

Efficient Detrending in the Presence of Fractional Errors*

Yixiao Sun
Department of Economics
Yale University

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

and

Chinchin Lee
Department of Economics
Yale University

November 1999

*Sun thanks the Cowles Foundation for summer support under a Graduate Student Fellowship.

ABSTRACT

This paper examines the efficiency gains of GLS detrending over OLS detrending in the model $y_t = z_t' \gamma + u_t$, where $z_t' = (1, t, \dots, t^{k-1})$ and $u_t \sim I(d)$, $-1/2 < d < 3/2$. A famous theorem on trend removal by OLS regression (usually attributed to Grenander and Rosenblatt (1957) gave conditions for the asymptotic equivalence of GLS and OLS in deterministic trend estimation with $I(0)$ errors. When the errors are fractionally integrated of order $d \neq 0$, this asymptotic equivalence no longer holds. In this case, the asymptotic relative efficiency depends on initialization. Under infinite past initialization, the GLS estimator, when consistent, is not only asymptotically more efficient for any value of d but also converges faster by a rate of $\sqrt{\log(T)}$ when $d = k - 1/2$, $k = 0, 1, 2, \dots$. Under the origin initialization, i.e. $u_t = 0$, for $t \leq 0$, GLS still achieves substantial efficiency gain, but in general, it does not enjoy a faster convergence rate for any d . Two exceptions are (1) a trending variable is of order $t^{d-1/2}$ and the GLS of this specific trend converges faster than OLS by a rate of $\sqrt{\log(T)}$. (2) $d = -0.5$ and GLS estimators of all coefficients converge faster than the OLS by a rate of $\sqrt{\log(T)}$.

JEL Classification: C13; C22

Keywords: Fractional Integration, GLS and OLS detrending, Asymptotic Efficiency, Grenander–Rosenblatt theorem.

1 INTRODUCTION

Grenander and Rosenblatt (1957) gave conditions for the asymptotic equivalence of GLS and OLS estimation in time series regression models with stationary errors such as the model $y_t = z_t' \gamma + u_t$ where $z_t' = (1, t, \dots, t^{k-1})$ and u_t is integrated of order zero, i.e. $I(0)$. This important result relies on the condition that the spectrum of u_t is continuous and nonzero at the origin (where the spectral mass of z_t is concentrated) and is satisfied in most models involving stationary time series. But, the condition is violated whenever u_t is $I(d)$, with $d \neq 0$. For in that case the spectral density, $f_u(\lambda)$, of u_t behaves like a multiple of λ^{-2d} as $\lambda \rightarrow 0$ and is unbounded or equal to zero at $\lambda = 0$. In such cases the asymptotic equivalence of GLS and OLS breaks down and we have efficiency gains in estimating the trend coefficients γ using GLS even in large samples.

This paper calculates the asymptotic efficiency gains of OLS relative to GLS in trend parameter estimation when u_t is $I(d)$ with $-\frac{1}{2} < d < \frac{3}{2}$. OLS is different from GLS unless $d = 0$. However, GLS is computationally more difficult since we must first calculate the covariance matrix of u_t while OLS is always available and consistent. Hence it is important to compare OLS with GLS for various error correlation structure. It turns out the efficiency comparison depends crucially on how the fractional error process is initialized. When the process is initialized at the infinite past, the case of $d = 1/2$ needs separate treatment. In this case, the OLS is infinitely deficient. When the process is initialized at the origin, we need to single out other cases, for example, the location model where $d = 1/2$ and $z_t = 1$. Since the asymptotic relative efficiency is an almost everywhere continuous function of d , we can infer its magnitude for a given d_0 by referring to the point near d_0 . $d = 1/2$ provides an interesting reference point. The recent studies of long memory process, including Bollerslev and Mikkelsen(1993) on stock market volatility, and Hassler and Wolters(1995) on inflation rates, suggest that the long memory coefficient in stock market volatility is around $\frac{1}{2}$. Therefore $I(\frac{1}{2})$ processes are not only of theoretical interests but also of empirical relevance.

It has been suggested that fractionally integrated processes model some time series better than the more widely used models where $d = 0$ or $d = 1$ (see, for instance, Granger and Joyeux (1980), and Baillie (1996)). Fractional models encompass both stationary and nonstationary processes depending on the value of the parameter d and include the weakly dependent and unit root processes as special cases. For these reasons, fractional processes are attractive to empirical analysts, providing much richer impulse response dynamics than the classical dichotomy of $I(0)$ and $I(1)$ processes. One central issue is to test when long memory really exists in economic time series. In the presence of deterministic trends, the size and power of such tests based on asymptotics certainly depend on whether we can efficiently extract the trends. Therefore efficiency consideration is related not only to the estimation of the trending parameter but also to the detection of long memory.

Phillips and Lee(1996) investigate the unit root case ($d = 1$) and relate it to the efficient unit root test. Yajima(1988), Beran(1994, ch. 9) and Samarov and Taqqu(1988) examine stationary long-memory cases with $0 \leq d < 1/2$. All those studies initialize long memory processes at the infinite past. We will investigate, under both initialization at the infinite

past and initialization at the origin, the efficiency gains of using GLS over OLS in such cases when u_t is both stationary ($-\frac{1}{2} < d < 1/2$) and nonstationary ($d \geq 1/2$). We will relate our results to those already in the literature.

The plan of the paper is as follows: Section 2 gives the corresponding theory for long-memory processes. Section 3 studies and graphs the efficiency gains in a linear trends model when the error process is initialized at the infinite past. Section 4 examines the efficiency gains when the error process is initialized at the origin. Section 5 concludes the paper. Proofs are given in the Appendix.

2 LIMIT THEOREMS FOR FRACTIONAL PROCESSES

Under infinite past initialization, a *stationary* fractionally integrated process u_t can be represented as

$$u_t = (1 - L)^{-d} v_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} v_{t-j}, \quad t = 1, 2, \dots, T \quad (1)$$

where $d \in [-\frac{1}{2}, \frac{1}{2})$, and v_t ($-\infty < t < \infty$) is stationary with zero mean and continuous spectrum $f_v > 0$. Throughout the paper, it will be convenient to assume that v_t follows a linear process of the form

$$v_t = \phi(L)\varepsilon_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j}, \quad \sum_{l=0}^{\infty} l|\phi_l| < \infty, \quad \phi = \sum_{l=0}^{\infty} \phi_l \neq 0$$

for all t with ε_t is iid($0, \sigma^2$). The above summability condition is satisfied by a wide class of parametric and nonparametric models for v_t , enables the use of Phillips-Solo device and ensures that the partial sums of v_t satisfy a functional central limit theorem.

Nonstationary fractional processes are defined as \bar{d} fold partial sums of stationary $I(d-\bar{d})$ processes, where \bar{d} is the integer closest to d and equal to $d + 1/2$ when $d - [d] = 1/2$. In view of (1), it is natural to define $f_u(\lambda) = |1 - e^{i\lambda}|^{-2d} f_v(\lambda) \sim \lambda^{-2d} f_v(\lambda)$ as $\lambda \rightarrow 0$. The function $f_u(\lambda)$ gives the spectrum of u_t when it exists and u_t is stationary (i.e. for $d < 1/2$ and under infinite past initialization of u_t as in (1)) and is the analogue of the spectrum in the nonstationary case when $d \geq 1/2$ even though it is not integrable. When $v_t = \varepsilon_t$, we can deduce from $f_u(\lambda)$ the autocovariances of u_t , γ_j , as

$$\gamma_j = \sigma^2 \frac{\Gamma(1-2d)\Gamma(d+j)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d+j)}, \quad d < 1/2$$

so the autocovariances have the same sign as d for $j \geq 1$ and satisfy

$$\gamma_j \sim \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} j^{2d-1}, \quad \text{as } j \rightarrow \infty$$

These autocovariances decay slower than those of stationary ARMA processes. When $d < 0$, autocovariances are absolutely summable and u_t is sometimes referred to as an intermediate memory process. When $0 < d < \frac{1}{2}$, autocovariances are not absolutely summable and u_t is a stationary long-memory process. When $d \geq 1/2$, the covariances are infinite and u_t is a nonstationary long-memory process. See Baillie(1996) for more details.

In both stationary and nonstationary cases, violation of the Grenander–Rosenblatt condition suggests that GLS will be more efficient than OLS when estimating trend parameters. For both cases, we need functional central limit theorems for long-memory processes. The limit theorems involve weak convergence of sample functionals to fractional Brownian motions processes, which were introduced and studied in Mandelbrot and Van Ness (1968). Fractional Brownian motion, denoted here by B_H , is a Gaussian process defined for $H \in (0, 1]$ on $\mathcal{D}[0, 1]$ such that $B_H(0) = 0$ with covariance kernel

$$E[B_H(r)B_H(s)] = (1/2)[r^{2H} - |r - s|^{2H} + s^{2H}].$$

The process $B_H(r)$ can be written as

$$B_H(r) = \begin{cases} c \left\{ \int_{-\infty}^0 [(r-s)^{H-1/2} - (-s)^{H-1/2}] dW(s) + \int_0^r (r-s)^{H-1/2} dW(s) \right\}, & H \in (0, 1) \\ rW(1), & H = 1 \end{cases}$$

with $c = \left\{ (1/(2H) + \int_{-\infty}^0 [(1-s)^{H-1/2} - (-s)^{H-1/2}]^2 ds \right\}^{-1/2}$, and it reduces to standard Brownian motion, $W(r)$, when $H = 1/2$. Samorodnitsky and Taqqu(1994) provide an extensive review of the subject.

Now we are ready to state the functional limit theorems for fractional processes. Theorems of this type are given in Taqqu(1975) and they have been used recently in econometrics by Sowell(1990) and Liu(1998) in the context of fractional unit root limit theory. Usually those theorems assume that v_t is *iid*. To derive a functional central limit theorem when v_t is weakly dependent, we employ the Beveridge-Nelson (B-N) decomposition(Phillips-Solo device) to get:

$$v_t = \phi \varepsilon_t - (1 - L)\tilde{\varepsilon}_t$$

where $\tilde{\varepsilon}_t$ is a stationary process defined by

$$\tilde{\varepsilon}_j = \sum_{l=0}^{\infty} \tilde{\phi}_l \varepsilon_{j-l}, \quad \tilde{\phi}_l = \sum_{k=l+1}^{\infty} \phi_k$$

Consider

$$T^{-(1+2d)/2} \sum_{t=1}^{[Tr]} u_t = \phi T^{-(1+2d)/2} \sum_{t=1}^{[Tr]} (1-L)^{-d} \varepsilon_t + T^{-(1+2d)/2} (1-L)^{-d} (\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[Tr]}), \quad d \in (-1/2, 1/2)$$

Now for $-1/2 < d < 1/2$, $(1-L)^{-d}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nr]})$ is a stationary process with mean 0 and finite variance, so that $T^{-(1+2d)/2}(1-L)^{-d}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[Tr]}) \rightarrow_p 0$. On the other hand, $(1-L)^{-d}\varepsilon_t$ is a

fractional process constructed from a $iid(0, \sigma^2)$ with $E|\varepsilon_t|^p < \infty$, hence $T^{-(1+2d)/2} \sum_{t=1}^{[Tr]} (1-L)^{-d} \varepsilon_t$ satisfies a functional central limit theorem and so does $T^{-(1+2d)/2} \sum_{t=1}^{[Tr]} u_t$. Let $K(d) = \{\Gamma(1-2d)/[(1+2d)\Gamma(1+d)\Gamma(1-d)]\}^{1/2}$, then we have:

Lemma 1 *If $d \in (-1/2, 1/2)$, ε_t is an independent white noise with $E|\varepsilon_t|^p < \infty$ for some $p \geq \max(4, -8d/(1+2d))$, then for $r \in (0, 1]$,*

$$T^{-(1+2d)/2} \sum_{t=1}^{[Tr]} u_t \Rightarrow \sigma \phi K(d) B_H(r)$$

with $H = d + 1/2$.

From Lemma 1 we get Lemma 2 immediately.

Lemma 2 *If $d \in (1/2, 3/2)$, ε_t is an independent white noise with finite absolute moments of order $p \geq \max(-8(d-1)/(2d-1), 4)$, Then*

$$T^{-(2d-1)/2} u_{[Tr]} \Rightarrow \sigma \phi K(d-1) B_H(r)$$

with $H = d - 1/2$.

Again using Phillips-Solo Device and Theorem 2.2 in Liu (1998), we obtain:

Lemma 3 *If $d = \frac{1}{2}$, ε_t is an independent white noise with $E|\varepsilon_t|^2 < \infty$, then for $r \in (0, 1]$,*

$$\frac{1}{T\sqrt{\log(T)}} \sum_{t=1}^{[Tr]} u_t \Rightarrow \frac{\sigma \phi}{\sqrt{\pi}} r W(1)$$

In Lemma 1, where u_t is stationary, the sum of u_t appropriately normalized behaves like a scaled fractional Brownian motion in the limit. In Lemma 2, u_t is nonstationary and, appropriately standardized, u_t itself behaves in the limit like a scaled fractional Brownian motion. In Lemma 3, u_t is just nonstationary in the sense that any process with $d = \frac{1}{2} - \epsilon$, for any positive ϵ , is stationary. In this case, the normalization factor has additional part of $\sqrt{\log(T)}$ comparing with the other points near $\frac{1}{2}$. This seemingly puzzling phenomenon can be explained by taking $d = \frac{1}{2} + \epsilon$. From Lemma 2, $T^{-\epsilon} K^{-1}(d-1) u_{[Tr]} = O_p(1)$. By noting that $T^{2\epsilon} K^2(d-1) = T^{2\epsilon} \Gamma(2-2\epsilon)/[2\epsilon \Gamma(1/2+\epsilon) \Gamma(3/2-\epsilon)] \sim \Gamma(2-2\epsilon)/[\Gamma(1/2+\epsilon) \Gamma(3/2-\epsilon)] \frac{T^{2\epsilon}-1}{2\epsilon}$ as $T \rightarrow \infty$ and that $(T^{2\epsilon}-1)/2\epsilon \rightarrow \log(T)$ as $\epsilon \rightarrow 0$, we deduce that $u_{[Tr]} = O_p(\sqrt{\log(T)})$ and thus $\sum u_{[Tr]} = O_p(T\sqrt{\log(T)})$ when $d = 1/2$. Extending functional central limit theorems to the cases of $d \geq 3/2$ by applying the continuous mapping theorem on the

basis of low-order limit results, we find that this feature remains for any $d = (2k + 1)/2$, $k \geq 0$. We will focus on $d \in (-1/2, 3/2)$, which is general enough for empirical analysis.

The above analysis relies on infinite past initialization. In contrast, Akonom and Gouriou(1987) initialize long memory processes at the origin. They assume $\varepsilon_t = 0$ for all $t \leq 0$ so $v_t = 0$ for all $t \leq 0$ and

$$u_t = (1 - L)^{-d}v_t = \sum_{j=0}^{t-1} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}v_{t-j}, t = 1, 2, \dots, T, \quad (2)$$

which essentially provides an alternative definition of fractional sum operator. Different specifications of fractional sum imply different definitions of fractional difference, as fractional integration of order d is defined as fractional difference of order $-d$. (1) implies

$$(1 - L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}L^j \quad (3)$$

while (2) implies

$$(1 - L)^{-d} = \sum_{j=0}^{t-1} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}L^j \quad (4)$$

Actually, (3) and (4) are two special cases of a large family of fractional sums

$$(1 - L)^{-d} = \sum_{j=0}^{t-1+a} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}L^j, a \geq 0$$

We will not consider the above general case. Instead we investigate two polar cases when $a = 0$ and $a = \infty$, which correspond to initialization at the origin and the infinite past. Our results can be easily extended to the case of $a = [n\kappa]$, so that the first observation x_0 has the same stochastic order as x_t for $t = O(n)$.

For $a = 0$, we first define a fractional Brownian motion initialized at the origin as

$$\vec{B}_H(r) = (2H)^{1/2} \int_0^r (r - s)^{H-1/2} dW(s), H \in (0, 1]$$

This definition differs from the previous one in that it ignores the prehistoric influence, which changes the nature of the process. In fact, $\vec{B}_H(r)$ and $B_H(r)$ have different covariance kernels:

$$E[\vec{B}_H(r)\vec{B}_H(s)] = 2H \int_0^{r \wedge s} (r - t)^{H-1/2}(s - t)^{H-1/2} dt.$$

This difference turns out to be very important for later variance calculation.

Akonom and Gouriou's (1987) gave a FCLT by employing a moving average representation of a fractional process. As it is well known, when a fractional process is initialized

at the infinite past and $d > 1$, it does not have a moving average representation. In contrast, when a process is initialized at the origin, it has a moving average representation, i.e., $(1 - L)^{-d} = \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j$ for any d , which is proved in the Appendix.

With valid moving average representations, the following lemmas follow smoothly from Akonom and Gourieroux(1987).

Lemma 4 *If $d \in (-1/2, 1/2)$, $\varepsilon_s = 0$ for $s \leq 0$ and ε_t is an independent white noise with finite absolute moments of order $p > \max((d + 1/2)^{-1}, 2)$, then*

$$T^{-(2d+1)/2} \sum_{t=1}^{[Tr]} u_t \Rightarrow (\sigma\phi/\Gamma(d+1))(2H)^{-1/2} \vec{B}_H(r), r \in (0, 1] \text{ with } H = d + 1/2.$$

Lemma 5 *If $d \in (1/2, 3/2)$, $\varepsilon_s = 0$ for $s \leq 0$ and ε_t is an independent white noise with finite absolute moments of order $p > \max((d - 1/2)^{-1}, 2)$, then*

$$\begin{aligned} T^{-(2d-1)/2} u_{[Tr]} &\Rightarrow \frac{\sigma\phi}{\Gamma(d)} \int_0^r (r-s)^{d-1} dW(s), r \in (0, 1] \\ &= (\sigma\phi/\Gamma(d))(2H)^{-1/2} c \text{ with } H = d - 1/2. \end{aligned}$$

Lemma 6 *If $d = \frac{1}{2}$, $\varepsilon_s = 0$ for $s \leq 0$ and ε_t is an independent white noise with $E|\varepsilon_t|^2 < \infty$, then for $r \in (0, 1]$,*

$$\frac{1}{T} \sum_{t=1}^{[Tr]} u_t \Rightarrow \sigma\phi \sqrt{\frac{2}{\pi}} \vec{B}_1(r)$$

Note that one can obtain Lemma 6 from Lemma 4 by letting $d \rightarrow 1/2$, while one can not get Lemma 3 in this way. This implies that, as we will see, at $d = 1/2$ the asymptotic efficiency is generally continuous with the origin initialization while unbounded with the infinite past initialization.

These FCLTs reveal that initialization has significant effects on the asymptotic distributions. For a stationary short memory $ARMA(p, q)$ process, the effects of different initializations are usually asymptotically negligible. For a fractionally integrated process, even if it is stationary, different initializations cause different asymptotics. More interestingly, when initialized at the origin the partial sums of an $I(1/2)$ process converge slower by a rate of $\sqrt{\log(T)}$. Meanwhile, initialization for a fractional process involves infinite many time points while initialization for a $ARMA(p, q)$ process and a unit root process involves at most finite many time points. Therefore, initialization is of vital importance for a fractional process. Because of long memory, historic changes could affect future trajectory dramatically. We thus expect that the efficiency of trend extraction depends on initialization assumed. Section 3 assumes the process is initialized at the infinite past while section 4 assumes it is initialized at the origin.

3 EFFICIENCY GAINS WITH INITIALIZATION AT THE INFINITE PAST

We first derive the limiting distributions of the OLS and GLS estimator of the trend coefficients when u_t is stationary. Consider the following data generating process

$$y_t = \alpha + (t, \dots, t^{k-1})\beta + u_t = z_t'\gamma + u_t, \quad t = 1, 2, \dots, T \quad (5)$$

where $z_t' = (1, t, \dots, t^{k-1})$, $u_t = (1-L)^{-d}v_t$, $\gamma' = (\alpha, \beta')$ and $v_t = \phi(L)\varepsilon_t = \sum_{j=0}^{\infty} \phi_l \varepsilon_{t-l}$, $\sum_{j=0}^{\infty} l|\phi_l| < \infty$ with $d \in (-1/2, 1/2)$ and $\varepsilon_t \sim iid(0, \sigma^2)$.

Theorem 1 *Assume $E|\varepsilon_t|^{2p} < \infty$ for some $p \geq \max(4, -8d/(1+2d))$, $d \in (-1/2, 1/2)$ and $H = d + 1/2$, the OLS estimator $\hat{\gamma}$ of γ has the limiting distribution*

$$\begin{aligned} D_T^{1/2}(\hat{\gamma} - \gamma)/T^d &\Rightarrow \sigma\phi K(d)Q^{-1} \left(\mathbf{1}B_H(1) - \int_0^1 \bar{g}(r)B_H(r)dr \right) \\ &\equiv N(0, \sigma^2\phi^2 K^2(d)Q^{-1}\Omega_{ols}Q^{-1}), \end{aligned}$$

where $D_T = \text{diag}(T^1, T^3, \dots, T^{2k-1})$, Ω_{ols} is a $k \times k$ matrix with elements $\{\omega_{ij}\}$,

$$\omega_{ij} = \begin{cases} 1/(i+j-1), & \text{if } d = 0, \\ d(1+2d) \int_0^1 \int_0^1 r^{i-1} s^{j-1} |r-s|^{2d-1} dr ds, & \text{otherwise} \end{cases}$$

$\bar{g}(r) = (0, 1, 2r, \dots, (k-1)r^{k-2})'$ and $\mathbf{1}$ is a column vector of ones.

The next result follows from and extends Theorem 2.3 of Yajima (1988), who considered the case $0 < d < 1/2$. The proof is omitted because of its similarity with the arguments there.

Theorem 2 *Assume $d \in [-1/2, 1/2)$. Let $\tilde{\gamma}$ be the GLS estimator of γ , then*

$$\lim_{T \rightarrow \infty} D_T^{1/2} E(\tilde{\gamma} - \gamma)(\tilde{\gamma} - \gamma)' D_T^{1/2} / T^{2d} = \sigma^2 \phi^2 (\Omega_{gls})^{-1},$$

where Ω_{gls} is a $k \times k$ matrix with elements $\{\omega_{ij}\}$,

$$\omega_{ij} = \Gamma(i-d)\Gamma(j-d)/[\Gamma(i-2d)\Gamma(j-2d)(i+j-1-2d)].$$

We define the asymptotic relative efficiency, *ARE* of $\hat{\gamma}$ to $\tilde{\gamma}$ as

$$ARE = \det[K^2(d)Q^{-1}\Omega_{ols}Q^{-1}] / \det[(\Omega_{gls})^{-1}].$$

When $z_t = 1$, from Theorems 1 and 2,

$$T^{(3-2d)/2}(\hat{\alpha} - \alpha) \Rightarrow \sigma\phi K B_H(1) \equiv N(0, \sigma^2\phi^2 K^2)$$

and

$$T^{3-2d}E(\tilde{\alpha} - \alpha)^2 = \sigma^2\phi^2[\Gamma(1-2d)]^2(1-2d)/[\Gamma(1-d)]^2,$$

so the asymptotic relative efficiency, ARE of $\hat{\alpha}$ to $\tilde{\alpha}$ in this case is:

$$ARE^c = \frac{\Gamma(1-d)}{(1+2d)\Gamma(1+d)\Gamma(2-2d)}$$

Figure 1 graphs how ARE^c varies with d over the interval $(-1/2, 1/2)$. It shows that the efficiency gain is very small for $d > 0$, achieving a maximum of only 1.019 when $d = 0.32$. This can be explained by the observation that too high positive error autocorrelation offsets the effect of efficient weighting. Therefore, the sample mean is reasonably efficient. Figure 1 also reveals that $ARE^c \rightarrow 1$ as $d \rightarrow 0.5$. The intuition is, as $d \rightarrow 0.5$, the population mean is almost unidentifiable in the sense that it can not be consistently estimated when $d = 0.5 + \epsilon$ for any $\epsilon \geq 0$. As $d \rightarrow 0.5$, the asymptotic variance of any consistent estimator diverges to infinite. For the OLS estimator and the GLS estimator, the divergence rates happen to be the same so $ARE^c \rightarrow 1$. In contrast, when $d < 0$, negative autocorrelation strengthens the effect of efficient weighting thus the GLS estimator gains much more efficiency. As $d \rightarrow -0.5$, $ARE^c \rightarrow \infty$. Note that, when $d = -0.5$, Theorem 2 implies that $\tilde{\alpha} - \alpha = O_p(T^{-1})$ and Theorem 2.3 in Liu(1998) implies that $\hat{\alpha} - \alpha = \sum_{t=1}^T u_t/T = O(T^{-1}\sqrt{\log T})$, thus GLS is infinite more efficient than OLS when $d = -0.5$.

When $z_t = t$, i.e. there is no fitted intercept, we can deduce that

$$T^{(3-2d)/2}(\hat{\beta} - \beta) \Rightarrow 3\sigma\phi K [B_H(1) - \int_0^1 B_H(r)dr] \equiv N(0, 9\sigma^2\phi^2 K^2/(3+2d))$$

and

$$T^{3-2d}E(\tilde{\beta} - \beta)^2 = \sigma^2\phi^2[\Gamma(2-2d)]^2(3-2d)/[\Gamma(2-d)]^2,$$

so the asymptotic relative efficiency of $\hat{\beta}$ to $\tilde{\beta}$ is

$$ARE^t = \frac{9\Gamma(1-2d)[\Gamma(2-d)]^2}{(1+2d)(9-4d^2)\Gamma(1+d)\Gamma(1-d)[\Gamma(2-2d)]^2}.$$

Figure 2 graphs ARE^t against d over the interval $d \in (-1/2, 1/2)$. It reveals that the efficiency gain increases with the increase of the absolute value of d , and $ARE^t \rightarrow \infty$ as $|d| \rightarrow 0.5$. The property of ARE^t as d approaches and equals to 0.5 is further explored later on in this section. For $d = -0.5$, we deduce from theorem 2 that $\tilde{\beta} - \beta = O_p(T^{-2})$ and from Theorem 2.3 in Liu(1998) that $\hat{\beta} - \beta = \sum tu_t / \sum t^2 = O_p(T \sum u_t / T^3) = O_p(T^{-2}\sqrt{\log T})$. So again GLS is infinite more efficient than OLS when $d = -0.5$.

Note that in all cases, $ARE = 1$ when $d = 0$, as we would expect from the Grenander-Rosenblatt theorem.

Theorems 1 and 2 dealt with stationary fractional process u_t . For $d \geq 1/2$, the process is nonstationary. In this case, we can only derive asymptotic properties conditioning on the sample path going through some point. Hence, without loss of generality, we assume that all the nonstationary process take value 0 when $t = 0$.

Theorem 3 and 4 give the limiting distributions of the OLS estimators $\hat{\gamma}$ when $d \geq 1/2$.

Theorem 3 *If $d = 1/2$ and ε_t satisfies the assumptions in Lemma 3, the OLS estimator $\hat{\gamma}$ of γ is distributed asymptotically as*

$$F_T(\hat{\gamma} - \gamma)/\sqrt{\log(T)} \Rightarrow N(0, \sigma^2 \phi^2 \pi^{-1} Q^{-1} \Omega_{ols} Q^{-1'}),$$

where $F_T = \text{diag}(1, T, T^2, \dots, T^{k-1})$, Ω_{ols} is a $k \times k$ matrix with elements $\{\omega_{ij}\}, \omega_{ij} = 1/(ij)$.

Theorem 4 *If $1/2 < d < 3/2$ and ε_t satisfies the assumptions in Lemma 2, the OLS estimator $\hat{\gamma}$ of γ is distributed asymptotically as*

$$\begin{aligned} D_T^{1/2}(\hat{\gamma} - \gamma)/T^d &\Rightarrow \sigma \phi K(d-1) Q^{-1} \int_0^1 g(r) B_H(r) dr \quad \text{for } H = d - 1/2, \\ &\equiv N(0, \sigma^2 \phi^2 K^2(d-1) Q^{-1} \Sigma Q^{-1'}) \end{aligned}$$

where $\Sigma = 1/2 \int_0^1 \int_0^1 g(r)(r^{2d-1} - |r-s|^{2d-1} + s^{2d-1})g(s)' ds dr$, and $g(r) = (1, r^1, \dots, r^{k-1})'$.

Theorem 5 presents the asymptotic variance of the GLS estimator.

Theorem 5 *Assume $1/2 \leq d < 3/2$ and ε_t satisfies the assumptions in Lemma 2 ($d > 1/2$) or Lemma 3 ($d = 1/2$), then the GLS estimator of α , $\tilde{\alpha}$ is inconsistent while the GLS estimator of β , $\tilde{\beta}$ satisfies*

$$\lim_{T \rightarrow \infty} (D_T^\beta)^{1/2} (E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' (D_T^\beta)^{1/2} / T^{2d} = \sigma^2 \phi^2 (\Omega_{gl_s}^\beta)^{-1},$$

where $D_T^\beta = \text{Diag}(T^3, \dots, T^{2k-1})$, $\Omega_{gl_s}^\beta$ is a $(k-1) \times (k-1)$ matrix with elements $\{\omega_{ij}\}$,

$$\omega_{ij} = ij\Gamma(i-d+1)\Gamma(j-d+1)/[\Gamma(i-2d+2)\Gamma(j-2d+2)(i+j-2d+1)].$$

The following corollary is a specialization of the preceding theorem.

Corollary 1 *If $d = 1/2$ and ε_t satisfies the assumptions in Lemma 3, then*

$$\lim_{T \rightarrow \infty} F_T^\beta E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' F_T^\beta = \sigma^2 \phi^2 (\Omega_{gl_s}^\beta)^{-1}, \quad F_T^\beta = \text{diag}(T, T^2, \dots, T^{k-1}).$$

When $d \geq 1/2$, neither OLS nor GLS can consistently estimate the intercept. Nevertheless, for other trending parameters, the GLS estimator converges faster by a rate of $\sqrt{\log(T)}$ than the OLS estimator when $d = 1/2$. Therefore, in this case the relative asymptotic efficiency of $\hat{\beta}$ to $\tilde{\beta}$, $ARE = O(\sqrt{\log(T)})$ and we say that OLS is infinite deficient. When $d > 1/2$ and $z_t = (1, t)$, we have

$$T^{(3-2d)/2}(\hat{\beta} - \beta) \Rightarrow N\left(0, \sigma^2 \phi^2 \frac{9(2d^2 + 3d - 1)\Gamma(3 - 2d)}{\Gamma(d)\Gamma(2 - d)2d(4d^2 - 1)(3 + 2d)}\right)$$

and

$$T^{(3-2d)/2}(\tilde{\beta} - \beta) \Rightarrow N(0, \sigma^2 \phi^2 (3 - 2d)\Gamma^2(3 - 2d)/\Gamma^2(2 - d)).$$

Therefore

$$ARE = \frac{9(2d^2 + 3d - 1)\Gamma(2 - d)}{2(9 - 4d^2)(4d^2 - 1)\Gamma(d + 1)\Gamma(3 - 2d)}.$$

When $d = 1$, this specializes to $ARE = 6/5$ corresponding to the result given in Phillips and Lee (1996) when u_t is integrated of order 1.

Figure 3 graphs the relative efficiency of $\hat{\beta}$ over $\tilde{\beta}$ when $d \in (1/2, 3/2)$. Observe that ARE monotonically decreases over the interval $(1/2, 3/2)$ and the ARE in figure 2 is "U" shaped. The relative efficiency of GLS over OLS approaches 1 as d approaches $3/2$. As the noise becomes stronger with the increase of d , the signal becomes relatively weaker and finally unidentifiable when $d = 3/2$, so the advantage of using GLS disappears eventually. The asymptote at $d = 1/2$ shows that GLS achieves its greatest efficiency gains over OLS as $d \rightarrow 1/2$ from above. Figure 2 confirms that this also applies as $d \rightarrow 1/2$ from below. Thus, the case where u_t is fractionally integrated and "just nonstationary" at $d = 1/2$ produces the greatest efficiency gains for GLS in trend estimation and these gains are infinite when $d = 1/2$. The intuition behind this result is that we can expect the gains from GLS to be greatest when u_t is stationary (and hence $O_p(1)$ as $t \rightarrow \infty$) but has the sharpest discontinuity in its spectrum at the origin. This occurs when $d \rightarrow 1/2$ from below.

4 EFFICIENCY GAINS WITH INITIALIZATION AT THE ORIGIN

Under the assumption that $\varepsilon_t = 0$ for all $t \leq 0$, the exact GLS estimator $\tilde{\gamma}$ is the same as the GLS-detrending estimators $\hat{\gamma}$, which is defined as

$$\hat{\gamma}_{gls} \equiv (\Sigma_1^T \tilde{z}_t [Var(v)]^{-1} \tilde{z}_t')^{-1} \Sigma_1^T \tilde{z}_t^{-1} [Var(v)]^{-1} \tilde{y}_t$$

where $\tilde{z}_1 = z_1$, $\tilde{y}_t = y_t$, $\tilde{z}_t = (1 - L)^d z_t$, $\tilde{y}_t = (1 - L)^d y_t$ for $t = 2, \dots, T$ and $Var(v)$ is the variance matrix of $(v_1, v_2, \dots, v_t)'$. By Grenander-Rosenblatt Theorem, GLS-detrending estimator is asymptotically equivalent to the hybrid-detrending estimator defined as

$$\hat{\gamma} \equiv (\Sigma_1^T \tilde{z}_t \tilde{z}_t')^{-1} \Sigma_1^T \tilde{z}_t \tilde{y}_t$$

Therefore as far as the asymptotic properties are concerned, we only need to consider the hybrid-detrening estimator.

With lemmas 4, 5 and 6, we now state the asymptotic distributions of the OLS and the GLS estimators. Proofs are omitted when they are essentially the same as corresponding proofs for theorems in section 3.

Theorem 6 *If $d \in (-1/2, 1/2)$, $H = d + 1/2$ and ε_t satisfies the assumptions in Lemma 4, then the OLS estimator $\hat{\gamma}$ of γ has the limiting distribution*

$$\begin{aligned} D_T^{1/2}(\hat{\gamma} - \gamma)/T^d &\Rightarrow \frac{\sigma\phi}{\Gamma(d+1)}(2H)^{-1/2}Q^{-1} \left(\mathbf{1}\vec{B}_H(1) - \int_0^1 \bar{g}(r)\vec{B}_H(r)dr \right) \\ &\equiv N(0, \sigma^2\phi^2(2H)^{-1}(\Gamma(d+1))^{-2}Q^{-1}V_{ols}Q^{-1}), \end{aligned}$$

where $V_{ols} = \text{Var} \left(\mathbf{1}\vec{B}_H(1) - \int_0^1 \bar{g}(r)\vec{B}_H(r)dr \right)$ and $\bar{g}(r) = (0, 1, 2r, \dots, (k-1)r^{k-2})'$

Using Lemma 6.7 (given in the appendix), we have Theorem 7.

Theorem 7 *If $d \in (-1/2, 1/2)$ and ε_t satisfies the assumptions in Lemma 4, then the GLS estimator $\tilde{\gamma}$ of γ is distributed asymptotically as*

$$\begin{aligned} D_T^{1/2}(\tilde{\gamma} - \gamma)/T^d &\Rightarrow \sigma\phi V_{gls}^{-1} \int_0^1 \tilde{g}(r)dW(r) \\ &\equiv N(0, \sigma^2\phi^2 V_{gls}^{-1}) \end{aligned}$$

where V_{gls} is a $k \times k$ matrix with elements $\{v_{ij}\}$, $v_{ij} = (i-1)!(j-1)!/[\Gamma(i-d)\Gamma(j-d)(i+j-2d-1)]$ and $\tilde{g}(r)$ is a column vector with its i -th elements $(i-1)![\Gamma(i-d)]^{-1}r^{i-1-d}$.

When $z_t = 1$ and $d \in (-1/2, 1/2)$, Theorems 6 and 7 tell us

$$T^{(3-2d)/2}(\hat{\alpha} - \alpha) \Rightarrow N(0, \sigma^2\phi^2/[(2d+1)\Gamma^2(d+1)])$$

and

$$T^{3-2d}E(\tilde{\alpha} - \alpha)^2 = \sigma^2\phi^2(1-2d)[\Gamma(1-d)]^2,$$

so the asymptotic relative efficiency, ARE of $\hat{\alpha}$ to $\tilde{\alpha}$ in this case is:

$$ARE^c = \frac{1}{(1-4d^2)\Gamma^2(1-d)\Gamma^2(1+d)}$$

Figure 4 graphs how ARE^c varies with d over the interval $(-1/2, 1/2)$. Note that $ARE^c \rightarrow \infty$ as $d \rightarrow -0.5$, which exactly happens in figure 1. When $d = -0.5$, $\text{Var}(\hat{\alpha} - \alpha) = \text{Var}(\sum_{t=1}^T u_t/T) = T^{-2} \sum_{t=1}^T \{\Gamma(t+d+1)/[\Gamma(t+1)\Gamma(d)]\}^2 = T^{-2}O(\sum_{t=1}^T t^{-2d}) = O(T^{-2} \log(T))$

so $\hat{\alpha} - \alpha = O_p(T^{-1}\sqrt{\log T})$ while the GLS estimator is asymptotically equivalent to the OLS estimator of α in $\tilde{y}_t = \alpha t^{-d} + v_t$ so $\hat{\alpha} - \alpha = O_p(T^{-1})$. This explains why $ARE^c \rightarrow \infty$ as $d \rightarrow -0.5$. Comparing with figure 1, the GLS estimator gains substantial efficiency not only for large negative d but also for large positive d . Therefore when we estimate the location model, the OLS estimator (simple average) is reasonably efficient only if $d > 0$ and we assume infinite past initialization. In fact when $d = 1/2$, the OLS estimator of the intercept is inconsistent while the GLS estimator converges by a rate of $\sqrt{\log(T)}$, which will be shown in Theorem 8 and 9.

When $z_t = t$, i.e. there is no fitted intercept, we can deduce that

$$\begin{aligned} T^{(3-2d)/2}(\hat{\beta} - \beta) &\Rightarrow 3\sigma\phi(2d+1)^{-1/2}\Gamma^{-1}(d+1)[\vec{B}_H(1) - \int_0^1 \vec{B}_H(r)dr] \\ &\equiv N(0, 9\sigma^2\phi^2\mu/[(2d+1)\Gamma^2(d+1)]) \end{aligned}$$

with

$$\mu = 1 - (4d+2) \left[\int_0^1 \int_0^r (r-t)^d(1-t)^d dt dr - \int_0^1 \int_0^s \int_0^r (r-t)^d(s-t)^d dt dr ds \right]$$

and

$$T^{3-2d}E(\tilde{\beta} - \beta)^2 = \sigma^2\phi^2(3-2d)\Gamma^2(2-d),$$

so the asymptotic relative efficiency of $\hat{\beta}$ to $\tilde{\beta}$ is

$$ARE^t = \frac{9\mu}{(3-2d)(2d+1)\Gamma^2(2-d)\Gamma^2(d+1)}.$$

Figure 5 graphs ARE^t against d over the interval $d \in (-1/2, 1/2)$. As before the efficiency gain of GLS increases with the absolute value of d . However, unlike in figure 2, the efficiency gain does not go to ∞ as $d \rightarrow 0.5$. This is because, when $d = 1/2$, the normalization factor in Lemma 6 does not have the additional part $\sqrt{\log(T)}$ as in Lemma 3. Note that $ARE^t \rightarrow \infty$ as $d \rightarrow -0.5$, which also happens in figure 2. The same explanation applies except we do not need the finite dimension convergence result in Liu(1998) as we can easily derive that $\sum_{t=1}^T u_t = O_p(\sqrt{\log T})$.

Theorem 8 and 9 provide the asymptotics when $d = 1/2$.

Theorem 8 *If $d = 1/2$ and ε_t satisfies the assumptions in Lemma 6, the OLS estimator $\hat{\gamma}$ of γ is distributed asymptotically as*

$$\begin{aligned} F_T(\hat{\gamma} - \gamma) &\Rightarrow \sqrt{\frac{2}{\pi}}\sigma\phi Q^{-1} \left(\mathbf{1}\vec{B}_1(1) - \int_0^1 \bar{g}(r)\vec{B}_1(r)dr \right) \\ &\equiv N(0, 2\sigma^2\phi^2\pi^{-1}Q^{-1}V_{ols}Q^{-1}). \end{aligned}$$

Theorem 9 If $d = 1/2$ and ε_t satisfies the assumptions in Lemma 6, the GLS estimator $\tilde{\gamma}$ of γ is distributed asymptotically as

$$[Diag(\log(T), T, \dots, T^{k-1})]^{1/2}(\tilde{\gamma} - \gamma) \Rightarrow N(0, \sigma^2 \phi^2 \Lambda_{gls}^{-1})$$

where Λ_{gls} is a $k \times k$ matrix with elements $\{\lambda_{ij}\}$,

$$\lambda_{ij} = \begin{cases} (i-1)!(j-1)!/[\Gamma(i-d)\Gamma(j-d)(i+j-2d-1)], & i > 1 \text{ and } j > 1 \\ 1/\Gamma^2(1-d), & i = j = 1 \\ 0, & \text{otherwise} \end{cases}.$$

Note that in Theorem 9, a constant regressor is included and it is indeed consistently estimable by the GLS. When $z_t = 1$ and $d = 1/2$, the GLS estimator of α is the same as the OLS estimator of α in $\tilde{y}_t = \alpha/t^{1/2} + v_t$. It is well known that in the latter regression the convergence rate is of $\sqrt{\log(T)}$. With initialization at the infinite past, neither the OLS nor the GLS estimator of the intercept is consistent. In contrast, with initialization at the origin, the OLS estimator is still inconsistent while the GLS estimator converges by a rate of $\sqrt{\log(T)}$. In effect, comparing with initialization at the origin, initialization at the infinite past strengthens the noise to the extent that even GLS can not consistently detect the signal. Therefore the relative asymptotic efficiency depends crucially on initialization.

Generally, with initialization at the Origin, GLS does not enjoy a faster convergence rate for any value of d . The exceptions are (1) a trend variable is of order $t^{d-1/2}$ (2) $d = -0.5$. In the first case, the GLS estimator of the corresponding coefficient converges faster by a rate of $\sqrt{\log(T)}$ than the OLS estimator while in the second case faster convergence applies to all the coefficients.

We now turn to the case of $d > 1/2$.

Theorem 10 If $d \in (1/2, 3/2)$, $H = d - 1/2$ and ε_t satisfies the assumptions in Lemma 5, the OLS estimator $\hat{\gamma}$ of γ is distributed asymptotically as

$$\begin{aligned} D_T^{1/2}(\hat{\gamma} - \gamma)/T^d &\Rightarrow \frac{\sigma\phi}{\Gamma(d)}(2H)^{-1/2}Q^{-1} \int_0^1 g(r) \vec{B}_H(r) dr \\ &\equiv N(0, \sigma^2 \phi^2 (\Gamma(d))^{-2} Q^{-1} \Xi Q^{-1}) \end{aligned}$$

where $\Xi = \int_0^1 \int_0^1 \int_0^{r \wedge s} g(r)(r-t)^{d-1}(s-t)^{d-1}g(s)' dt ds dr$.

Theorem 11 If $d \in (1/2, 3/2)$ and ε_t satisfies the assumptions in Lemma 5, the GLS estimator $\tilde{\gamma}$ of γ is distributed asymptotically as

$$\begin{aligned} D_T^{1/2}(\tilde{\gamma} - \gamma)/T^d &\Rightarrow \sigma\phi V_{gls}^{-1} \int_0^1 \tilde{g}(r) dW(r) \\ &\equiv N(0, \sigma^2 \phi^2 V_{gls}^{-1}) \end{aligned}$$

If $d > 1/2$ and $z_t = t$, we have

$$T^{(3-2d)/2}(\widehat{\beta} - \beta) \Rightarrow N\left(0, \sigma^2 \phi^2 \frac{9\nu}{[\Gamma(d)]^2(2d-1)}\right)$$

with $\nu = \int_0^1 \int_0^1 \int_0^{r \wedge s} r(r-t)^{d-1}(s-t)^{d-1} s dt ds dr$ and

$$T^{(3-2d)/2}(\widetilde{\beta} - \beta) \Rightarrow N(0, \sigma^2 \phi^2 (3-2d)[\Gamma(2-d)]^2).$$

Therefore

$$ARE = \frac{9\nu}{[\Gamma(d)\Gamma(2-d)]^2(3-2d)}.$$

Figure 6 graphs the relative efficiency of $\widehat{\beta}$ for $z_t = t$ over $d \in (1/2, 3/2)$. It shows that $ARE \rightarrow \infty$ as $d \rightarrow 3/2$, This is in sharp contrast with the case of infinite past initialization where $ARE \rightarrow 1$ as $d \rightarrow 3/2$. Actually, when $z_t = t$ and $d = 3/2$, the OLS is inconsistent while the GLS converges by a rate of $\sqrt{\log(T)}$, the same as when $z_t = 1$ and $d = 1/2$.

5 CONCLUSION

When the errors are fractionally integrated, the asymptotic equivalence of GLS detrending and OLS detrending no longer holds. In this case, the asymptotic relative efficiency depends initialization. Under infinite past initialization, the GLS estimator, when consistent, is not only asymptotically more efficient for any value of d but also converges faster by a rate of $\sqrt{\log(T)}$ when the errors are integrated of order $k - 1/2$, $k = 0, 1, 2, \dots$. Under the origin initialization, GLS still achieves substantial efficiency gain, but it, in general, does not enjoy a faster convergence rate. The exceptions are (1) a trending variable is of order $t^{d-1/2}$ and GLS estimator of this trend converges faster than OLS by a rate of $\sqrt{\log(T)}$. (2) $d = -0.5$ and GLS estimators of all coefficients converge faster the OLS by a rate of $\sqrt{\log(T)}$.

In this paper we consider only two alternative initializations, a general initialization would be letting $v_t = 0$ for all $t < -[T\kappa]$ for some $\kappa \in (0, 1)$, *i.e.*, the process is initialized at the past that comparable to sample size. This generalization is straightforward. We only need to redefine a fractional Brownian motion as

$$\vec{B}_H(r) = (2H)^{1/2} \int_{-\kappa}^r (r-s)^{H-1/2} dW(s), \quad H \in (0, 1]$$

and all the asymptotics in section 4 apply. In addition, we assume the long memory parameter is known, more effort is needed to carry out the asymptotics for feasible GLS and compare the efficiency of OLS to feasible GLS.

6 APPENDIX

6.1 Lemma

$$(1 - L)^{-d}v_t = \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j v_t \text{ for any } d.$$

The proof is based on the representation for $-1/2 \leq d < 1/2$. we only prove relevant the case $1/2 \leq d < 3/2$. Other cases follow similarly. Since

$$\begin{aligned} (1 - L)^{-d}v_t &= \sum_{k=1}^t (1 - L)^{-(d-1)}v_k \\ &= \sum_{k=1}^t \sum_{j=0}^{k-1} \frac{\Gamma(j+d-1)}{\Gamma(d-1)\Gamma(j+1)} v_{k-j} \\ &= \sum_{m=1}^t \sum_{k=m}^t \frac{\Gamma(k+d-1-m)}{\Gamma(d-1)\Gamma(k-m+1)} v_m \\ &= \sum_{m=1}^t \sum_{k=1}^{t-m+1} \frac{\Gamma(k+d-2)}{\Gamma(d-1)\Gamma(k)} v_m \\ &= \sum_{m=1}^t \frac{\Gamma(t-m+d)}{\Gamma(d)\Gamma(t-m+1)} v_m \\ &= \sum_{j=0}^{t-1} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j v_t \end{aligned}$$

as required. Here we have used $\sum_{k=1}^{t-m+1} \frac{\Gamma(k+d-2)}{\Gamma(k)} = \frac{\Gamma(t-m+d)}{(d-1)\Gamma(t-m+1)}$ which can be proved by induction on m .

6.2 Proof of Theorem 1:

Let $S_t = \sum_1^t u_j$ and set $S_0 = 0$. Let i be any integer in the interval $[1, k-1]$. Rewrite $\sum_1^T t^i u_t$ as

$$\begin{aligned} \sum_1^T t^i u_t &= \sum_1^T t^i (S_t - S_{t-1}) \\ &= \sum_1^{T-1} S_t (t^i - (t+1)^i) + S_T T^i \\ &= S_T T^i - i \sum_1^{T-1} t^{i-1} S_t + o_p(T^{(1+2d+2i)/2}). \end{aligned}$$

Therefore

$$\begin{aligned} T^{-(1+2d+2i)/2} \Sigma_1^T t^i u_t &= T^{-(1+2d)/2} S_T - iT^{-1} \Sigma_1^{T-1} T^{-(1+2d)/2} r^{i-1} S_t + o_p(1) \\ &\Rightarrow \sigma \phi K(d) (B_H(1) - i \int_0^1 r^{i-1} B_H(r) dr) \end{aligned}$$

where $H = d+1/2$ using Lemma 1 and the continuous mapping theorem. Hence,

$$\begin{aligned} D_T^{1/2} (\hat{\gamma} - \gamma) / T^d &= (D_T^{-1/2} \Sigma_1^T z_t z_t' D_T^{-1/2})^{-1} D_T^{-1/2} \Sigma_1^T z_t u_t / T^d \\ &\Rightarrow \sigma \phi Q^{-1} K(d) (\mathbf{1} B_H(1) - \int_0^1 \bar{g}(r) B_H(r) dr). \end{aligned}$$

When $d \neq 0$, we can show that, after some manipulation,

$$\begin{aligned} &E[(B_H(1) - i \int_0^1 r^{i-1} B_H(r) dr)(B_H(1) - j \int_0^1 s^{j-1} B_H(s) ds)] \\ &= d(1+2d) \int_0^1 \int_0^1 r^i s^j |r-s|^{2d-1} dr ds \\ &= \varpi_{i+1, j+1}^{ols}, \quad i \geq 1 \text{ and } j \geq 1 \end{aligned}$$

and

$$\begin{aligned} &E[(B_H(1)(B_H(1) - j \int_0^1 s^{j-1} B_H(s) ds)] \\ &= d(1+2d) \int_0^1 \int_0^1 s^j |r-s|^{2d-1} dr ds \\ &= \omega_{1, j+1}^{ols}, \quad j \geq 1, \end{aligned}$$

so that

$$D_T^{1/2} (\hat{\gamma} - \gamma) / T^d \Rightarrow N(0, \sigma^2 \phi^2 K^2 Q^{-1} \Omega_{ols} Q^{-1'}) \text{ for } d \neq 0$$

When $d = 0$,

$$\begin{aligned} D_T^{1/2} (\hat{\gamma} - \gamma) / T^d &\Rightarrow \sigma \phi Q^{-1} K(d) \int_0^1 g(r) dW(r) \\ &\equiv N(0, \sigma^2 \phi^2 K^2 Q^{-1} \Omega_{ols} Q^{-1'}) \end{aligned}$$

with $\Omega_{ols} = Q$.

6.3 Proof of Theorem 3:

Let $S_t = \Sigma_1^t u_j$ and set $S_0 = 0$. Let i be any integer in the interval $[1, k-1]$. As in 6.1 rewrite $\Sigma_1^T t^i u_t$ as

$$\begin{aligned} \Sigma_1^T t^i u_t &= \Sigma_1^T t^i (S_t - S_{t-1}) \\ &= \Sigma_1^{T-1} S_t (t^i - (t+1)^i) + S_T T^i \\ &= S_T T^i - i \Sigma_1^{T-1} t^{i-1} S_t + o_p(T^{(1+i)} \sqrt{\log(T)}). \end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\sum_1^T t^i u_t}{T^{(1+i)}\sqrt{\log(T)}} &= \frac{S_T}{T\sqrt{\log(T)}} - i\frac{1}{T}\sum_1^{T-1}\left(\frac{t}{T}\right)^{i-1}\frac{S_t}{T\sqrt{\log(T)}} + o_p(1) \\
&\Rightarrow \sigma\phi/\sqrt{\pi}W(1)(1-i\int_0^1 r^i dr) \\
&= \sigma\phi/\sqrt{\pi}W(1)/(i+1)
\end{aligned}$$

using Lemma 2 and the continuous mapping theorem. Together with Lemma 3, we have

$$\begin{aligned}
F_T(\hat{\gamma} - \gamma)/\sqrt{\log(T)} &= (T^{-1}F_T^{-1}\Sigma_1^T z_t z_t' F_T^{-1})^{-1} F_T^{-1}\Sigma_1^T z_t u_t \left(T\sqrt{\log(T)}\right)^{-1} \\
&\Rightarrow N(0, \sigma^2\phi^2\pi^{-1}Q^{-1}\Omega_{ols}Q^{-1'})
\end{aligned}$$

6.4 Proof of Theorem 4:

Write $\hat{\gamma} - \gamma = (\Sigma_1^T z_t z_t')^{-1}\Sigma_1^T z_t u_t$. Now

$$\begin{aligned}
\Sigma_1^T z_t u_t &= \Sigma_1^T T \int_{(t-1)/T}^{t/T} z_{[Ts]} u_{[Ts]} ds, \text{ where } \frac{t-1}{T} \leq s < \frac{t}{T} \\
&= D_T^{1/2} T^d \Sigma_1^T \int_{(t-1)/T}^{t/T} g(s) T^{(1-2d)/2} u_{[Ts]} ds \\
&= D_T^{1/2} T^d \int_0^1 g(s) T^{(1-2d)/2} u_{[Ts]} ds,
\end{aligned}$$

hence, using Lemma 2 and the continuous mapping theorem we have

$$D_T^{-1/2}\Sigma_1^T z_t u_t / T^d \Rightarrow \sigma\phi K(d-1) \int_0^1 g(s) B_H(s) ds \text{ where } H = d - 1/2, d \in (1/2, 3/2),$$

and

$$\begin{aligned}
D_T^{1/2}(\hat{\gamma} - \gamma)/T^d &= (D_T^{-1/2}\Sigma_1^T z_t z_t' D_T^{-1/2})^{-1} D_T^{-1/2}\Sigma_1^T z_t u_t / T^d \\
&\Rightarrow \sigma\phi K(d-1) Q^{-1} \int_0^1 g(s) B_H(s) ds.
\end{aligned}$$

Since $E[\int_0^1 g(s) B_H(s) ds] = 0$, and

$$\begin{aligned}
&E\left[\int_0^1 g(s) B_H(s) ds\right] \left[\int_0^1 g(r) B_H(r) dr\right]' \\
&= 1/2 \int_0^1 \int_0^1 g(r) (r^{2d-1} - |r-s|^{2d-1} + s^{2d-1}) g(s)' ds dr = \Sigma,
\end{aligned}$$

we deduce that $D_T^{1/2}(\hat{\gamma} - \gamma)/T^d \Rightarrow N(0, \sigma^2\phi^2 K^2(d-1)Q^{-1}\Sigma Q^{-1'})$.

6.5 Proof of Theorem 5:

The GLS estimator of γ in $y_t = \gamma' z_t + u_t$ is the same as the GLS estimator of γ in $(1-L)y_t = (1-L)z_t\gamma + (1-L)u_t = (1-L)(t, \dots, t^{k-1})\beta$ as long as we assume all nonstationary processes take value zero at $t = 0$. Note that the latter regression satisfies the assumption in Theorem 2 with $(1-L)u_t \sim I(d-1)$ and $(1-L)t^k = kt^{k-1} + o(t^{k-1})$, the theorem follows immediately.

6.6 Proof of Theorem 6 and 8:

Follow the same procedure as in the proof of theorem 1.

To establish theorem 7, we need the following lemma.

6.7 Lemma:

For any finite integer $n \geq 0$ and $d \geq -1/2$,

$$(1-L)^d t^n \sim [n!/\Gamma(n+1-d)]t^{n-d} \text{ as } t \rightarrow \infty .$$

Remark (1): Actually Lemma 6.5 holds for non integer indices as follows

$$(1-L)^d t^\alpha = (\Gamma(\alpha+1)/\Gamma(\alpha+1-d))t^{\alpha-d} + o(t^{\alpha-d}) .$$

Remark (2): Note that with this rule we can evaluate the composite operator

$$\begin{aligned} (1-L)^{1-d}(1-L)^d t^n &\sim (n!/\Gamma(n+1-d))(1-L)^{1-d} t^{n-d} \\ &\sim (n!/\Gamma(n+1-d))\Gamma(n+1-d)/\Gamma(n+1-d-(1-d))t^{n-d-(1-d)} \\ &= (n!/\Gamma(n))t^{n-1} = nt^{n-1}, \end{aligned}$$

corresponding to the simple differencing operator $(1-L)t^n = t^n - (t-1)^n = nt^{n-1} + o(t^{n-1})$, as expected.

To prove Lemma 6.7, we need the following elementary combinatorial result for which we could find no reference. In the proof of this result we relate the fractional operator $(1-L)^d$ to fractional differentiation to facilitate the argument.

6.8 Lemma:

$$\sum_{j=0}^n \binom{n}{j} (-1)^{n-j} (\gamma + n - j)^{-1} = n! \prod_{j=0}^n (\gamma + j)^{-1}.$$

Proof of Lemma 6.8: Proceed by induction and assume the result holds for $n-1$, i.e.

$$\sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^{n-j-1} 1/(\gamma + n - j - 1) = (n-1)! \prod_{j=0}^{n-1} (\gamma + j)^{-1}. \quad (\text{A1})$$

Multiple (A1) by $1/(\gamma+n)$ and the left side is

$$\begin{aligned} & \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^{n-j-1} [(\gamma+n-j-1)(\gamma+n)]^{-1} \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^{n-j-1} (1/(j+1)) [\gamma+n-j-1]^{-1} - (\gamma+n)^{-1}. \end{aligned} \quad (\text{A2})$$

Note that

$$(j+1)^{-1} \binom{n-1}{j} = \frac{(n-1)!}{(j+1)!(n-j-1)} = n^{-1} \binom{n}{k} \text{ with } k = j+1$$

so that (A2) becomes

$$n^{-1} \sum_{k=1}^n \binom{n}{k} (-1)^{n-k} [(\gamma+n-k) - (\gamma+n)^{-1}]. \quad (\text{A3})$$

But

$$\sum_{k=1}^n \binom{n}{k} (-1)^{n-k} + \binom{n}{0} (-1)^n = (1-1)^n = 0,$$

so

$$\sum_{k=1}^n \binom{n}{k} (-1)^{n-k} = -(-1)^n = (-1)^{n-1}.$$

Hence (A3) is

$$\begin{aligned} & n^{-1} \sum_{k=1}^n \binom{n}{k} (-1)^{n-k} (\gamma+n-k)^{-1} - n^{-1} (-1)^{n-1} / (\gamma+n) \\ &= n^{-1} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} (\gamma+n-k)^{-1}. \end{aligned}$$

It follows that (A2) is

$$n^{-1} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} (\gamma+n-k)^{-1} = (n-1)! \prod_{j=0}^n (\gamma+j)^{-1},$$

as required for the induction from $n-1$ to n .

When $n=1$, we have for the left side of the equation (A1),

$$((\gamma+1)^{-1}(-1) + 1/\gamma) = 1/[\gamma(\gamma+1)],$$

showing that the result holds for $n=1$.

6.9 Proof of Lemma 6.7:

For large t we can write the operation $(1 - L)^d t^n$ in terms of a fractional derivative as follows. We set $\alpha = m - d$, where m is the smallest integer greater than d . Then we let $D = d/dt$ be the differential operator, and define the fractional power $D^{m-\alpha}$ of D as the product of D^m and $D^{-\alpha}$. The latter can be simply defined as a fractional integral using the Gamma function (see, e.g. Ross (1974, p. 16)). We then have:

$$\begin{aligned}
(1 - L)^d t^n &= D^{m-\alpha} t^n + o(t^{n-d}) \\
&= D^m D^{-\alpha} t^n + o(t^{n-d}) = D^m [(\Gamma(\alpha))^{-1} \int_0^t e^{-Ds} s^{\alpha-1} ds] t^n + o(t^{n-d}) \\
&= D^m [(\Gamma(\alpha))^{-1} \int_0^t (t-s)^n s^{\alpha-1} ds] + o(t^{n-d}) \\
&= n! [(n-m)! \Gamma(\alpha)]^{-1} \int_0^t (t-s)^{n-m} s^{\alpha-1} ds + o(t^{n-d}) \\
&= n! [(n-m)! \Gamma(\alpha)]^{-1} \int_0^t \sum_{j=0}^{n-m} \binom{n-m}{j} t^j (-s)^{n-m-j} s^{\alpha-1} ds + o(t^{n-d}) \\
&= n! [(n-m)! \Gamma(\alpha)]^{-1} \sum_{j=0}^{n-m} \binom{n-m}{j} t^j (-1)^{n-m-j} \int_0^t s^{n-m-j+\alpha-1} ds + o(t^{n-d}) \\
&= n! [(n-m)! \Gamma(\alpha)]^{-1} \sum_{j=0}^{n-m} \binom{n-m}{j} t^j (-1)^{n-m-j} t^{n-m-j+\alpha} / (\alpha-j+n-m) + o(t^{n-d}) \\
&= n! [(n-m)! \Gamma(\alpha)]^{-1} t^{\alpha+n-m} \sum_{j=0}^{n-m} \binom{n-m}{j} (-1)^{n-m-j} / (\alpha-j+n-m) + o(t^{n-d}) \\
&= n! (\Gamma(\alpha))^{-1} t^{\alpha+n-1} \prod_{j=0}^{n-m} (\alpha+j)^{-1} + o(t^{n-d}) \\
&= n! (\Gamma(\alpha+n-m+1))^{-1} t^{\alpha+n-1} + o(t^{n-d}) = n! (\Gamma(n+1-d))^{-1} t^{n-d} + o(t^{n-d}),
\end{aligned}$$

as required.

6.10 Proof of Theorem 7:

Using Lemma 6.7 we proceed as follows: The (i, j) th element of $\Sigma_1^T \tilde{z}_t \tilde{z}'_t$,

$$\begin{aligned}
[\Sigma_1^T \tilde{z}_t \tilde{z}'_t]_{ij} &\sim \Sigma_1^T \frac{(i-1)!(j-1)! t^{i+j-2d-2}}{\Gamma(i-d)\Gamma(j-d)} \text{ for } t \text{ large} \\
&\sim \frac{(i-1)!(j-1)! T^{i+j-2d-1}}{\Gamma(i-d)\Gamma(j-d)(i+j-2d-1)}.
\end{aligned}$$

Then $\lim_{T \rightarrow \infty} (D_T^{1/2} \Sigma_1^T \tilde{z}_t \tilde{z}_t' D_T^{1/2} T^{2d})^{-1} \rightarrow V_{gls}$. Also the i th row of $\sum_{t=1}^T \tilde{z}_t v_t$,

$$\begin{aligned} \left[\sum_{t=1}^T \tilde{z}_t v_t \right]_i &\sim \sum_{t=2}^T (i-1)! / \Gamma(i-d) t^{i-1-d} v_t \\ &= T^{1/2} (\Gamma(i-d))^{-1} (i-1)! \Sigma_2^T t^{i-1-d} \int_{t/T}^{(t+1)/T} dX_T(r) \\ &= T^{1/2} T^{i-1-d} (\Gamma(i-d))^{-1} (i-1)! \Sigma_2^T \int_{t/T}^{(t+1)/T} r^{i-1-d} dX_T(r), \end{aligned}$$

where $X_T(r) = T^{-1/2} \sum_{i=1}^{\lfloor Tr \rfloor} v_i$.

To derive the asymptotic distribution when $i \geq 2$, we can integrate the above integral by part and then apply continuous mapping theorem. However when $i = 0, 1$, the CMT is not valid and we have to employ Akonom and Gourieroux (1987) type argument, i.e. to prove convergence of finite dimension distributions and tightness. In any case, we have:

$$T^{-(-1+2i-2d)/2} [\Sigma_1^T \tilde{z}_t v_t]_i \Rightarrow \sigma \phi(i-1)! (\Gamma(i-d))^{-1} \int_0^1 r^{i-1-d} dW(r),$$

and

$$\begin{aligned} D_T^{1/2} (\tilde{\gamma} - \gamma) / T^d &= (D_T^{-1/2} \Sigma_1^T \tilde{z}_t \tilde{z}_t' D_T^{-1/2} T^{2d})^{-1} D_T^{-1/2} T^d \Sigma_1^T \tilde{z}_t v_t \\ &\Rightarrow \sigma \phi V_{gls}^{-1} \int_0^1 \tilde{g}(r) dW(r) \\ &\equiv N(0, \sigma^2 \phi^2 V_{gls}^{-1}), \end{aligned}$$

since $E[\int_0^1 \tilde{g}(r) \tilde{g}(r)' dr] = V_{gls}$, giving the stated result.

6.11 Proof of Theorem 9:

As in 6.10, the (i, j) th element of $\Sigma_1^T \tilde{z}_t \tilde{z}_t'$ when $i > 1$ and $j > 1$ satisfies

$$[\Sigma_1^T \tilde{z}_t \tilde{z}_t']_{ij} / T^{i+j-2d-1} \sim \frac{(i-1)!(j-1)!}{\Gamma(i-d)\Gamma(j-d)(i+j-2d-1)}.$$

For the $(1, 1)$ th element, we have

$$\begin{aligned} [\Sigma_1^T \tilde{z}_t \tilde{z}_t']_{11} / \log(T) &\sim \Sigma_1^T \frac{t^{-1}}{\Gamma(1-d)\Gamma(1-d)\log(T)} \text{ for } t \text{ large} \\ &\sim \frac{1}{\Gamma^2(1-d)}. \end{aligned}$$

For the (i, j) th element of $\Sigma_1^T \tilde{z}_t \tilde{z}_t'$ when either $i = 1$ or $j = 1$,

$$[\Sigma_1^T \tilde{z}_t \tilde{z}_t']_{ij} / (T^{i+j-2d-1} \log T) \sim 0$$

Also the i th row of $\sum_{t=1}^T \tilde{z}_t \varepsilon_t$ with $i > 1$ satisfies

$$T^{-(-1+2i-2d)/2} [\sum_1^T \tilde{z}_t v_t]_i \Rightarrow \sigma(i-1)! (\Gamma(i-d))^{-1} \int_0^1 r^{i-1-d} dW(r),$$

and the first row of the $\sum_{t=1}^T \tilde{z}_t v_t$

$$[\sum_{t=1}^T \tilde{z}_t v_t]_1 / \sqrt{\log(T)} \sim \sum_{t=1}^T t^{-1/2} v_t / [\Gamma(1-d) \sqrt{\log(T)}] \Rightarrow N(0, \frac{\sigma^2 \phi^2}{[\Gamma(1-d)]^2})$$

Combine the above results, the theorem follows immediately.

6.12 Proof of Theorem 11:

Follow the same procedure as in 6.10.

Fig1: Relative Efficiency with the Infinite Past Initialization
When $-0.5 < d < 0.5$ and $Z_t = 1$

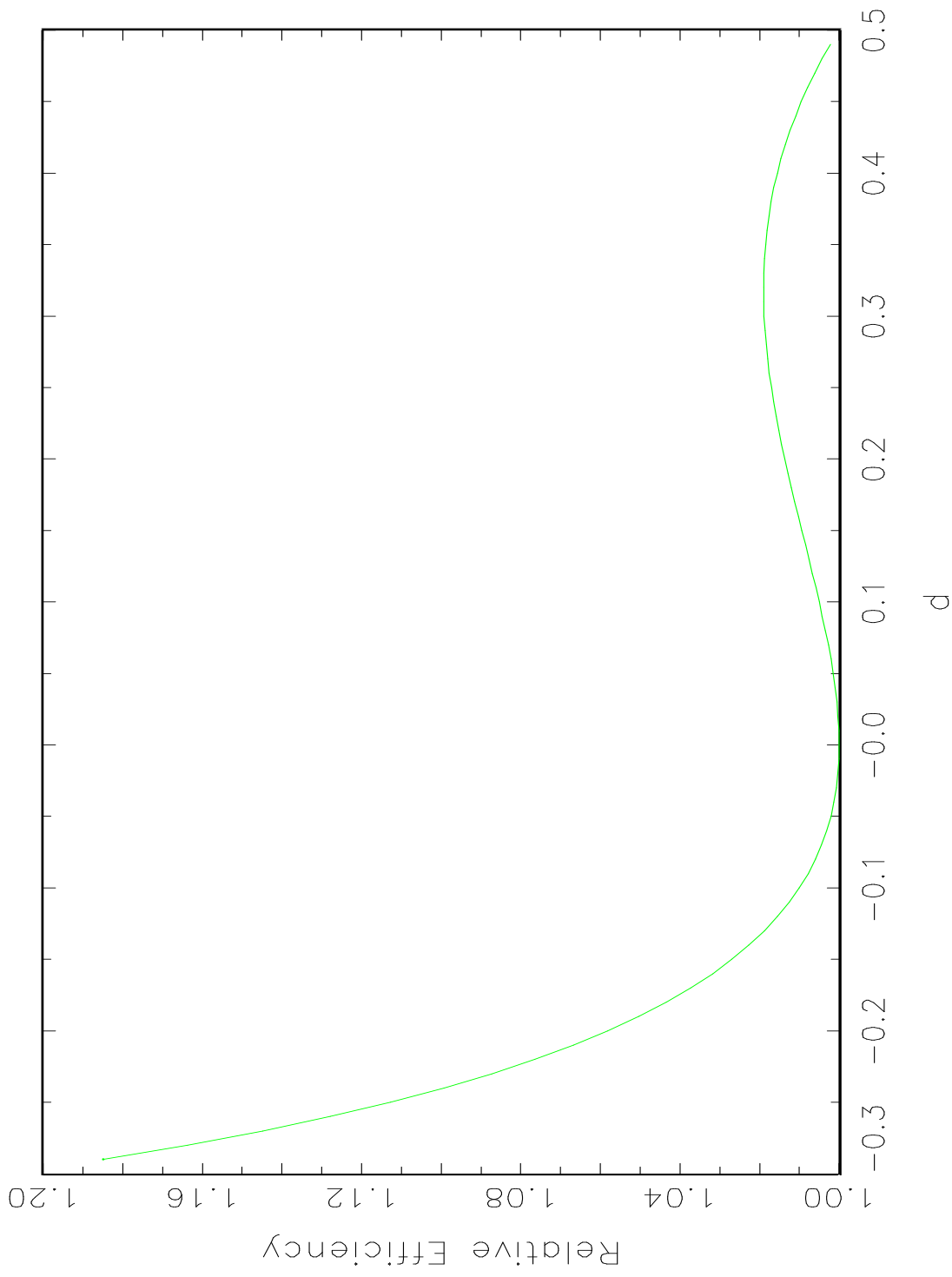


Fig2: Relative Efficiency with the Infinite Past Initialization
When $-0.5 < d < 0.5$ and $Z_t = t$

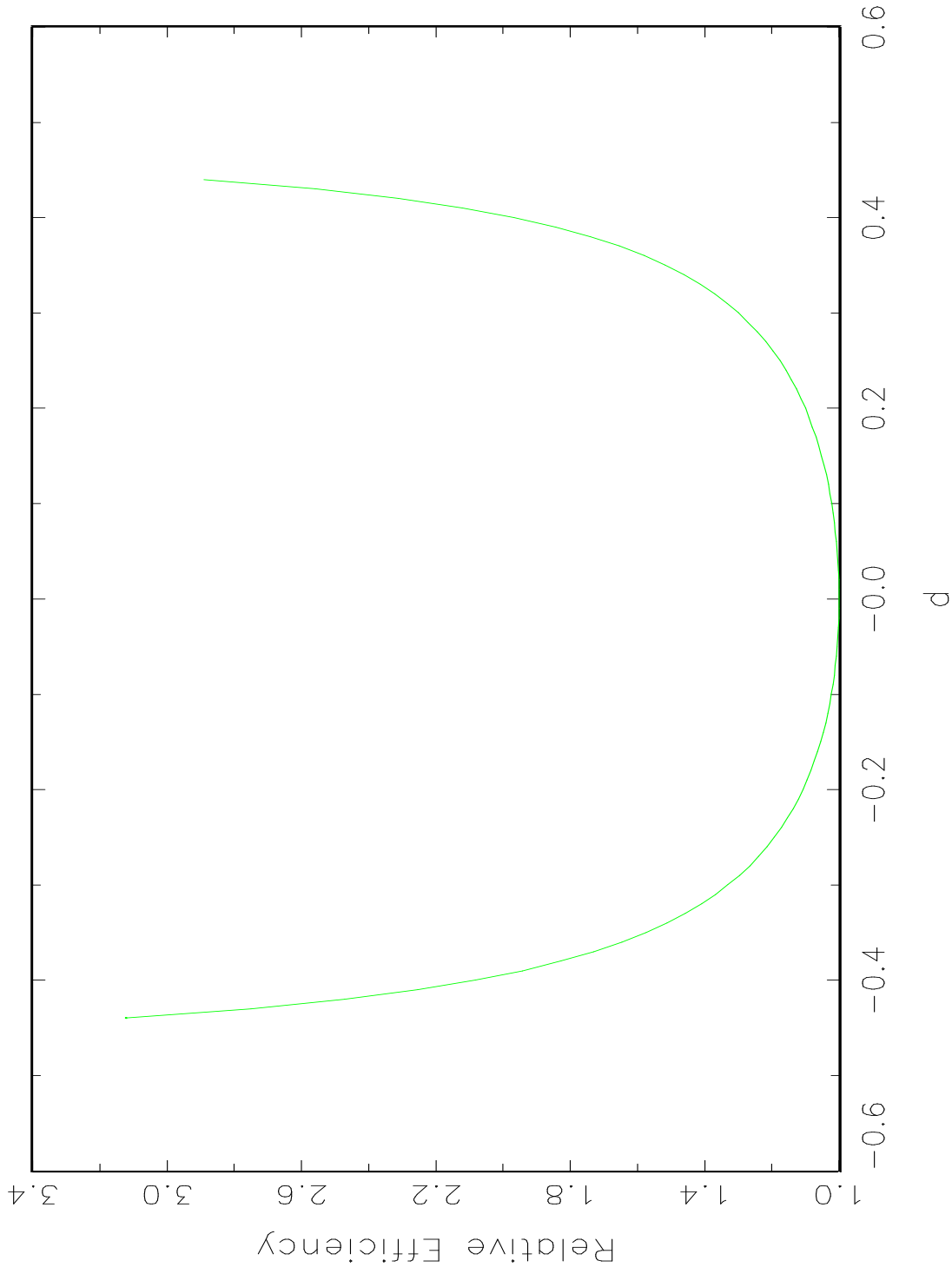


Fig3: Relative Efficiency with the Infinite Past Initialization
When $0.5 < d < 1.5$ and $Z_t = t$

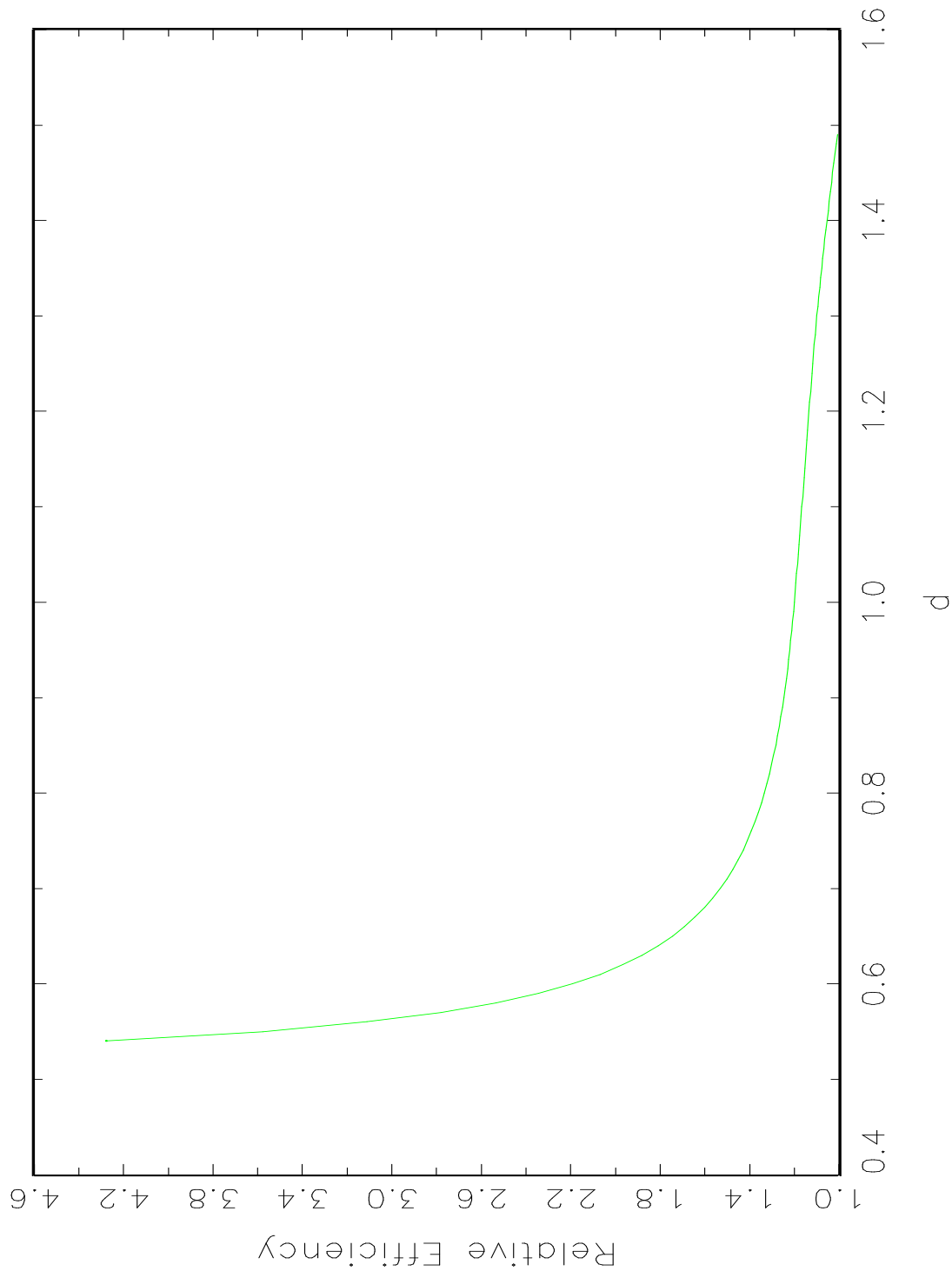


Fig4: Relative Efficiency With Initializations at the Origin
When $-0.5 < d < 0.5$ and $Zt=1$

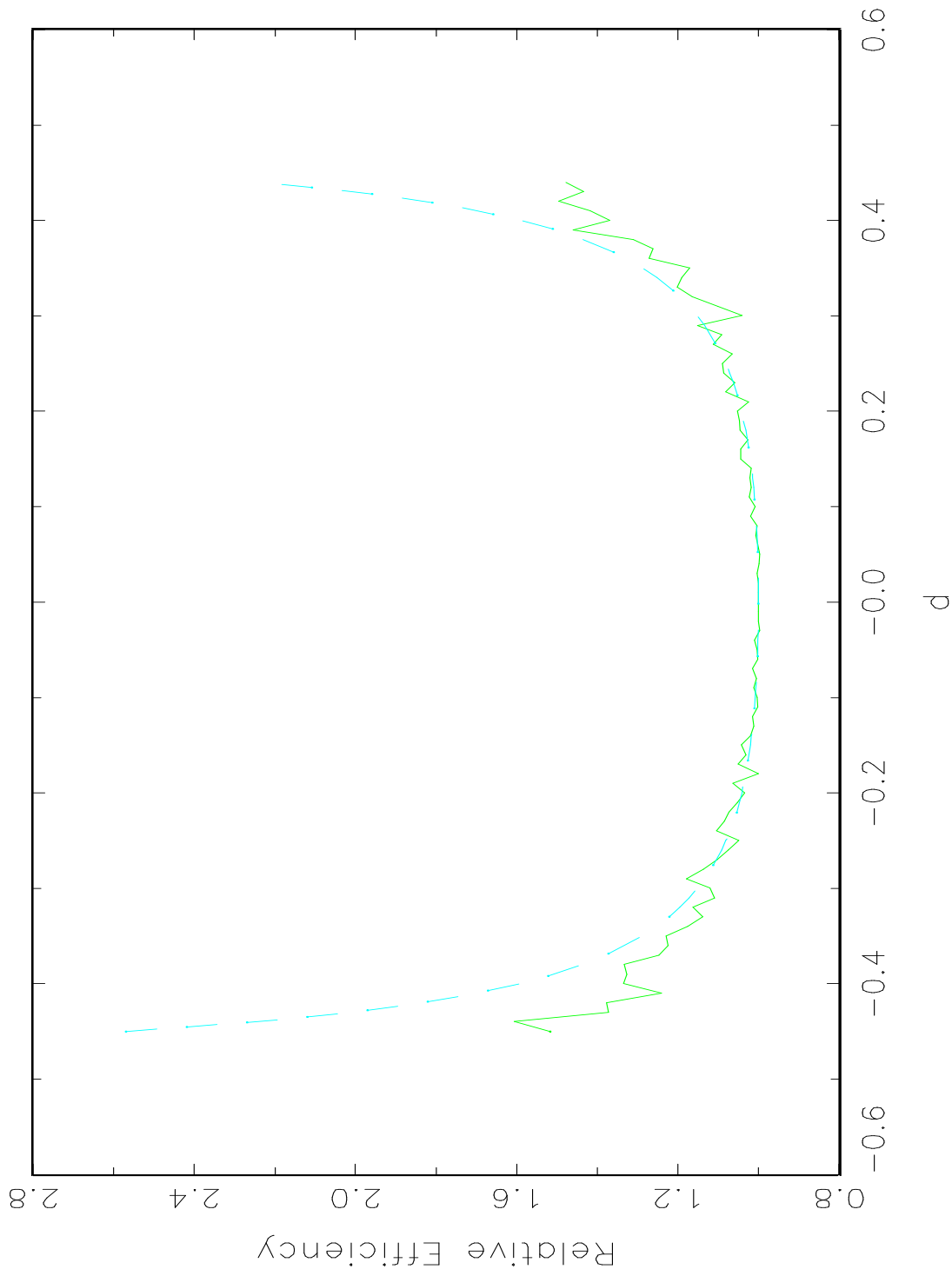


Fig5: Relative Efficiency With Initialization at Origin
When $-0.5 < d < 0.5$ and $Zt=t$

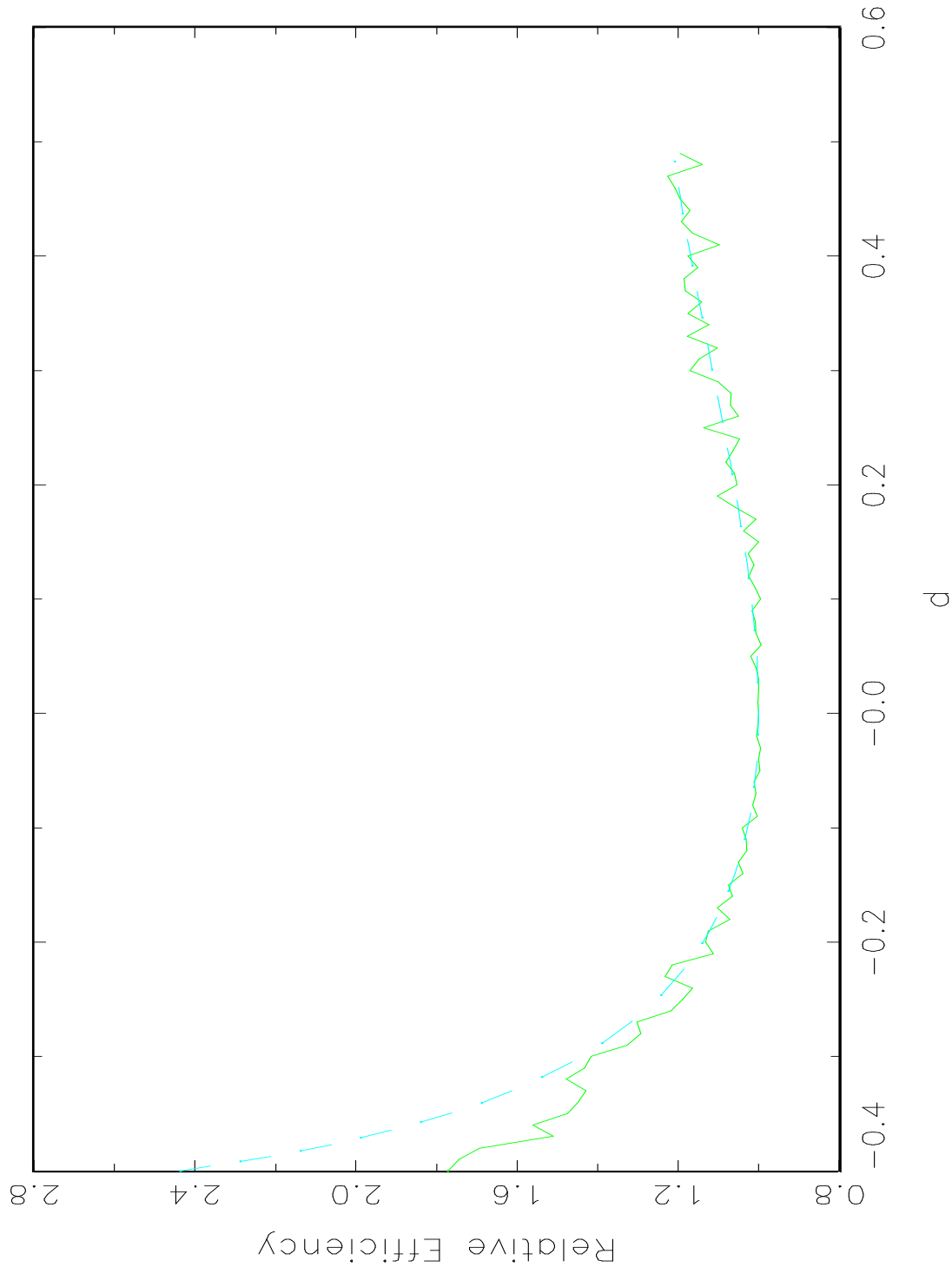
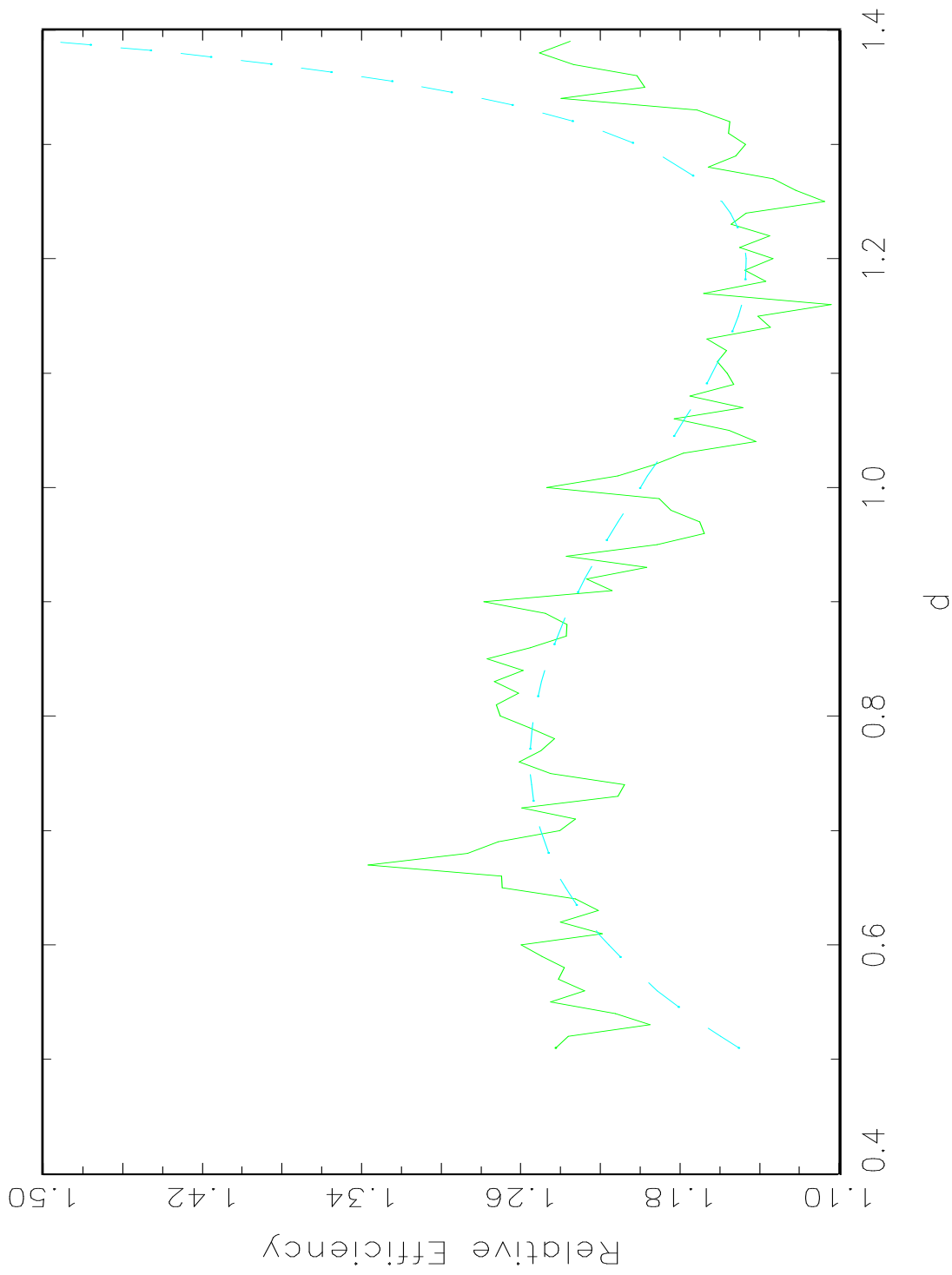


Fig6: Relative Efficiency With Initialization at Origin
When $0.5 < d < 1.5$ and $Zt=t$



7 REFERENCES

- Akonom, J. and C. Gouriéroux (1987). “A functional limit theorem for fractional processes,” unpublished manuscript.
- Baillie, R. (1996). “Long memory processes and fractional integration in econometrics.” *Journal of Econometrics*, 73, 5–59.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall.
- Bollerslev, T. and H.O. Mikkelsen (1993). “Modeling and pricing long memory in stock market volatility.” *Journal of Econometrics* 73(1), 151-184.
- Granger, C. W. J. and R. Joyeux (1980). “An introduction to long-memory time series models and fractional differencing,” *Journal of Time Series Analysis*, 1, 15–29.
- Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- Hassler, U. and J. Wolters (1995). “Long memory in inflation rates: international evidence.” *Journal of Business and Economics Statistics* 13(1), 37-45.
- Liu, M. (1998). “Asymptotics of nonsationary fractional integrated series.” *Econometric theory*, 14, 641-662.
- Mandelbrot, B. and J. Van Ness (1968). “Fractional Brownian motions, fractional Brownian noises and applications,” *SIAM Review*, 10, 422–437.
- Phillips, P.C.B. and C.C. Lee (1996). “Efficiency Gains from Quasi-Differencing under Nonstationarity.” in P. M. Robinson and M. Rosenblatt (Eds.) *Athens Conference on Applied Probability and Time Series: Volume II Time Series Analysis in memory of E. J. Hannan*.
- Phillips, P.C.B. and V. Solo (1994) “Asymptotics for linear processes”, *Annals of Statistics*, Vol. 20, No. 2, June 1992, pp. 971-1001.
- Samarov, A. and M. S. Taqqu (1988). “On the efficiency of the sample mean in long memory noise,” *Journal of Time Series Analysis*, 9, 191–200.
- Samorodnitsky, G. and M. S. Taqqu (1994). *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. London: Chapman and Hall.
- Sowell, F. (1990). “The fractional unit root distribution,” *Econometrica*, 58, 495–505.
- Taqqu, M. (1975). “Weak convergence to fractional Brownian motion and to the Rosenblatt process,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31, 287–302.
- Yajima, Y. (1988). “On estimation of a regression model with long-memory stationary errors,” *Annals of Statistics*, 16, 791–807.