left(1 + 2 \cdot \sum_{\ell=1}^{k-1} r_\ell^2 \right) /n \right]^{1/2}$$

The lag-ℓ partial autocorrelation of an observed series $\{y_t\}_{t=1}^n$ is related to $\phi_{\ell,\ell}$ in the order-ℓ autoregressive model fit to the data by linear regression:

$$y_t = \phi_{\ell,0} + \phi_{\ell,1} y_{t-1} + \cdots + \phi_{\ell,\ell} y_{t-\ell} + \epsilon_t$$

Thus, $\phi_{\ell,\ell}$ significantly different from zero suggests that an order-ℓ autoregressive model may be recommended. The partial autocorrelation p_ℓ may be computed from the sequence r_ℓ of previously computed autocorrelations by letting $p_{1,1} = r_1$, $v_1 = 1 - r_1^2$ and iteratively solving for $k = 1, \cdots \ell - 1$

$$\begin{aligned} p_{k+1,k+1} &= (r_{k+1} - p_{k,1} r_k - p_{k,2} r_{k-1} - \cdots - p_{k,k} r_1)/v_k \\ p_{k+1,j} &= p_{k,j} - p_{k+1,k+1} p_{k+1,k+1-j}, \qquad j = 1, \cdots k \\ v_{k+1} &= v_k \left(1 - p_{k+1,k+1}^2 \right) \end{aligned}$$

Then $p_\ell = p_{\ell,\ell}$. The 5% significance level of the k<u>th</u> partial autocorrelation is $2/\sqrt{n}$. The lag-1 autocorrelation and lag-1 partial autocorrelation are equal.

Parameter estimation. For a given set of macroevolutionary data, y_1, y_2, $\cdots$, y_n, the parameters of equations

$$y_t = \rho y_{t-1} + \epsilon_t \qquad [\mathbf{1}]$$

$$y_t = a_0 + \rho_a \, y_{t-1} + \epsilon_t \qquad [\mathbf{2}]$$

$$y_t = b_0 + b_1 \, (t - n/2) + \rho_b \, y_{t-1} + \epsilon_t, \qquad [\mathbf{3}]$$

are estimated by linear regression as follows (1).

Define $n-1$ dimensional vectors

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \mathbf{t} = \begin{bmatrix} 1 - n/2 \\ 2 - n/2 \\ \vdots \\ n - 1 - n/2 \end{bmatrix} \qquad \mathbf{Y_t} = \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{Y_{t-1}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}.$$

Define the matrices

$$\mathbf{U_1} = [\, \mathbf{Y_{t-1}} \,] \qquad \mathbf{U_2} = [\, \mathbf{1} \;\; \mathbf{Y_{t-1}} \,] \qquad \mathbf{U_3} = [\mathbf{1} \;\; \mathbf{t} \;\; \mathbf{Y_{t-1}}].$$

Using three independent linear regressions, the parameters of Eqs. **1-3** are estimated to be

$$\left[\, \hat{\rho} \,\right] = (\, \mathbf{U_1}'\mathbf{U_1} \,)^{-1} \, \mathbf{U_1}' \, \mathbf{Y_{t-1}}$$

$$\begin{bmatrix} \hat{a}_0 \\ \hat{\rho}_a \end{bmatrix} = (\, \mathbf{U_2}'\mathbf{U_2} \,)^{-1} \, \mathbf{U_2}' \, \mathbf{Y_{t-1}}$$

$$\begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{\rho}_b \end{bmatrix} = (\, \mathbf{U_3}'\mathbf{U_3} \,)^{-1} \, \mathbf{U_3}' \, \mathbf{Y_{t-1}}$$

The transpose and inverse of a matrix A are denoted by A' and A^{-1}, respectively, and we denote the matrix $(\, \mathbf{U_k}' \, \mathbf{U_k} \,)^{-1}$ by $\mathbf{c_k}$.

The residual mean squares of the three previous regressions are, respectively,

$$S_{ek}^2 = (n - 1 - k)^{-1} \left[\mathbf{Y_t'} \left(\mathbf{I} - \mathbf{U_k} \, (\mathbf{U_k}'\mathbf{U_k})^{-1} \, \mathbf{U_k}' \right) \mathbf{Y_t} \right] \qquad [\mathbf{4}]$$

for $k = 1, 2,$ and 3. The SE, $S_{e,x}$, of a parameter, x, estimated in one of the previous regressions is $\left(S_{ek}^2 \mathbf{c_k}(i,i)\right)^{1/2}$, where k is the number of the equation in which x appears and i is the index of x within the regression vector for that equation.

Regression t statistics for the conditions $\hat{a}_0 \neq 0$, $\hat{b}_0 \neq 0$, and $\hat{b}_1 \neq 0$ are

$$t_{\hat{x}} = \hat{x}/S_{e,x} \qquad [\mathbf{5}]$$

where x is a_0, b_0, or b_1. Regression t statistics for the hypotheses $\rho = 1$, $\rho_a = 1$, or $\rho_b = 1$, depending on which of the Eqs. **1-3** has been selected, are

$$t_{\hat{x}} = (\hat{x} - 1)/S_{e,x} \qquad [\mathbf{6}]$$

where x is ρ, ρ_a, or ρ_b.

Modifications by Phillips and Perron. Dickey and Fuller (1,2) analyzed the regression t statistics in Eqs. **5** and **6** for time series data. They assumed in their analysis that the innovations $\{\epsilon_t\}$ are normal, independent and of constant variance σ^2. Phillips (3) and Phillips and Perron (4) greatly relaxed this assumption about $\{\epsilon_t\}$ and allowed a weak dependence between the members of $\{\epsilon_t\}$ and slowly changing variance σ_t^2. They defined a modified t statistic, $Z(t_{\hat{x}})$, which has the same limiting distribution as $t_{\hat{x}}$. A summary of the equations defining the t statistics $Z(t_{\hat{x}})$ follows.

Phillips (3) Phillips and Perron (4) allowed innovations, ϵ_t, satisfying the following conditions (E denotes expected value):

1. $E(\epsilon_t) = 0$
2. $\sup_t E(|\epsilon_t|^{\beta+\gamma}) < \infty$ for some $\beta < 2$ and $\gamma > 0$.
3. As $n \to \infty$, $\sigma^2 = \lim E(n^{-1} S_n^2)$ exists and $\sigma^2 > 0$, where $S_n = \epsilon_1 + \cdots + \epsilon_n$.
4. $\{\epsilon_t\}$ is strong mixing with mixing coefficients α_m that satisfy $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$.

These conditions insure the existence of

$$\sigma_\epsilon^2 = \lim_{n\to\infty} \frac{1}{n} \sum{}_{t=1}^{n} E(\epsilon_t^2). \qquad [\mathbf{7}]$$

The Phillips and Perron (4) conditions 1-4 greatly expand the collection of time series that may be considered over those for which the innovations ϵ_t are normal, independent, and of constant variance, and the reader is referred to their paper for discussions of the conditions. They refer to Hall and Heyde (5) for discussion of strong mixing and mixing coefficients.

Phillips and Perron (4) define the parameter

$$\lambda = \frac{1}{2}(\sigma^2 - \sigma_\epsilon^2)$$

With the Dickey and Fuller (1,2) hypothesis that the ϵ_t are independent random variables with mean zero and constant variance, that constant variance is σ^2 and also is σ_ϵ^2, so that $\lambda = 0$.

Phillips and Perron (4) estimate σ_ϵ^2 to be the SE of the regressions, S_{e1}^2, S_{e2}^2, or S_{e3}^2, corresponding to the model 1, 2, or 3 that is being fitted.

They estimate σ^2 by $S_{nk,j}^2 \doteq \sigma^2$ where

$$S_{nk,j}^2 = (n-1)^{-1} \sum_{t=1}^{n-1} u_j(t) + 2(n-1)^{-1} \sum_{s=1}^{k} \left(1 - \frac{s}{k+1}\right) \sum_{t=s+1}^{n-1} u_j(t)u_j(t-s),$$

n is the length of the time series, and $u_j(t)$ is the residual of the regression: $u_1(t) = y_{t+1} - \hat{\rho} y_t$, $u_2(t) = y_{t+1} - \hat{a}_0 - \hat{\rho}_a y_t$ and $u_3(t) = y_{t+1} - \hat{b}_0 - \hat{b}_1(t - n/2) - \hat{\rho}_b y_t$. The selection of k in $\sum_{s=1}^{k}$ is left to the user. Phillips and Perron (4), p 343, suggest that for series of length 100, $k = 8$ is appropriate. The Phanerozoic time series that we analyze are of lengths 90 or 108, and we have used $k = 8$.

Some additional parameters are defined, with $\sum = \sum_{t=1}^{n}$.

$$m_{yy} = n^{-2} \sum y_t^2 \qquad \overline{y} = n^{-1} \sum y_t \qquad \overline{m}_{yy} = n^{-2} \sum (y_t - \overline{y})^2$$

$$m_y = n^{-3/2} \sum y_t \qquad m_{ty} = n^{-2.5} \sum t y_t$$

$$M = \left(1 - \frac{1}{n^2}\right) m_{yy} - 12 m_{ty}^2 + 12\left(1 + \frac{1}{n}\right) m_{ty} m_y - \left(4 + \frac{6}{n} + \frac{2}{n^2}\right) m_y^2$$

$$\lambda_j' = \lambda_j / S_{nk,j}^2 \qquad j = 1, 2, 3$$

The equation for $Z(t_{\hat{\rho}})$ is based on Phillips (3), Eq 22. The remaining equations are from Phillips and Perron (4).

$$Z(t_{\hat{a}_0}) = (S_{e2}/S_{nk,2})\, t_{\hat{a}_0} + \lambda_2' S_{nk,2} m_y \left[\overline{m}_{yy}\, m_{yy}\right]^{-1/2}$$

$$Z(t_{\hat{b}_0}) = (S_{e3}/S_{nk,3})\, t_{\hat{b}_0} - \lambda_3' S_{nk,3} m_y \left[M(M + m_y^2)\right]^{-1/2}$$

$$Z(t_{\hat{b}_1}) = (S_{e3}/S_{nk,3})\, t_{\hat{b}_1} - \lambda_3' S_{nk,3} \left(\frac{1}{2} m_y - m_{ty}\right) \left[\overline{m}_{yy}\, M/12\right]^{-1/2}$$

$$Z(t_{\hat{\rho}}) = (S_{e1}/S_{nk,1})\, t_{\hat{\rho}} - \lambda_1' S_{nk,1}\, (m_{yy})^{-1/2}$$

$$Z(t_{\hat{\rho}_a}) = (S_{e2}/S_{nk,2})\, t_{\hat{\rho}_a} - \lambda_2' S_{nk,2}\, (\overline{m}_{yy})^{-1/2}$$

$$Z(t_{\hat{\rho}_b}) = (S_{e3}/S_{nk,3})\, t_{\hat{\rho}_a} - \lambda_3' S_{nk,3}\, M^{-1/2}$$

Under the Dickey and Fuller hypothesis (1,2) that ϵ_t be of constant variance, then $Z(t_{\hat{x}}) = t_{\hat{x}}$ because $S_{ej} = S_{nk,j}$, $\lambda_j = 0$ and therefore $\lambda_j' = 0$ for $j = 1, 2, 3$.

Analysis of estimated Phanerozoic CO_2 levels. We illustrate the previous methods using the time series of estimates of Phanerozoic CO_2 levels (6) which was shown (7) to be significantly correlated with Sepkoski's fractional origination rates (8,9). The CO_2 estimates are evenly spaced at 10-Myr intervals based on a time scale in which the Cambrian began 570 mya. The time series has 58 numbers and to compare with Sepkoski's data (8,9) we have linearly rescaled the time scale to 545-0 mya. Time series of length 58 are short for random walk analysis, but within this limitation we will find that the null hypothesis that the time series is a random walk is not rejected. Analysis of windows of $\approx$ 450 Myr shows that the time series in the early windows is a random walk and in the more recent windows the time series is not a random walk.

The Phanerozoic CO_2 estimates are presented (Fig. 3) as RCO_2, the ratio of the estimate of partial pressure of CO_2 at a time in the past to the partial pressure of CO_2 today. Autocorrelation and partial autocorrelation functions for RCO_2 (Fig. 4) show that none of the partial autocorrelations beyond that of lag-1 are significantly different from zero, suggesting that a first-order autoregressive time series may model the data.

The tables and graphs of the printed article all show values of $Z(t_{\hat{x}})$ for $x = b_1$ and ρ_b, a_1 and ρ_a, or ρ, depending on the model being used. For a fit of the linear drift model (Eq. **3**) to the RCO_2 time series, $\hat{b}_1 = -0.0533$, $Z(t_{\hat{b}_1}) = -2.18$, $\hat{\rho}_b = 0.8587$ and $Z(t_{\hat{\rho}_b}) = -2.54$. From Table III of Dickey and Fuller (2) with our sample size of 58, in order for $\hat{b}_1$ to be (90 %) significantly different from zero we should have $\left|Z(t_{\hat{b}_1})\right| > 2.38$. With $\left|Z(t_{\hat{b}_1})\right| = 2.18$ we do not accept the linear drift model.

For a fit of the constant drift model (Eq. **2**) to the RCO_2 time series, $\hat{a}_0 = 0.1731$, $Z(t_{\hat{a}_0}) = 0.4586$, $\hat{\rho}_a = 0.9585$ and $Z(t_{\hat{\rho}_a}) = -1.06$. From Table III of Dickey and Fuller (2) for $\hat{a}_0$ to be (90 %) significantly different from zero $Z(t_{\hat{\rho}_a})$ should be greater than 2.75, and we do not accept the constant drift model.

For a fit of the no drift model (Eq. **1**) to the RCO_2 time series, $\hat{\rho} = 0.9708$ and $Z(t_{\hat{\rho}}) = -1.14$. From Table 8.5.2 of Fuller (10) we see that for $\hat{\rho}$ to be (90 %) significantly different from 1, $Z(t_{\hat{\rho}})$ should be less than -1.95 or greater than 1.31. With

$Z(t_{\hat{\rho}}) = -1.14$ we do not reject null hypothesis that that the Phanerozoic RCO_2 time series is a random walk.

Windows of width 48 of the RCO_2 time series (approximately 450 Myr) were analyzed individually. The linear and constant drift models were not accepted in any window. The values of $Z(t_{\hat{\rho}})$ plotted at the last times of the windows is shown in Fig. 5 where it can be seen that the random walk null hypothesis is not rejected in the early windows but is rejected in the more recent windows.

Models in Deviation Form. Dickey (personal communication) suggested that we use autoregressive models in deviation form so that the no-drift, constant-drift, and linear-drift equations would be written as

$$y_t = \rho y_{t-1} + \epsilon_t \qquad [\mathbf{8}]$$

$$y_t - a_0 = \rho_a (y_{t-1} - a_0) + \epsilon_t \qquad [\mathbf{9}]$$

$$y_t - b_0 - b_1 (t - n/2) = \rho_b (y_{t-1} - b_0 - b_1 (t - 1 - n/2)) + \epsilon_t \qquad [\mathbf{10}]$$

The no-drift form, Eq. **8**, is the same as the standard no-drift form, Eq. **1**. The parameters of these models are fit by linear regression as with the previous forms. Thus, the linear drift model is to be written

$$y_t = (1 - \rho) b_0 + \rho b_1 + (1 - \rho) b_1 (t - n/2) + \rho y_{t-1} + \epsilon_t$$

The parameters are determined by linear regression as above except now the constant term is $(1 - \rho) b_0 + \rho b_1$ and the coefficient of $t - n/2$ is $(1 - \rho) b_1$. A curious form arises in the case of $\rho = 1$ in which we have interest. The coefficient of $t - n/2$ in this autoregression is zero and the model is the constant drift model of Eq. **2** (with constant term b_1 instead of a_1). If that constant term is significantly different from zero it is b_1, the coefficient of $t - n/2$ in the deviations linear drift model, Eq. **10**. In light of this, Dickey suggested that we use the linear-drift model in all of our analysis. The estimates of ρ are the same whether using Eq. **3** or Eq. **10**.

If we use the linear drift model we find that for total diversity, $\hat{\rho} = 1.0131$, $Z(t_{\hat{\rho}}) = 1.18$; for accumulated origination, $\hat{\rho} = 1.0451$, $Z(t_{\hat{\rho}}) = 0.18$; and for accumulated extinction, $\hat{\rho} = 0.9572$, $Z(t_{\hat{\rho}}) = -2.06$. From Table 8.5.2 of (10), we see that in order for ρ to be significantly different from 1, $Z(t_{\hat{\rho}})$ should be less than -3.45 or greater than -0.90. We conlude that total diversity and accumulated origination are not random walks over the Phanerozoic, but that accumulated extinction is a random walk throughout the Phanerozoic. The graphs of $Z(t_{\hat{\rho}_b})$ for windows of width 90 analogous to those of Figure 2 in the main text appear in Fig. 6 *A-C*. It will be seen that the conclusions are the same as in Figure 2 of the main text: diversity and accumulated origination are random walks in the early windows and are not random walks in the recent windows; accumulated extinction is a random walk in all windows.

1. Dickey, D. A. & Fuller, W. A. (1979) *J. Am. Stat. Assoc.* **74**, 427-431.
2. Dickey, D. A. & Fuller, W. A. (1981) *Econometrica* **49**, 1057-1072.
3. Phillips, P. C. B. (1987) *Econometrica* **55**, 277-301.
4. Phillips, P. C. B. & Perron, P. (1988) *Biometrika* **75**, 335-346.
5. Hall, P. & Heyde, C. C. (1980) *Martingale Limit Theory and its Applications* (Academic, New York).
6. Berner, R. A. & Kothavala, Z. (2001) *Am. J. Sci.* **301**, 182-204.
7. Cornette, J. L, Lieberman, B. S. & Goldstein, R. H. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 7832-7835.
8. Sepkoski, J. J.Jr. (1998) *Phil. Trans. R. Soc. London B.* **353**, 315-326.
9. Sepkoski, J. J. Jr. (2002) *Bull. Am. Paleontol.* **363**, 5-560.
10. Fuller, W. A. (1976) *Introduction to Statistical Time Series* (Wiley, New York).