Simple (but effective) tests of long memory versus structural breaks

Katsumi Shimotsu
Department of Economics
Queen's University

Outline of the talk

- $I(d)$ (long memory) processes and semiparametric estimation
- Spurious long memory models
- True versus spurious $I(d)$: sample splitting-based diagnosis
- True versus spurious $I(d)$: test using dth differencing
- Empirical example with S&P 500 daily realized volatility

$I(d)$ (long memory) process

$$(1-L)^d X_t = u_t, \quad X_t = (1-L)^{-d} u_t = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\, k!} u_{t-k}$$

u_t : weakly dependent $(0, f_u(\lambda))$; $\quad d$: real number

- Slowly (hyperbolically) decaying impulse response when $d > 0$

$I(d)$ (long memory) process

$$(1-L)^d X_t = u_t, \quad X_t = (1-L)^{-d} u_t = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\, k!} u_{t-k}$$

u_t : weakly dependent $(0, f_u(\lambda))$; d : real number

- Slowly (hyperbolically) decaying impulse response when $d > 0$

- Suitable for modeling persistent autocorrelations commonly observed in financial volatility data: Ding, Granger and Engle (93), Baillie et al. (96), Andersen and Bollerslev (97), Breidt et al. (98), Bollerslev and Wright (00), Andersen et al. (03), Deo et al. (05), Hurvich et al. (05) (just to name a few)

Semiparametric estimators of d

- Focuses on the estimation of d by using the periodogram ordinates only in the vicinity of the origin. Agnostic to the short-run dynamics of the time series, easy to implement.

Semiparametric estimators of d

- Focuses on the estimation of d by using the periodogram ordinates only in the vicinity of the origin. Agnostic to the short-run dynamics of the time series, easy to implement.

- We use the local Whittle estimator (Künsch 1987, Robinson 1995b) that uses the frequency domain Gaussian likelihood function localized to the origin under the assumption $f_x(\lambda) \sim G\lambda^{-2d}$

$$(\widehat{G}, \widehat{d}) = \arg\min \frac{1}{m} \sum_{j=1}^{m} \left[\log \left(G\lambda_j^{-2d} \right) + \frac{\lambda_j^{2d}}{G} I_x(\lambda_j) \right],$$

where $I_x(\lambda_j)$ is the periodogram of X_t and m (bandwidth) satisfies $m/n \to 0$.

Spurious long memory models

- A process with structural breaks can be easily confused with a long memory process (Diebold and Inoue (01), Gourieroux and Jasiak (01), Granger and Hyung (04), Perron and Qu (06))

- Spurious long memory: their periodogram tends to have a pole at the origin and the semiparametric estimates of d tend to be positive.

Spurious long memory models

- A process with structural breaks can be easily confused with a long memory process (Diebold and Inoue (01), Gourieroux and Jasiak (01), Granger and Hyung (04), Perron and Qu (06))

- Spurious long memory: their periodogram tends to have a pole at the origin and the semiparametric estimates of d tend to be positive.

- Mean-plus-noise model with small p

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + v_t, \quad v_t = \begin{cases} 0 & \text{with probability } 1-p, \\ w_t & \text{with probability } p, \end{cases}$$

where $\varepsilon_t \sim iidN(0, \sigma_\varepsilon^2)$ and $w_t \sim iidN(0, 1)$.

- Stochastic permanent break (STOPBREAK) model with large γ

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma + \varepsilon_{t-1}^2}\varepsilon_{t-1}, \quad \varepsilon_t \sim iidN(0,1).$$

  – a smoother version of the mean-plus-noise model.

- Stochastic permanent break (STOPBREAK) model with large γ

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma + \varepsilon_{t-1}^2}\varepsilon_{t-1}, \quad \varepsilon_t \sim iidN(0,1).$$

– a smoother version of the mean-plus-noise model.

- Markov-switching model with large p_{00}, p_{11}

$$X_t \sim N(\mu_t, \sigma^2),$$

$\mu_t \in \{0,1\}$: first-order Markov process with transition probability

$$M = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}.$$

- Stochastic permanent break (STOPBREAK) model with large γ

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu_{t-1} + \frac{\varepsilon_{t-1}^2}{\gamma + \varepsilon_{t-1}^2}\varepsilon_{t-1}, \quad \varepsilon_t \sim iidN(0,1).$$

  – a smoother version of the mean-plus-noise model.

- Markov-switching model with large p_{00}, p_{11}

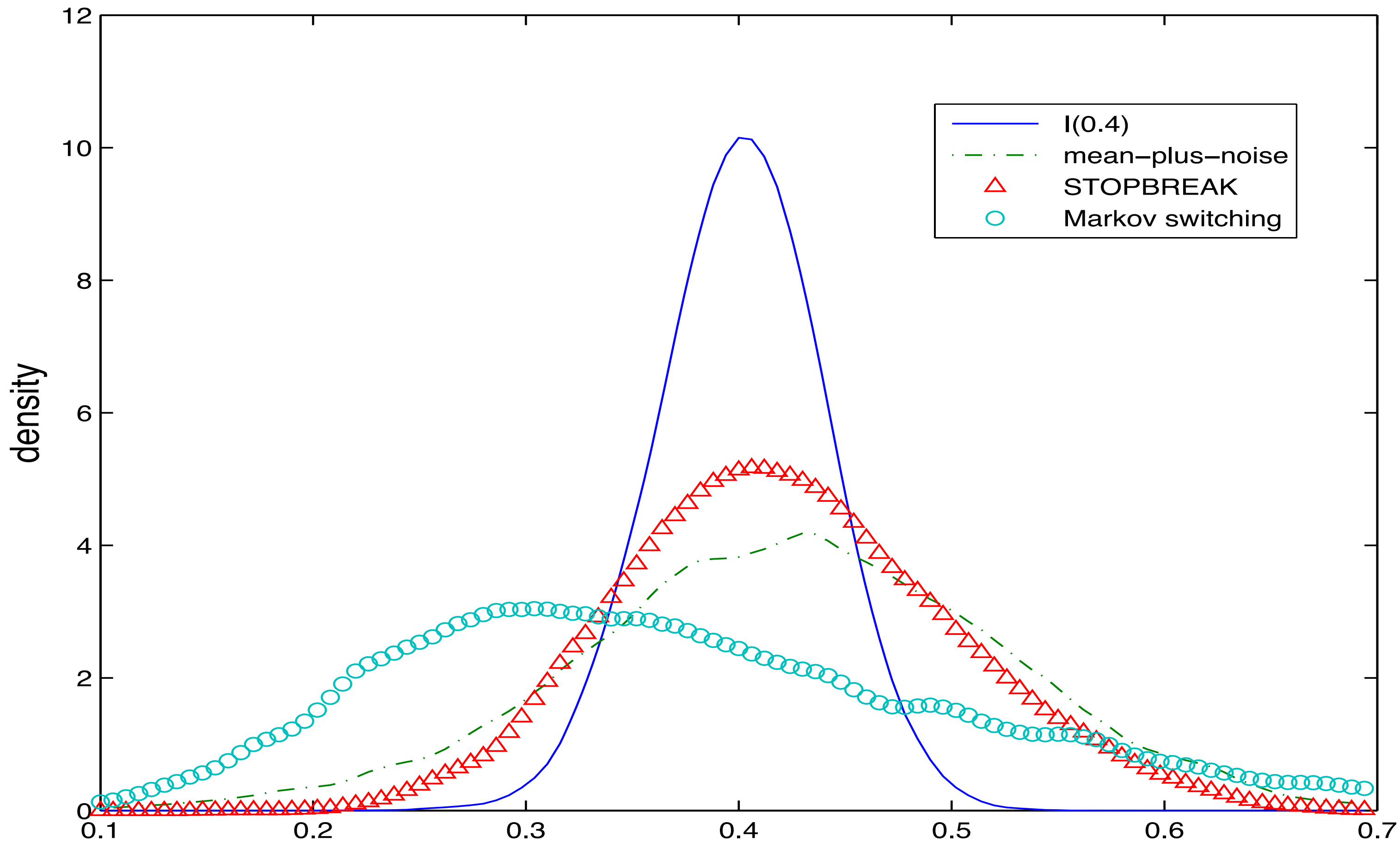
$$X_t \sim N(\mu_t, \sigma^2),$$

  $\mu_t \in \{0,1\}$: first-order Markov process with transition probability

$$M = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}.$$

- Mean-plus-noise model and STOPBREAK model are essentially $I(1)$, while the Markov switching model is essentially $I(0)$.

Densities of $\hat{d}$ with 5,000 observations and m=200



It is all spurious?

- Evidence that long memory models provide significantly better out-of-sample predictions than AR, MA, ARMA, GARCH and related models (Andersen et al. (03), Bhardwaj and Swanson (06))

- Positive estimate of d even for different degrees of temporal aggregation (Andersen et al. (00), Ohanissian et al. (04))

- Evidence of long memory even after removing structural breaks (Choi and Zivot (05))

It is all spurious?

- Evidence that long memory models provide significantly better out-of-sample predictions than AR, MA, ARMA, GARCH and related models (Andersen et al. (03), Bhardwaj and Swanson (06))

- Positive estimate of d even for different degrees of temporal aggregation (Andersen et al. (00), Ohanissian et al. (04))

- Evidence of long memory even after removing structural breaks (Choi and Zivot (05))

- We propose two simple tests of true $I(d)$ against spurious $I(d)$.

- Effective with 5,000 observations ($\simeq$ 20 years of daily data) and 240 observations.

- Other tests of $I(d)$ versus structural breaks: Berkes et al. (06). Giraitis et al. (06), Ohanissian et al. (04), Dolado et al. (05), Mayoral (06), Perron and Qu (06).

Test A: sample splitting

- Let b be an even integer and split the sample into b blocks, so that each block has n/b observations. $\{X_t : t = (a-1)n/b + 1, \ldots, an/b\}$.

- Define $\hat{d}^{(a)}$, $a = 1, \ldots, b$, be the LW estimator of d computed from the ath subsample and with the bandwidth m/b.

- Define $\bar{d} = b^{-1} \sum_{a=1}^{b} \hat{d}^{(a)}$: the average of the estimates from the subsamples.

Test A: sample splitting

- Let b be an even integer and split the sample into b blocks, so that each block has n/b observations. $\{X_t : t = (a-1)n/b + 1, \ldots, an/b\}$.

- Define $\hat{d}^{(a)}$, $a = 1, \ldots, b$, be the LW estimator of d computed from the ath subsample and with the bandwidth m/b.

- Define $\bar{d} = b^{-1} \sum_{a=1}^{b} \hat{d}^{(a)}$: the average of the estimates from the subsamples.

- If X_t is $I(d)$, then $\bar{d}$ should be close to $\hat{d}$ and $\hat{d}^{(a)}$ should be close to each other.

- $m/b \Rightarrow$ the same amount of frequency domain information as the estimation by the total sample

- $m/b \Rightarrow$ same bias from short-run dynamics as estimation from the total sample

Sample splitting: (standard) assumptions

- Assumption A1: for some $\beta \in (0, 2]$, $f(\lambda) \sim G_0 \lambda^{-2d_0}(1 + O(\lambda^{\beta}))$ as $\lambda \to 0+$, $G_0 \in (0, \infty)$.

- Assumption A2: $X_t - EX_0 = C(L)\,\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} c_j^2 < \infty$, where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$ a.s., $t = 0, \pm 1, \ldots$ Plus some technical assumptions on ε_t, and the smoothness of $C(e^{i\lambda})$.

- Assumption A3: as $n \to \infty$, $1/m + m^{1+2\beta} n^{-2\beta} \log^2 m \to 0$.

Sample splitting: (standard) assumptions

- Assumption A1: for some $\beta \in (0,2]$, $f(\lambda) \sim G_0 \lambda^{-2d_0}(1 + O(\lambda^{\beta}))$ as $\lambda \to 0+$, $G_0 \in (0,\infty)$.

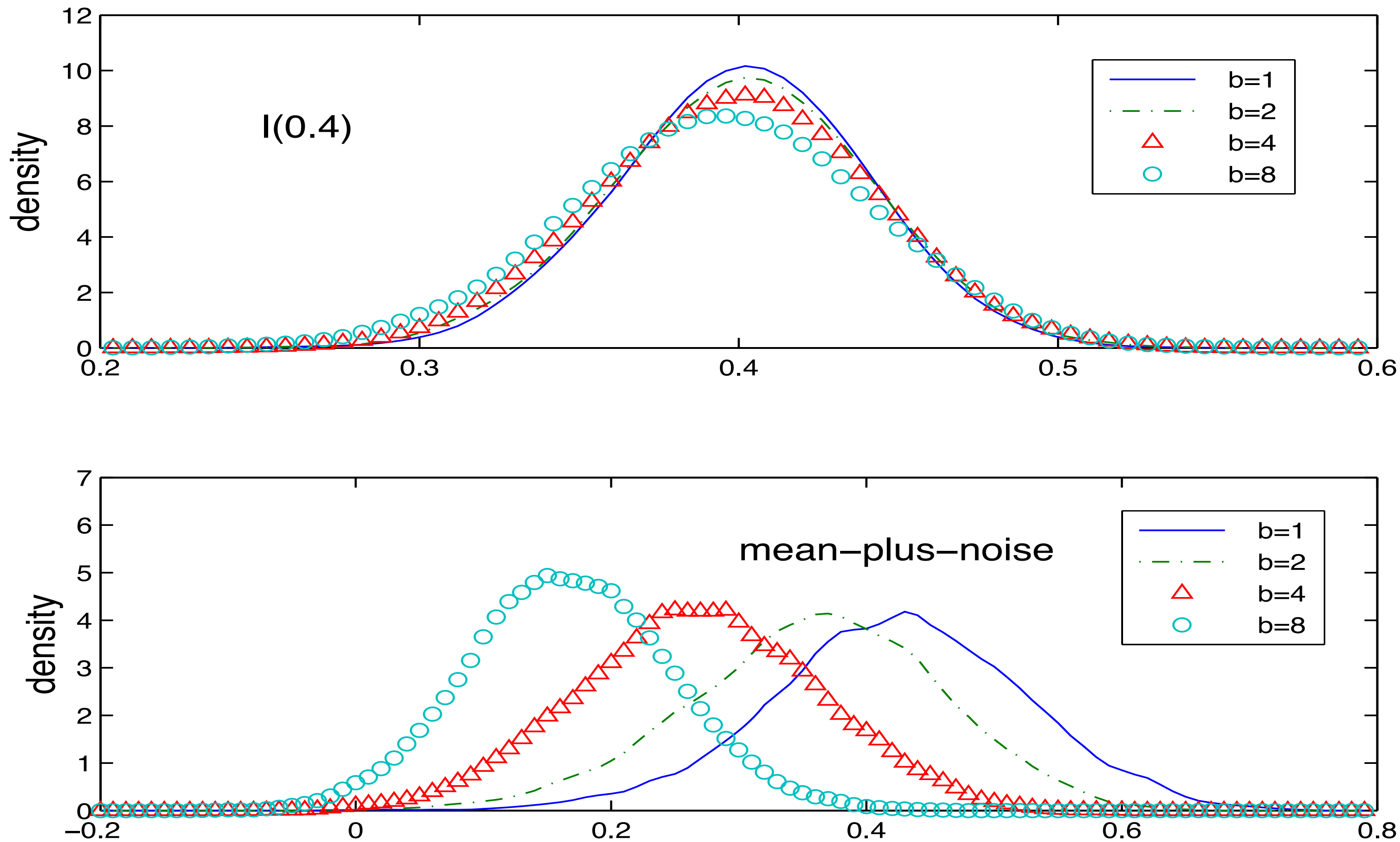
- Assumption A2: $X_t - EX_0 = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$, $\quad \sum_{j=0}^{\infty} c_j^2 < \infty$, where $E(\varepsilon_t | F_{t-1}) = 0$, $E(\varepsilon_t^2 | F_{t-1}) = 1$ a.s., $t = 0, \pm 1, \ldots$ Plus some technical assumptions on ε_t, and the smoothness of $C(e^{i\lambda})$.

- Assumption A3: as $n \to \infty$, $1/m + m^{1+2\beta} n^{-2\beta} \log^2 m \to 0$.

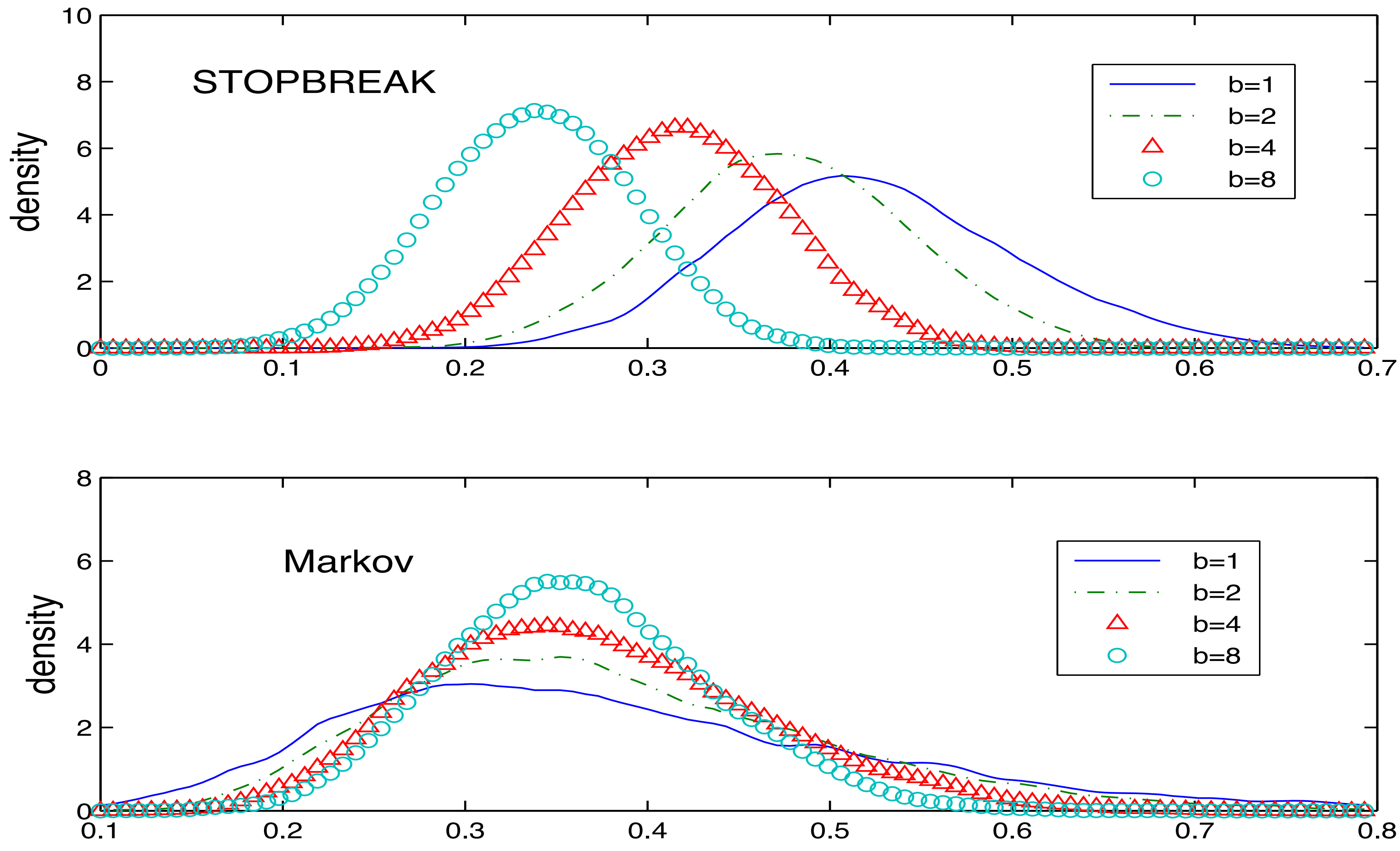
Lemma 1. *Suppose Assumptions A1-A3 hold. Then* $\sqrt{m}(\bar{d} - d_0) \to_d N(0, 1/4)$ *as* $n \to \infty$.

- The limiting variance of $\bar{d}$ is the same as that of $\widehat{d}$, but $\bar{d}$ has a larger variance than $\widehat{d}$ in a finite sample.

Densities of $\bar{d}$ with 5,000 observations and m=240



Densities of $\bar{d}$ with 5,000 observations and m=240



Test statistic for the parameter constancy

- Tests the hypothesis $H_0 : d_0 = d_{0,1} = \cdots = d_{0,b}$.

- Define

$$\hat{d}_b = \begin{pmatrix} \hat{d} - d_0 \\ \hat{d}^{(1)} - d_0 \\ \vdots \\ \hat{d}^{(b)} - d_0 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & \iota_b' \\ \iota_b & bI_b \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & -1 \end{pmatrix},$$

where ι_b is a $b \times 1$ vector of ones.

- Define the Wald statistic for testing H_0 as

$$W_c = 4m \cdot \alpha(m) A\hat{d}_b (A\Omega A')^+ (A\hat{d}_b)'.$$

where $(A\Omega A')^+$ denotes a generalized inverse of $A\Omega A'$ ($A\Omega A'$ has rank $b-1$) and $\alpha(m)$ is the finite sample adjustment factor (Hurvich and Chen (00)) with $\alpha(m) \to 1$.

- Define the Wald statistic for testing H_0 as

$$W_c = 4m \cdot \alpha(m) A\hat{d}_b (A\Omega A')^{+} (A\hat{d}_b)'.$$

where $(A\Omega A')^{+}$ denotes a generalized inverse of $A\Omega A'$ ($A\Omega A'$ has rank $b-1$) and $\alpha(m)$ is the finite sample adjustment factor (Hurvich and Chen (00)) with $\alpha(m) \to 1$.

Lemma 2. *Under the assumptions that are slightly stronger than A1 and slightly weaker than A3, $W_c \to_d \chi^2(b-1)$ as $n \to \infty$.*

We can use a slightly larger m than in estimation of d because the bias of the all elements of $\hat{d}_b$ are the same.

Simulation results

$X_t = (1-L)^{-0.4} u_t$, $u_t \sim iidN(0,1)$, $n = 5,000$

m	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c) 5% asy. critical value		
		$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	0.401	0.401	0.399	0.394	0.060	0.070	0.097
400	0.401	0.401	0.401	0.399	0.056	0.066	0.081
600	0.399	0.400	0.400	0.399	0.053	0.057	0.075
800	0.397	0.397	0.397	0.397	0.055	0.057	0.071

Simulation results

$X_t = (1-L)^{-0.4} u_t$, $u_t \sim iidN(0,1)$, $n = 5,000$

m	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c) 5% asy. critical value		
		$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	0.401	0.401	0.399	0.394	0.060	0.070	0.097
400	0.401	0.401	0.401	0.399	0.056	0.066	0.081
600	0.399	0.400	0.400	0.399	0.053	0.057	0.075
800	0.397	0.397	0.397	0.397	0.055	0.057	0.071

$X_t \sim$ mean-plus-noise model, $\sigma_\varepsilon^2 = 2$, $n = 5,000$

m	p	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c)		
			$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	0.002	0.424	0.359	0.268	0.167	0.456	0.748	0.796
400	0.004	0.397	0.350	0.285	0.201	0.445	0.725	0.881
600	0.006	0.379	0.338	0.283	0.211	0.482	0.725	0.895
800	0.010	0.388	0.352	0.304	0.242	0.493	0.728	0.878

Simulation results

$X_t \sim$ STOPBREAK model, $n = 5,000$

m	γ	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c)		
			$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	180	0.423	0.377	0.316	0.239	0.265	0.286	0.242
400	120	0.398	0.358	0.307	0.243	0.351	0.444	0.408
600	90	0.392	0.356	0.309	0.251	0.414	0.544	0.556
800	70	0.395	0.362	0.319	0.266	0.459	0.620	0.656

Simulation results

$X_t \sim$ STOPBREAK model, $n = 5,000$

m	γ	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c)		
			$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	180	0.423	0.377	0.316	0.239	0.265	0.286	0.242
400	120	0.398	0.358	0.307	0.243	0.351	0.444	0.408
600	90	0.392	0.356	0.309	0.251	0.414	0.544	0.556
800	70	0.395	0.362	0.319	0.266	0.459	0.620	0.656

$X_t \sim$ Markov-switching model, $n = 5,000$

m	p_{00}, p_{11}	mean($\hat{d}$)	mean($\bar{d}$)			rejection frequencies (W_c)		
			$b=2$	$b=4$	$b=8$	$b=2$	$b=4$	$b=8$
200	0.93	0.389	0.381	0.370	0.362	0.479	0.603	0.620
400	0.86	0.414	0.406	0.395	0.384	0.596	0.774	0.839
600	0.75	0.390	0.384	0.374	0.362	0.622	0.813	0.884
800	0.66	0.414	0.408	0.401	0.391	0.670	0.861	0.933

Sample splitting

- Even visual examination provides a good idea on true/spurious long memory
- The test has good power
- Quick “reality check”
- Spurious $I(d)$ is not robust to sample splitting
- Spurious $I(d)$ is not robust to the number of the periodogram ordinates (m)

Extension to nonstationary $I(d)$ processes

$$X_t - \mu_0 = (1 - L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}$$

- One can use either the local Whittle estimator (for $d_0 < 3/4$) of the exact local Whittle estimator (Shimotsu and Phillips (05), Shimotsu (06)).

- The same asymptotic results hold.

Test using dth differencing

- The dgp is $X_t - \mu_0 = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}$.

- Fact: if an $I(d)$ process is differenced d times, then the resulting time series is an $I(0)$ process.

- Test the null of $X_t \sim I(d)$ by applying a test of stationarity and a unit root test to the dth differenced data and its partial sum.

Test using dth differencing

- Obtain a consistent estimate of d.
- Demean the data. Demeaning (estimation of μ_0) needs to be done carefully, because the sample average is not a good estimate of μ_0 when d is large:

$$\hat{\mu}\,(d) = w(d)\overline{X} + (1 - w(d))X_1,$$

  where $w(d)$ is a smooth (twice continuously differentiable) weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq \frac{3}{4}$.

Test using dth differencing

- Obtain a consistent estimate of d.
- Demean the data. Demeaning (estimation of μ_0) needs to be done carefully, because the sample average is not a good estimate of μ_0 when d is large:

$$\hat{\mu}\,(d) = w(d)\overline{X} + (1 - w(d))X_1,$$

  where $w(d)$ is a smooth (twice continuously differentiable) weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq \frac{3}{4}$.

- Take the $\hat{d}$th difference of the demeaned data:

$$\hat{u}_t = (1 - L)^{\hat{d}}(X_t - \hat{\mu}(\hat{d})).$$

Test using dth differencing

- Apply the KPSS test to the differenced data:

$$\hat{\eta}_\mu = n^{-2} \sum_{t=1}^{n} S_t^2 / s^2(l), \quad S_t = \sum_{k=1}^{t} e_k,$$

e_t: the residual from regressing $\hat{u}_t$ on a constant
$s^2(l)$: an estimate of $lrvar(e_t)$ with lag length l

- Apply the Phillips-Perron Z_t test (with an intercept) to the partial sum of the differenced data.

Test using dth differencing

- Apply the KPSS test to the differenced data:

$$\hat{\eta}_\mu = n^{-2} \sum_{t=1}^{n} S_t^2 / s^2(l), \quad S_t = \sum_{k=1}^{t} e_k,$$

  e_t: the residual from regressing $\hat{u}_t$ on a constant
  $s^2(l)$: an estimate of $lrvar(e_t)$ with lag length l

- Apply the Phillips-Perron Z_t test (with an intercept) to the partial sum of the differenced data.

- Spurious $I(d)$ processes are either overdifferenced or underdifferenced after differenced $\hat{d}$ times.

Test using dth differencing: assumptions

- Assumption C1: $X_t - \mu_0 = (1-L)^{-d_0} u_t \mathbf{1}\{t \geq 1\}$, $d_0 \in (-1/2, 3/2)$.

- Assumption C2: $\hat{d} - d_0 = o_p((\log n)^{-1})$.

- Assumption C3: The partial sum process $n^{-d-1/2} \sum_{t=1}^{[nr]} X_t$ converges to a fBM.

- Assumption C4: (a) $l(\hat{d} - d_0) \to_p 0$. (b) $\tilde{s}^2(l) \to_p \omega^2$, where $\tilde{s}^2(l)$ is the version of $s^2(l)$ constructed with u_t instead of $\hat{u}_t$.

- Assumptions C2, C3: technical conditions are provided in Robinson (1995), Phillips and Shimotsu (2005), Marrinuchi and Robinson (2000), and Hosoya (2005)

Test using dth differencing

Lemma 3. *Suppose Assumptions C0-C4 hold. Then the limiting distribution of Z_t and $\hat{\eta}_\mu$ are given by replacing $W(r)$ (standard BM) in their (standard) limiting distribution by $W(r;d)$, where*

$$W(r;d) = W(r) - w(d)(\Gamma(2-d)\Gamma(d+1))^{-1} r^{1-d} \int_0^1 (1-s)^d dW(s).$$

($W(r;d)$ is a Brownian bridge if $d = 0$)

Test using dth differencing

Lemma 3. *Suppose Assumptions C0-C4 hold. Then the limiting distribution of* Z_t *and* $\hat{\eta}_\mu$ *are given by replacing* $W(r)$ *(standard BM) in their (standard) limiting distribution by* $W(r;d)$*, where*

$$W(r;d) = W(r) - w(d)(\Gamma(2-d)\Gamma(d+1))^{-1} r^{1-d} \int_0^1 (1-s)^d dW(s).$$

($W(r;d)$ *is a Brownian bridge if* $d=0$*)*

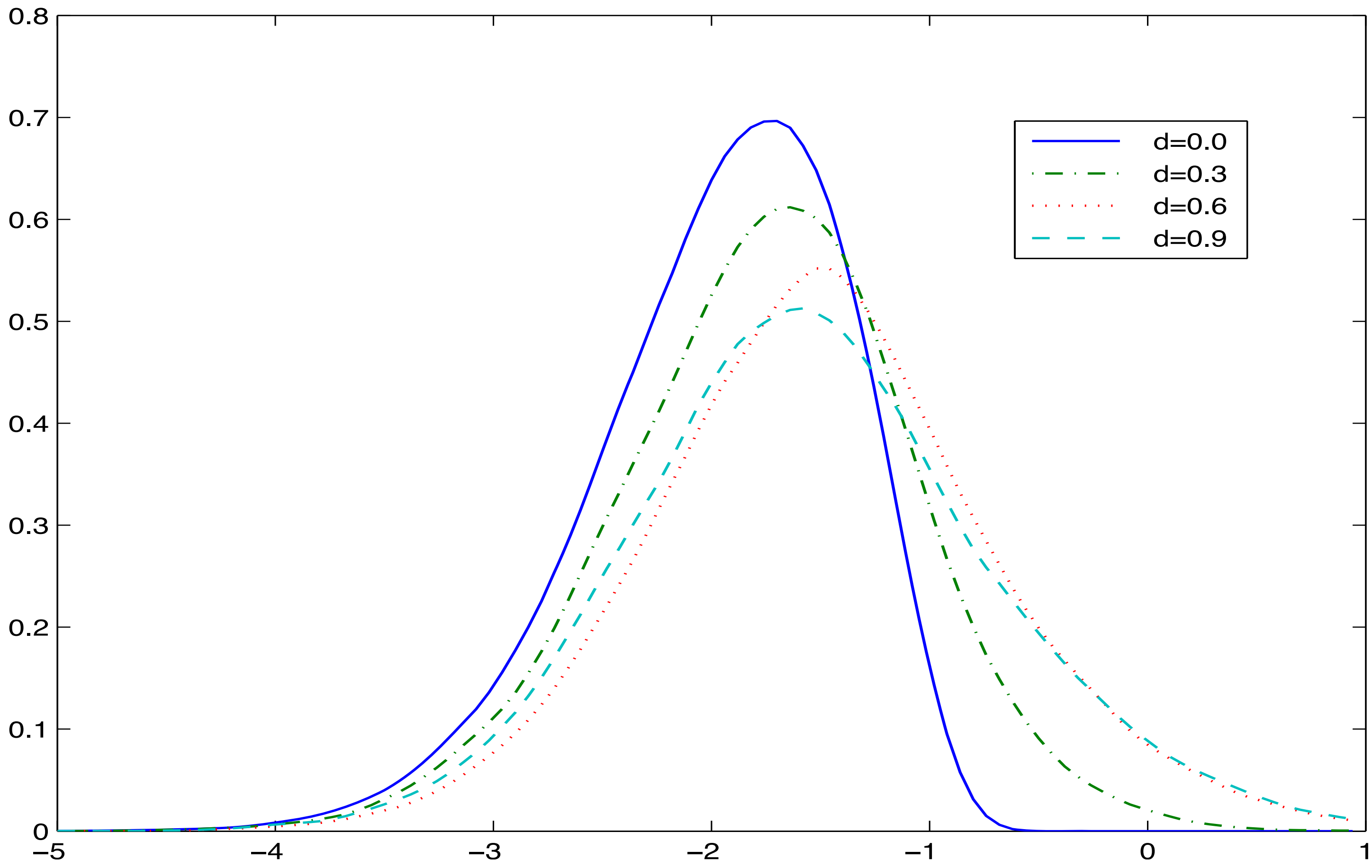
- The additional term appears because of $\hat{\mu}(d)$.
- When X_t has a linear trend, we need to detrend the data first and use a different critical value.

Test using dth differencing

Table 1. Simulated critical values

	Z_t			KPSS		
d	10%	5%	1%	10%	5%	1%
0.0	-2.750	-3.025	-3.556	0.347	0.460	0.736
0.1	-2.710	-2.989	-3.532	0.344	0.460	0.737
0.2	-2.678	-2.960	-3.500	0.342	0.453	0.731
0.3	-2.640	-2.932	-3.469	0.337	0.446	0.715
0.4	-2.600	-2.893	-3.432	0.335	0.440	0.702
0.5	-2.558	-2.850	-3.398	0.334	0.435	0.699
0.7	-2.550	-2.838	-3.430	0.340	0.451	0.721
1.0	-2.563	-2.849	-3.424	0.347	0.460	0.737
DF	-2.57	-2.86	-3.42	0.347	0.463	0.739

Test using dth differencing: densities of Z_t



Test using dth differencing

- Easy to implement

- The Z_t test has power against $I(0)$ spurious long memory, because the $\hat{d}$th differenced data are overdifferenced.

- The KPSS test has power against $I(1)$ spurious long memory, because the $\hat{d}$th differenced data are underdifferenced.

Simulation results with n=5000

$$X_t = (1-L)^{-0.4} u_t \mathbf{1}\{t \geq 1\},\ u_t \sim iidN(0,1)$$

m	mean$(\hat{d})$	rejection freq. (Z_t)			rejection freq. $(\hat{\eta}_\mu)$		
		10%	5%	1%	10%	5%	1%
200	0.400	0.079	0.032	0.004	0.077	0.031	0.002
400	0.400	0.088	0.040	0.005	0.092	0.039	0.004
600	0.399	0.090	0.042	0.007	0.098	0.045	0.007
800	0.397	0.087	0.041	0.007	0.103	0.050	0.009

Simulation results with n=5000

$$X_t = (1-L)^{-0.4} u_t \mathbf{1}\{t \geq 1\},\ u_t \sim iidN(0,1)$$

m	mean($\hat{d}$)	rejection freq. (Z_t)			rejection freq. ($\hat{\eta}_\mu$)		
		10%	5%	1%	10%	5%	1%
200	0.400	0.079	0.032	0.004	0.077	0.031	0.002
400	0.400	0.088	0.040	0.005	0.092	0.039	0.004
600	0.399	0.090	0.042	0.007	0.098	0.045	0.007
800	0.397	0.087	0.041	0.007	0.103	0.050	0.009

$X_t \sim$ mean-plus-noise model, $p = 0.002$

m	mean($\hat{d}$)	rejection freq. (Z_t)			rejection freq. ($\hat{\eta}_\mu$)		
		10%	5%	1%	10%	5%	1%
200	0.424	0.001	0.000	0.000	0.713	0.563	0.252
400	0.324	0.000	0.000	0.000	0.973	0.948	0.872
600	0.276	0.000	0.000	0.000	0.989	0.977	0.936
800	0.247	0.000	0.000	0.000	0.993	0.985	0.953

Simulation results with n=5000

$X_t \sim$ STOPBREAK model, $\gamma = 180$

m	mean$(\hat{d})$	rejection freq. (Z_t)			rejection freq. $(\hat{\eta}_\mu)$		
		10%	5%	1%	10%	5%	1%
200	0.423	0.000	0.000	0.000	0.739	0.581	0.234
400	0.323	0.000	0.000	0.000	0.977	0.955	0.870
600	0.275	0.000	0.000	0.000	0.992	0.981	0.941
800	0.246	0.000	0.000	0.000	0.996	0.988	0.960

Simulation results with n=5000

$X_t \sim$ STOPBREAK model, $\gamma = 180$

m	mean$(\hat{d})$	rejection freq. (Z_t) 10%	5%	1%	rejection freq. $(\hat{\eta}_\mu)$ 10%	5%	1%
200	0.423	0.000	0.000	0.000	0.739	0.581	0.234
400	0.323	0.000	0.000	0.000	0.977	0.955	0.870
600	0.275	0.000	0.000	0.000	0.992	0.981	0.941
800	0.246	0.000	0.000	0.000	0.996	0.988	0.960

$X_t \sim$ Markov-switching model, $p_{00} = p_{11} = 0.93$

m	mean$(\hat{d})$	rejection freq. (Z_t) 10%	5%	1%	rejection freq. $(\hat{\eta}_\mu)$ 10%	5%	1%
200	0.388	0.766	0.650	0.432	0.005	0.003	0.001
400	0.609	0.993	0.988	0.969	0.003	0.002	0.000
600	0.704	0.997	0.996	0.991	0.002	0.001	0.000
800	0.748	0.997	0.997	0.995	0.002	0.001	0.000

Simulation results with n=240

$X_t = (1-L)^{-0.4} u_t \mathbf{1}\{t \geq 1\}$

m	mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
		$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
20	0.387	0.079	0.050	0.073	0.040	0.013	0.068	0.025	0.003
40	0.389	0.070	0.103	0.062	0.024	0.002	0.072	0.025	0.002
60	0.385	0.065	0.092	0.074	0.031	0.004	0.093	0.037	0.004

$X_t \sim$ mean-plus-noise model

m	mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
		$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
20	0.388	0.141	0.054	0.104	0.069	0.036	0.141	0.065	0.007
40	0.397	0.152	0.169	0.004	0.001	0.000	0.499	0.345	0.100
60	0.403	0.161	0.178	0.001	0.000	0.000	0.687	0.548	0.287

Simulation results with n=240

$X_t \sim$ STOPBREAK model

m	mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
		$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
20	0.405	0.100	0.043	0.103	0.069	0.033	0.137	0.053	0.005
40	0.397	0.120	0.124	0.003	0.001	0.000	0.503	0.341	0.092
60	0.377	0.145	0.142	0.001	0.000	0.000	0.707	0.574	0.312

$X_t \sim$ Markov-switching model

m	mean($\hat{d}$)	rej. freq. (W_c)		rej. freq. (Z_t)			rej. freq. ($\hat{\eta}_\mu$)		
		$b=2$	$b=4$	10%	5%	1%	10%	5%	1%
20	0.364	0.107	0.026	0.053	0.023	0.005	0.063	0.040	0.017
40	0.363	0.176	0.146	0.348	0.179	0.038	0.021	0.013	0.002
60	0.470	0.263	0.280	0.696	0.522	0.205	0.011	0.006	0.000

Estimation and test results with S&P 500 log realized standard deviation: 1/2/1985 – 10/25/2004, 5,000 obs

m	$\hat{d}$	$\bar{d}$		W_c		Z_t	$\hat{\eta}_\mu$
		$b=2$	$b=4$	$b=2$	$b=4$		
200	0.525	0.528	0.560	0.849	2.012	-1.443	0.126
400	0.466	0.466	0.489	0.729	3.029	-1.017	0.256
600	0.429	0.428	0.448	0.214	7.848*	-0.802	0.376
800	0.391	0.389	0.402	0.441	4.837	-0.624	0.534*

Note: * indicates rejection of the null at the 5% level.
$\chi^2_{0.95}(1) = 3.84, \chi^2_{0.95}(3) = 7.82.$

m	$\hat{d}$	$\bar{d}$		W_c		Z_t	$\hat{\eta}_\mu$
		$b=2$	$b=4$	$b=2$	$b=4$		
subperiod 1: 1/2/1985 – 12/14/1988							
40	0.549	0.531	0.340	0.068	3.825	-1.496	0.101
100	0.562	0.541	0.483	3.525	8.628*	-1.499	0.096
160	0.449	0.419	0.366	2.520	8.907*	-1.029	0.258
subperiod 2: 12/15/1988 – 11/27/1992							
40	0.410	0.443	0.255	0.757	1.890	-1.379	0.200
100	0.358	0.375	0.304	0.015	1.709	-0.997	0.300
160	0.308	0.324	0.269	0.131	3.234	-0.775	0.404
subperiod 3: 11/30/1992 – 11/11/1996							
40	0.367	0.390	0.364	0.207	1.536	-1.873	0.133
100	0.302	0.307	0.261	0.000	4.580	-1.327	0.246
160	0.274	0.269	0.236	0.010	8.101*	-1.162	0.305
subperiod 4: 11/12/1996 – 10/27/2000							
40	0.485	0.539	0.376	0.054	6.976	-2.572	0.058
100	0.386	0.391	0.321	0.213	12.770*	-1.631	0.128
160	0.341	0.338	0.302	0.935	9.807*	-1.347	0.176
subperiod 5: 10/30/2000 – 10/25/2004							
40	0.656	0.611	0.561	0.013	3.804	-1.761	0.051
100	0.597	0.571	0.527	0.469	2.168	-1.089	0.142
160	0.480	0.465	0.435	0.052	1.873	-0.226	0.517*

Estimation and test results with S&P 500 log realized standard deviation

- The evidence against true $I(d)$ is not strong in view of the power of the tests
- However, both $\hat{d}$ and $\bar{d}$ decrease as m increases, which is consistent with the mean-plus-noise model and STOPBREAK model, suggesting a possibility of a presence of jumps and/or structural breaks in the data.

Estimation and test results with S&P 500 log realized standard deviation

- The evidence against true $I(d)$ is not strong in view of the power of the tests
- However, both $\hat{d}$ and $\bar{d}$ decrease as m increases, which is consistent with the mean-plus-noise model and STOPBREAK model, suggesting a possibility of a presence of jumps and/or structural breaks in the data.
- A pure $I(d)$ process may not explain all of the persistence in the logarithm of the realized volatility, but the data do not support an extreme view that structural breaks account for all the observed persistence.
- A strong possibility of local variation in d
- We may obtain different results with different datasets (e.g. 50 years of daily data)