Sensitivity of Detrended Long-Memory Processes

Anurag N Banerjee,\textsuperscript{1}

Department of Economics and Finance, Durham Business School,
University of Durham,
23/26 Old Elvet, Durham, DH1 3HY, UK.

Telephone: +44 (0) 191 33 46356,
Fax: +44 (0) 191 33 46341
E-mail: a.n.banerjee@durham.ac.uk

October 28, 2011

\textsuperscript{1}I would like to thank ESRC (grant number RES-000-22-0646) and the British Academy for the financial support. I would like to thank the anonymous referee for very useful comments.
Abstract

Long memory tests are often complicated by the presence of deterministic trends. Hence an additional step of detrending the data is necessary. The typical way to detrend a suspected long-memory series is to use OLS or BSP residuals. Applying the method of sensitivity analysis we address the question of how robust these residuals are in presence of potential long memory components. Unlike short memory ARMA process long memory I(d) processes causes sensitivity to OLS/BSP residuals. Therefore, we develop a finite sample measure of the sensitivity of a detrended series based on the residuals. Based on our sensitivity measure we propose a "A rule of thumb" for practitioners to choose between the two methods of detrending, has been provided in this paper.

**JEL Codes:** C22

**Keywords:** sensitivity measure, detrending, long memory
1 Introduction

Long memory I(d) processes generalise linear ARIMA models by allowing for non-integer differencing powers and thereby provide a more flexible framework for analysing time series data. This flexibility enables fractional processes to model stronger data dependence than what is allowed in stationary ARMA models without resorting to non-stationary unit root processes. The autocorrelation functions of ARMA processes decay faster (exponentially) than the fractionally integrated processes which decay hyper geometrically and are therefore known as long memory processes. However, estimators of the fractional model exhibit larger bias and are computationally more demanding (Sowell (1992)). It is therefore, beneficial to discriminate fractionally integrated processes from ARMA specifications in a robust modelling step. For this purpose the literature frequently utilizes different statistical tests for both frequency and time domain.  

1 In reality, the performance most of these tests are complicated by the presence of deterministic trends. Hence an additional step of detrending the data is necessary. There are typical ways to detrend the series 1) to use OLS residuals as the detrended data to be tested for long memory. The alternative method 2) to detrend the series is by using finite differencing also known as BSP (Bhargava-Schmidt-Philips) residuals. These methods served the purpose to test alternative ARMA processes and for unit root tests against stationary alternatives (for example Schmidt and Phillips (1992)) as the residuals are consistent estimates of the errors. There have been some studies on how the methods of detrending affect the tests for long memory. Most of them look for consistency properties of long memory tests. The KPSS test is based on data detrended in levels (OLS) and is known to be consistent against unit root alternatives and also against fractionally integrated alternatives. Whereas Su (1998) shows the revised KPSS test (Schmidt 1993, using BSP residuals) are inconsistent “no matter whether we test ‘short memory against stationary long memory’, ‘short memory against nonstationary long memory’ or ‘short memory against unit root...’”. We are therefore, interested in studying the robustness properties of these

\footnote{For example, the Geweke and Porter-Hudak (1983) test, the modified rescaled range test of Lo (1991) and Lagrange multiplier (Robinson (1991, 1994)). Dolado, Gonzalo and Mayoral (2002b) developed a fractional Dickey-Fuller test using an auxiliary non-linear regression along the lines of Dickey-Fuller test but is inefficient. Recently Lobato and Velasco (2007) proposed an efficient Wald test for fractional unit root.}
two different ways to detrend.

In this article, we propose a direct measure of robustness of the detrended series in finite sample, applying the method of sensitivity. The method of sensitivity analysis was developed by Banerjee and Magnus (1999), who investigated the sensitivity of the OLS estimators to the white noise assumption of the error process. The standard linear regression model \( y = X\beta + u \) was considered assuming \( u \sim (0, \sigma^2\Omega(\theta)) \), with unknown autocorrelation parameters \( \theta \). A sensitivity statistics \( B_1 \) measure was proposed as the sensitivity of \( \hat{y}(\theta) = X\hat{\beta}(\theta) \) the predictor, when autocorrelation parameters \( \theta \) of a stationary ARMA disturbance process moves away from white noise. The authors derived the distribution of \( B_1 \). One of their main conclusions is the OLS residuals are not sensitive to misspecification of ARMA\((p,q)\) disturbances in the error process. This implies that OLS residuals can be used to robustly detrend a short memory time series, which is usually done in practice.

The objective of this paper is to analyse and compare the two methods of detrending data using, BSP and OLS by comparing their sensitivity to misspecification of the unknown differencing parameter \( d \). Intuitively we should expect the BSP residuals are less sensitive when \( d \) (unknown) is close to 1 (sensitivity BSP at 1 is fixed at 5%) and OLS residuals are less sensitive when \( d \) is near 0 (sensitivity OLS at 0 is fixed at 5%).

In this paper we construct the sensitivity statistic of the OLS residuals \( BD_1 \), when \( u \) is distributed as an \( I(d) \) process. Our sensitivity measure \( BD_1 \) shows that the OLS residuals are sensitive to long memory process. Hence we conclude that in general detrending a time series suspected of long memory properties using OLS residuals is not a robust procedure. We also develop \( BD_2 \) as the measure the sensitivity for the the BSP residuals, when \( u \) is distributed as an \( I(d) \) process (i.e. after differencing becomes \( I(d-1) \) process). We show that for a range of \( d \), OLS residuals are less sensitive than BSP residuals. Though in general the BSP residuals are more robust to long-memory specifications for the same \( I(d) \) process. We provide a rule of thumb for the practitioner to choose between OLS or BSP to detrend the series.

The paper is organised as follows: section two sets out the framework of our analysis, in section three we describe and analyse the sensitivity measures of the detrending methods, and finally
section four we draw some conclusions.

2 Sensitivity of the residuals.

Let us consider the following model:

\[ y_t = a + bt + u_t \quad (t = 1, \ldots, T), \]  
(2.1)

where \( u_t \) are distributed as distributed as \( I(d) \) process:

\[ u_t = \Delta^{-d} \varepsilon_t \]  
(2.2)

with innovations \( \varepsilon_1, \ldots, \varepsilon_T \sim \text{i.i.d.} \ (0, \sigma^2) \).

There are two ways to detrend the time series.

1) Assume that the disturbances are \( \text{i.i.d.} \) (i.e. \( d = d_1 = 0 \)) and use the residuals of the OLS regression. Generally the GLS residuals of (2.1) are:

\[ \hat{u}_{1,t}(d) = y_{1,t} - \hat{y}_{1,t}(d) = y_{1,t} - \hat{a}_1(d) - \hat{b}_1(d) t \]

\[ \hat{\beta}_1(d) = (\hat{a}_1(d), \hat{b}_1(d)) = (X_1' \Omega(d) X_1)^{-1} X_1' \Omega(d) y_1 \]

where \( X_1 = ((1, t)) \), \( y_1 = (y_t) \), \( \hat{u}_1(d) = (\hat{u}_{1,t}(d)) \) and \( \Omega(d) \) is the covariance matrix of \( u_t \) process, such that \( \Omega(0) = I \).

2) The BSP (Bhargava-Schmit -Philips) residuals are where we assume, the disturbances have an unit root (i.e. \( d = d_2 = 1 \)), difference \( y_t \) and use the OLS residuals by regressing against a constant and an unit vector. The general implication is that the model (2.1) reduces to,

\[ \Delta y_t = a e_{1,t} + b + \Delta u_t, \quad (t = 1, \ldots, T) \]  
(2.3)

where \( e_1 = (1, 0, \ldots, 0)' \) and \( \Delta u_t = \Delta \Delta^{-d} \varepsilon_t = \Delta^{-(d-1)} \varepsilon_t \) an \( I(d - 1) \) process. Therefore the GLS
residuals of (2.3) are:

\[ \hat{u}_{2,t}(d) = y_t - \hat{a}_2(d) - \hat{b}_2(d) (t-1) \]

\[ \hat{\beta}_2(d) = (\hat{a}_2(d), \hat{b}_2(d)) = (X'_2 \Omega (d-1) X_2)^{-1} X'_2 \Omega (d-1) y_2. \]

where \( X_2 = ((e_1, 1)), y_2 = (\Delta y_t), \hat{u}_2(d) = (\hat{a}_{2,t}(d)) \) and \( \Omega (d-1) \) is the covariance matrix of \( \Delta u_t \) process, such that \( \Omega (0) = I \). Note that \( \hat{b}_2(1) = \Delta y_t \) and \( \hat{a}_{2,t}(1) = y_t - y_1 - (t-1) \hat{b}_2(1) \) are the BSP residuals.

Asymptotically the normalised OLS residuals, \( T^{-1/2-d} \hat{u}_{1,t}(0) \), are a fractional Brownian motion (under the assumption of normality \( \varepsilon_t \)'s (Lee and P. Schmidt 1996)). The normalised BSP residuals, \( T^{-1/2-(d-1)} \hat{u}_{2,t}(1) \), are a Brownian Bridge (Su 1998). Here we are interested in deviations of the GLS ( of the respective methods) residuals from their respective reference values 0 and 1. We are interested in small sample results.

Under both methods we can write the variance matrix as,

\[ \Omega (d - d_i) = \sum_{h=0}^{T-1} \omega_h (d - d_i) T^{(h)}, i = 1, 2. \]  \hspace{1cm} (2.4)

where we denote by \( T^{(h)}, 0 \leq h \leq T - 1 \), the \( T \times T \) symmetric Toeplitz matrix with

\[ T^{(h)}_{(i,j)} = \begin{cases} 1 & \text{if } |i-j| = h, \\ 0 & \text{otherwise}. \end{cases} \]

and the coefficients \( \omega_h(d - d_i) \), (with \( d_1 = 0 \) and \( d_2 = 1 \)) are given by the autocovariance generating function:

\[ g(d, z) = [(1 - z) \left(1 - z^{-1}\right)]^{-d} = \sum_{h=-\infty}^{\infty} \omega_h (d) z^h. \]

**Theorem 1** Let \( \Omega (d) \) be the covariance matrix of \( u_1, \ldots, u_T \) and \( \Omega (d-1) \) be the covariance
matrix of \( \Delta u_1, \ldots, \Delta u_T \). Then:

\[
A = \frac{\partial \Omega (d - d_i)}{\partial d} \bigg|_{d=d_i} = \sum_{h=1}^{T-1} \frac{1}{h} T^{(h)}, d_i \in \{0, 1\}. \tag{2.5}
\]

Proof of Theorem 1: See Appendix.

The theorem can be generalised easily to general ARFIMA processes (see Appendix Theorem (3))

Finally we define,

\[
C_i = (I - M_i)AM_i
\]

where \( M_i = I - X_i(X'_iX_i)^{-1}X'_i \), \( i = 1, 2 \) be the usual idempotent matrix for the different methods of detrending.

Sensitivity analysis of a given statistic asks the question about how sensitive the statistic is to changes in nuisance parameter specifications. We ask how far \( \hat{u}_i(d) \) is removed from \( \hat{u}_i(d_i) \), where \( d_1 = 0 \) (i.e. if OLS residuals are used) and \( d_2 = 1 \) (if BSP residuals are used). We define the sensitivity of the detrended series \( \hat{u}_i(d) \) with respect to \( d \) as:

\[
z_i = \frac{\partial \hat{u}_i(d)}{\partial d_i} \bigg|_{d=d_i}, d_i \in \{0, 1\}. \tag{2.6}
\]

In order to use the \( T \times 1 \) vector \( z_i \) as a sensitivity statistic, it is transformed into a quadratic form in the usual way as:

\[
BD_i = z'_i(C_iC'_i)^{-1}z_i \cdot \frac{1}{(T-2)\hat{\sigma}_i^2(0)} \tag{2.7}
\]

where \( \hat{\sigma}_i^2(0) \) is the OLS/BSP estimator of the variance.

The following theorem gives us the distribution of \( BD_i \):

**Theorem 2** We have,

1. \( z_i = -C_iy \); such that \( E(z_i) = 0 \) and \( Var(z_i) = \sigma^2C_iC'_i \).

2. \( BD_i = \frac{y'C_i(C_iC'_i)^{-1}C_iy}{y'M_iy} \);
3. In addition if \( u \sim N(0, \sigma^2 \Omega (d - d_i)) \) then

\[
BD_i \sim \frac{u' \left( C_i' \left( C_i - C_i^\prime \right)^{-1} C_i \right) u}{u'M_iu}, \quad i = 1, 2
\]

a ratio of quadratic forms and when the distribution of \( u \) is evaluated at \( d = d_i \),

\[
BD_i \sim Beta(1, (T - 2)/2), \quad i = 1, 2.
\]

**Proof of Theorem 2:** See Appendix 1.

Note that the results of the theorem (1) and (2) part 1 and 2 does not depend of the distribution. In particular equation (2.5) is true for any \( d_i \), although we are concerned with the sensitivity when \( d_1 = 0 \) and \( d_2 = 1 \). The distribution of \( BD_i \) is Beta only \( u_t \sim I(d_i) \). When \( d \neq d_i \), \( BD_i \) is computed using IMHOF method.

Additional comments are in order. Generalising the results to ARFIMA(1,d,1) series with parameter \( \theta = (\rho, d, \phi) \) and consider approximating the residuals with OLS residuals \( \hat{u}_i(0) \). The short memory deviations can be separated from the long memory components as

\[
\hat{u}_i(\theta) - \hat{u}_i(0) \approx d \left. \frac{\partial \hat{u}_i(\theta)}{\partial d_i} \right|_{\theta=0} + \rho \left. \frac{\partial \hat{u}_i(\theta)}{\partial \rho} \right|_{\theta=0} + \phi \left. \frac{\partial \hat{u}_i(\theta)}{\partial \phi} \right|_{\theta=0} + ...
\]

Banerjee and Magnus (1999) shows that the sensitivity of OLS residuals \( \left. \frac{\partial \hat{u}_i(\theta)}{\partial \phi} \right|_{\theta=0} \) and/or \( \left. \frac{\partial \hat{u}_i(\theta)}{\partial \phi} \right|_{\theta=0} \) is measured by the statistic \( B1 \) which is different from the statistic \( BD_1 \) considered here. Since Banerjee and Magnus (1999) shows that residuals are not sensitive to short-memory deviations (ARMA process) we shall only concentrate on long memory deviations for our analysis.

### 3 Properties of \( BD_1 \) and \( BD_2 \)

Since we are interested in knowing how close are \( \hat{u}_i(d) \) to \( \hat{u}_i(d_i) \) we need to study how close \( BD_i \) is to zero. Since the sensitivity measure \( BD_i \) will generally be random variables we study
the following probabilities as a measure of “closeness” to zero,

\[ \pi_i(d) = \Pr_d \left( BD_i \geq B_\alpha \right), \quad i = 1, 2. \]  

(3.1)

where \( \Pr_d \) is the probability measure associated with the random variable \( u \sim N(0, \sigma^2 \Omega (d - d_i))^2 \).

The cutoff point \( B_\alpha \) is obtained from the equation,

\[ \pi_i(d) = \Pr_d \left( BD_i \geq B_\alpha \right) = \alpha, \quad 0 < \alpha < 1. \]  

(3.2)

where \( \Pr_d i \) is the probability measure of the \textit{Beta}—distribution (Theorem 2) associated with white noise \((d_1 = 0)\) or the integrated process \((d_2 = 1)\).

This probability function \( \pi_i(d) \) is essentially a robustness function of the residuals \( \hat{u}_i(d) \), against the violation of white noise assumption of the error process against long-memory or short-memory alternatives. These probabilities give an indication of how close to zero the sensitivity measures are. The greater the probability mass of the sensitivity measures around zero, closer is the distance between \( \hat{u}_i(d) \) and \( \hat{u}_i(d_i) \). In order to have a sharper bound for the sensitivity we can choose a lower value of \( \alpha \) (in this paper, we chose \( \alpha = 0.05 \)). Higher the value of \( \pi_i(d) \), the higher is the probability of sensitivity at \( d \). Theorem 2 gives the distribution of \( BD_i \) statistic, which helps us to compute the \( B_\alpha \) at \( d = d_i \). Noting that the \( BD_i \) are ratios of quadratic forms \( \pi_i(d) \) is calculated using IMHOF method, assuming the innovations \( \varepsilon_i \)'s, are normally distributed (see Mathai and Provost (1992)). Asymptotically \( \lim_{T \to \infty} BD_i = 0 \), hence \( \lim_{T \to \infty} \pi_i(d) = 1 \) or \( 0 \), for any \( B_\alpha > 0 \). This implies for large sample the sensitivity measure will always show sensitivity. Therefore, we shall confine ourselves to small sample results.

We consider the model (2.1) with \( T = 25 \) and \( \varepsilon_i \sim N(0, \sigma^2) \). Figure 1 shows the \( \pi_i(d), \ i = 1, 2 \) curves for the OLS and the BSP residuals. One obvious observation is that both OLS and BSP residuals are sensitive to the departure of \( d \) from their respective default values 0 and 1 respectively. But the figures also show the relative flatness of the \( \pi_2(d) \) to \( \pi_1(d) \). This

\footnote{Normality is not crucial for the definition.}
indicates the insensitivity of the BSP residuals with respect to an \( I(d) \) process compared to OLS residuals. In contrast \( \pi_1 (d) \) is steep and goes above \( \pi_2 (d) \) at around \( d = 0.3 \). So if the long memory parameter \( d < 0.3 \) OLS residuals are less sensitive than BSP residuals. This makes intuitive sense since the assumption behind using OLS detrending is to assume that the process is white noise.

Now let us consider an ARFIMA(1,d,1) series where we consider approximating the residuals with OLS residuals \( \hat{u}_t(0) \). Figure 2 shows how our sensitivity measures perform against general ARFIMA processes; we further compute the \( \pi_i (d) \) curves when the disturbances follows the ARFIMA(1,d,1): \( \Delta^d u_t = 0.6\Delta^d u_{t-1} + 0.2\varepsilon_{t-1} + \varepsilon_t \) process. We note that the OLS residuals show sensitivity even at \( d = 0 \) (\( \pi_1 (0) = 0.52 \)), which is in contrast with the results of insensitivity of OLS to shortmemory deviations in Banerjee and Magnus (1999). There are two reasons for this apparent deviation. Firstly, we should be using \( B_1 \) to use measure short memory deviation. Indeed if \( B_1 \) is used the sensitivity at \( d = 0 \), i.e. when \( u_t = 0.6u_{t-1} + 0.2\varepsilon_{t-1} + \varepsilon_t \), the OLS sensitivity is \( \pi(B_1) = 0.12 \). Secondly, we show the \( BD_1 \) statistic as showing sensitivity because \( I(d) \) processes can be approximated by ARMA(1,1) process as shown in Basak et. al. (2001). This shows that the sensitivity of BSP residuals is lower than the OLS residuals for most values of \( d \).

We then analyse the robustness of our findings to the normality assumptions as in theorem (2) to see if our \( \pi \) functions are robust. We assume \( \Delta^d u_t = \varepsilon_t, \varepsilon_t \sim t_2 \) (Student T-Distribution with 2 df). Figure 3, shows that the \( \pi_i (d) \) curves are qualitatively very similar to the curves in Figure 1. The \( \pi_i (d) \) curves in Figure 3, are computed using 2000 independent draws.

General conclusion from the analysis is that detrending the series using OLS estimates is a non robust process most of the time and that, we should use BSP residuals. But sometimes OLS performs better than BSP residuals, therefore we can also propose the following rule of thumb to
determine which detrending method we should use. We would choose OLS if \( \Pr (BD_1 \geq B_o) < \Pr (BD_2 \geq B_o) \), otherwise we choose BSP. This is similar to saying \( BD_1 \leq BD_2 \). Therefore our rule of thumb is:

<table>
<thead>
<tr>
<th>If ( BD_1 \leq BD_2 )</th>
<th>Use OLS residuals to detrend</th>
</tr>
</thead>
<tbody>
<tr>
<td>or otherwise</td>
<td>Use BSP residuals to detrend.</td>
</tr>
</tbody>
</table>

The calculations of these statistic(s) are simple and can be done using an appropriate software package like MATLAB or Gauss.

4 Conclusions

In this article, two sensitivity statistics \( BD_1 \) and \( BD_2 \) are designed to decide whether to detrend a potential long memory series using OLS residuals or BSP residuals. Our results show that the OLS residuals are sensitive to the fractionally integrated process misspecification, which is in contrast to the results of Banerjee and Magnus (1999), where they conclude that the OLS predictor/residuals is robust to short memory ARMA specification, in general both over-differencing or under differencing matters. One important outcome of this observation is that we give a rule of thumb when to use a OLS residuals to detrend a long-memory series.

References


**Appendix 1:**

**Proof of Theorem 1:**

1) Let the power series expansion be

\[ g(d, z) = \left[ (1 - z) (1 - z^{-1}) \right]^{-d} = \sum_{h=-\infty}^{\infty} \omega_h(d) z^h. \]
Note that $\omega_h (d)$ can be obtained from by using the Integral transform:

$$\omega_h (d) = \frac{1}{2\pi} \int_0^{2\pi} \left[ (1 - z) (1 - z^{-1}) \right]^{-d} z^{-h} dz$$

Differentiation under integral at $d = 0$ gives us:

$$\frac{\partial \omega_h (d)}{\partial d} \bigg|_{d=0} = \frac{1}{2\pi} \int_0^{2\pi} \ln (1 - z) (1 - z^{-1}) z^{-h} dz$$

Let $z = \exp (-ix)$, where $i = \sqrt{-1}$ then $(1 - z) (1 - z^{-1}) = 2 \sin^2 \left( \frac{x}{2} \right)$, therefore:

$$\frac{\partial \omega_h (d)}{\partial d} \bigg|_{d=0} = \frac{1}{2\pi} \int_0^{2\pi} \ln 2 \sin^2 \left( \frac{x}{2} \right) \exp (-ixh) dx.$$ 

Since the Fourier coefficients of the sine part of the transform is zero, then

$$\frac{\partial \omega_h (d)}{\partial d} \bigg|_{d=0} = \frac{1}{2\pi} \int_0^{2\pi} \ln 2 \sin^2 \left( \frac{x}{2} \right) \cos (hx) dx = \frac{1}{h}.$$ 

QED.

**Proof of Theorem 2:** a) Using standard results of differential calculus we obtain

$$\frac{\partial \mathbf{\hat{u}}_i (d)}{\partial d} = \mathbf{X}_i \left( \mathbf{X}_i \mathbf{\Omega} (d - d_i)^{-1} \mathbf{X}_i \right)^{-1} \frac{\partial \mathbf{\Omega} (d - d_i)^{-1}}{\partial d} \left( \mathbf{y}_i - \mathbf{X}_i \mathbf{\hat{\beta}}_i (d) \right).$$

Since $\mathbf{\Omega} (0) = \mathbf{I}$, $i = 1, 2$ we have

$$z_i = \frac{\partial \mathbf{\hat{u}}_i (d)}{\partial d} \bigg|_{d=d_i} = -\mathbf{X}_i \left( \mathbf{X}_i \mathbf{X}_i \right)^{-1} \mathbf{X}_i \mathbf{A} \mathbf{M} \mathbf{y}_i = -\mathbf{C}_i \mathbf{y}_i.$$ 

b) and c) follows from Banerjee and Magnus (1999) Theorem 2.
4.1 Additional Results:

If the $y_t$ process is covariance stationary, then $\Omega(\theta)$ can be written as

$$\Omega(\theta) = I + \sum_{h=1}^{T-1} \omega_h(\theta) T^{(h)}, \quad (4.1)$$

where $T^{(h)}$, $0 \leq h \leq T - 1$, the $T \times T$ symmetric Toeplitz matrix with

$$T^{(h)}(i,j) = \begin{cases} 1 & \text{if } |i-j| = h, \\ 0 & \text{otherwise.} \end{cases}$$

If the errors $u_t$ are distributed as ARFIMA(p,d,q) process:

$$\Delta^d u_t = \sum_{i=1}^{p} \phi_i \Delta^d u_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t \quad (4.2)$$

with innovations $\varepsilon_1, \ldots, \varepsilon_T \sim \text{i.i.d. } (0, \sigma^2)$ and $\theta = (\phi_1, \ldots, \phi_p, d, \psi_1, \ldots, \psi_q)$.

**Theorem 3** Let $\sigma^2 \Omega(\theta)$ be the covariance matrix of $u_1, \ldots, u_n$. Then,

$$\frac{\partial \Omega(\theta)}{\partial \phi_i} \bigg|_{\theta=0} = \frac{\partial \Omega(\theta)}{\partial \psi_i} \bigg|_{\theta=0} = T^{(i)}, \quad (4.3)$$

and

$$\frac{\partial \Omega(\theta)}{\partial d} \bigg|_{\theta=0} = \sum_{t=1}^{T-1} \frac{1}{t} T^{(t)}. \quad (4.3)$$

**Proof of Theorem 3:** Following Harvey (1993, p. 29) we introduce the autocovariance generating function

$$g_\theta(z) = \sum_{h=-\infty}^{\infty} \omega_h(\theta) z^h,$$

where $\omega_h(\theta)$ is the autocovariance at lag $h$ and $z$ is a complex number. Note $g_\theta(L) = \Omega(\theta)$ where $L$ is the lag-operator. For the ARFIMA(p,d,q) model we have

$$g_\theta(z) = \frac{\psi(z)(1-z)^{-d}(1-z^{-1})^{-d}}{\phi(z)\phi(z^{-1})}.$$
where $\psi(z) = 1 + \sum_{h=1}^{q} \psi_h z^h$ and $\phi(z) = 1 - \sum_{h=1}^{q} \phi_h z^h$ and $\theta = (\phi_1, ..., \phi_p, d, \psi_1, ..., \psi_q)$. Note that $g_0(z) = 1$. The first derivatives $g_{\theta}(z)$ with respect to $\psi_h$ give

$$\frac{\partial g_{\theta}(z)}{\partial \psi_h} = \frac{(1 - z)^{-d} (1 - z^{-1})^{-d}}{\phi(z) \phi(z^{-1})} (\psi(z^{-1}) z^h + \psi(z) z^{-h})$$

The first derivatives $g_{\theta}(z)$ with respect to $\phi_h$ give

$$\frac{\partial g_{\theta}(z)}{\partial \phi_h} = g_{\theta}(z) \frac{(\phi(z^{-1}) z^h + \phi(z) z^{-h})}{\phi(z) \phi(z^{-1})}$$

The first derivatives $g_{\theta}(z)$ with respect to $d$ gives

$$\frac{\partial g_{\theta}(z)}{\partial d} = -g_{\theta}(z) \left[ \log (1 - z^{-1}) + \log (1 - z^{-1}) \right]$$

and hence, at $\theta = 0$,

$$\left. \frac{\partial g_{\theta}(z)}{\partial \phi_h} \right|_{\theta=0} = \left. \frac{\partial g_{\theta}(z)}{\partial \psi_h} \right|_{\theta=0} = (z^h + z^{-h}),$$

and

$$\left. \frac{\partial g_{\theta}(z)}{\partial d} \right|_{\theta=0} = \log (1 - z) (1 - z^{-1}) = \sum_{h=-\infty}^{\infty} \frac{1}{|h|} z^h$$

The last two expressions are obtained from lemma (1).

QED.

**Figures**
Figure 1: Sensitivity measure $\pi(d)$ of $BD_1$ and $BD_2$ : under $\Delta^d u_t = \varepsilon_t$. 
Figure 2: Sensitivity measure $\pi(d)$ of $BD_1$ and $BD_2$: under $\Delta^d u_t = 0.6 \Delta^d u_{t-1} + 0.2 \varepsilon_{t-1} + \varepsilon_t$
Figure 3: Sensitivity measure $\pi(d)$ of $BD_1$ and $BD_2$ : under $\Delta^d u_t = \varepsilon_t$, $\varepsilon_t \sim t_2$