Estimation of Linear Model with One Time-varying Parameter via Wavelets

Michael C. Tseng, Ramazan Gençay

August 2015

Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada.

Abstract

We consider a linear model with one time-varying parameter, under standard strict stationarity assumptions on the regressors and error terms. The error terms can have short-range serial correlation. No restrictions are placed on the types of time-varying behavior, allowing for arbitrary combinations of smooth structural changes and abrupt structural breaks. The ability of wavelets to represent parsimoniously such spatially inhomogeneous functions yield a minimax estimator, which is a non-linear orthogonal series estimator. Building on previous results, it is shown that the wavelet estimate of the time varying parameter is without spurious jumps with probability one asymptotically, thereby faithfully captures the structure breaks in the parameter.

JEL classification: C14; C32.

Keywords: Time varying parameter. Smooth time-varying trend. Abrupt breaks. Jumps. Minimax estimator. Orthogonal series estimator. Wavelets.

* Corresponding author: Tel.: +1-778-782-3934. *Email addresses:* rgencay@sfu.ca (Ramazan Gençay).

1 Introduction

Instability of parameters is an issue confronted by econometricians everyday. Coefficient instability reflects, for example, an economic system exposed to exogenous shocks or an economic agent exhibiting behavioral changes (the latter as pointed out by the Lucas critique Lucas Jr (1976)). Indeed, the empirical literature has found models with constant parameters to be unsatisfactory in various contexts and devised various models to accommodate such structural changes in the underlying data generating process.

To the best of our knowledge, theoretical literature on linear models with time-varying coefficients predominantly employs kernel smoothing methods. See, for example, Cai (2007), Cai *et al.* (2009), Kristensen (2012), and Chen and Hong (2012). The kernel approach requires the model assumption that the functional parameters be at least as smooth as the kernel, thus ruling out abrupt structural breaks or jump behavior. While our model only admits a single time varying parameter as a source of non-stationarity, the standard smoothness assumption on the time-varying parameter is dropped completely. Formally, this means that the functional space containing the functional parameter is enlarged significantly. When the time-varying coefficient corresponds to an intercept term, our model is a linear model with arbitrary time trend and overlaps with the partial linear model from Robinson (1988) in the time series setting.

Our approach consists of two steps. The time-invariant part of the model is estimated consistently by kernel method. With these consistent estimates in hand, the problem of estimating the time-varying parameter is converted into a de-noising problem. One has a time series of noisy observations of the time-varying parameter. The objective is to remove the noise from the time series to reveal the true underlying trend. In our case, this trend may be an arbitrary function of time that is, for example, nonlinear and has jumps. We apply an orthogonal series estimator to this problem. In particular, we choose the orthonormal basis to be a wavelet basis and make use of their descriptive power in encoding spatial-inhomogeneity in the parameter, such as varying degree of smoothness between possible discontinuities.

Wavelets was first discovered by Daubechies in seeking an orthonormal basis for square-integrable functions that is both compactly supported and has certain smoothness properties Daubechies (1988). The multi-resolution structure inherent in their construction makes the associated filters natural tool in time series analysis from a spectral perspective. A sizable body of results now testifies to the effective of wavelets in this regard. For example, the unit root test of Fan and Gençay (2010) and serial correlation test of Gençay and Signori (2015) originate from this perspective. In contrast, the ability of wavelets to represent parsimoniously a wide variety of functions remain relatively unexplored in the econometrics. Lee and Hong (2001) introduces a test of serial correlation for covariance stationary time series based on a linear wavelet estimator of the spectral density. In principle, this nonparametric approach via wavelets allows for testing for serial correlation of arbitrary form, without smoothness assumptions on the spectral density, resulting in a powerful test. This was extended to the panel model setting by Hong and Kao (2004).

In our methodology, following Donoho (1995a), we make use of a nonlinear wavelet estimator that is both minimax over a wide Besov scale and preserve the smoothness properties of the underlying true parameter with probability one asymptotically. The language of Besov spaces allows one to speak precisely of smooth properties more general than differentiability, or, indeed, continuity (Besov *et al.* (1978)), and this language translates the ability of the wavelet estimator to faithfully extract true jump from a noisy signal containing many spurious jumps. In addition to satisfactory theoretical properties, the nonlinear wavelet estimator also offers computational advantages due to the pyramid structure of the associated filter algorithm.

We also exploit the whitening effect of wavelets on serially correlated processes when the error terms in the model have short memory instead of being mere white noise. In this more general setting, the utility of wavelets is two-fold: first to decorrelate serial correlation in error terms, second to encode efficiently the underling time-varying coefficient. Both of these properties derive from the time-frequency localization of wavelets, which is unique among available basis. Bases such as the classical Fourier basis or B-splines do not have this property.

There are also extensive literature on testing for structural breaks. The flexibility of wavelet estimator, being minimax in a global sense rather than pointwise, can complement existing testing methods by supplying required information within the class of models being considered. The seminal paper Andrews (1993) introduced tests for structural break in a general partial sample general method of moments framework. The Andrews test is formally designed against alternatives where the structural change occurs in a specified interval. However, information about location of change point for the parameter in question may not be available to the econonometrican. For example, the issue at hand could be general model adequacy. Or the structural break may be caused by a policy change, the lag-effect of which is is unknown. For the class of model considered in this paper, the Andrews alternative is the special case where the time-varying coefficient is a piecewise constant function of time. A minimax estimator that captures the true jumps can be very useful in providing the econonometrican with a possible locations of structure breaks in the Andrews alternative. When the time-varying coefficient is piecewise constant, our model is also a special case of the linear models with partial structural breaks considered in Bai and Perron (1998) and Perron and Yamamoto (2014) where the number of breaks are assumed to be known. Bai and Perron (1998) also provides a test for whether the model has l versus $l+1$ structural breaks. Once the number of breaks are determined, estimating the location of the breaks then becomes a dynamic programming problem over the set of possible partitions. In procedures of estimating the number and then locations of the breaks, considerable computation may be circumvented by inspecting the minimax wavelet estimate as a preliminary step.

In empirical research, the parameter instability might be addressed by either introducing time series specifications for the parameter or fitting elaborate non-linear alternatives. Neither approach directly confronts the issue if the true source of misspecification is parameter instability in time. Furthermore, the econometrician is forced to impose additional structures on the model, which may not be relevant. In such situations, a simple linear model that allows for both abrupt structural breaks and smooth structural changes, should one be available, deserves consideration before other alternatives. We highlight some examples where our model is applicable below.

Our model specializes to a capital asset pricing model (CAPM) with time-varying beta. Ample empirical evidence exists for the inadequacy of the static CAPM model and the asset pricing literature contains extensive investigations of the CAPM model with time-varying beta. Chen and Keown (1981) extracts the component of the unsystematic risk due to non-stationary beta under the random walk assumption on beta. One can also adopt a model where the returns follow a time series specification as in Bodurtha and Mark (1991). Time series specifications carry the implication that changes in the joint distribution of all the securities in the market are captured by the parameters within the model. This may prove to be inadequate, either due to simple misspecification or that the parameters themselves vary in time. Ghysels (1998) shows that the latter case does indeed arise. It was shown that when both the static model and a parametric time-varying beta model are misspecified, there are situations where the static model could outperform the parametric time-varying beta model in pricing risk. The parameters themselves can exhibit non-constancy and result in a sufficiently severe misspecification to overcome any additional flexibility gained by allowing for parametric time-variation. Modelling beta as a deterministic arbitrary, not necessarily smooth, function removes such misspecification issues. Also of interest is a model allowing for time-varying alpha rather than beta. For example, Evans (1994) shows that the contribution of time-varying beta to variation in returns is small compared to the contributions of changing risk premia. This would also be a special case of our model.

The predictive regression literature studies the regression of excess returns on dividend yield ratio, taking as given stylized facts persistence and endogeneity of valuation predictors, assuming a correctly specified model. Cochrane (1999) asserts stock returns are predictable over long horizons. Recent research has found evidence both for and against this assertion by considering different non-linear alternatives. The non-linear alternative used in Kilian and Taylor (2003) and Rapach and Wohar (2005) features a regressor time series with a mean-reversion property. They found longer forecasting horizon in predictive regression leads to more power. On the other hand, Ang and Bekaert (2007) constructs a present value model where the dividend yield is a nonlinear function of interest rates, excess returns, and cash flows and showed no power gain for long run predictive regression under this misspecification. Similarly Maynard and Ren (2014) fits the predictive linear regression to data generated a linear model where the coefficients switch in a Markov fashion and found no power gain of long range regression over short range. Dangl and Halling (2012) considers a framework where the regressors in the predictive regression follows a random walk. As in the asset pricing case, a time varying coefficient linear model would remedy the potential misspecification of parametric nonlinear models condition on existing data and suggest reasonable candidates of nonlinear alternatives for forecasting purposes.

Nellis and Longbottom (1981) conducts an econometric analysis of United Kingdom housing prices. They found that a regression model of housing prices on consumer prices and income fails the Chow test for parameter constancy. Data shows that the relationship between real house prices and real disposable income is not constant, which can be attributed to changes to the economy during the period examined such as financial deregulation and interest rate fluctuation. In addressing this, various time series specifications, such as state space or GARCH, have been devised. See, for example, Brown *et al.* (1997). Guirguis *et al.* (2005) considers the same issue for United States housing prices. They also found, in a linear model setting, the failure of Chow test provides evidence of susceptibility of housing prices to structural changes. Time series specifications introduces additional structure in the model that is not implied by rejection of the Chow test null hypothesis. A linear model with structure breaks is the Chow alternative.

The rest of the paper is organized as follows. In Section 2, we specify the model and recall standard facts in estimating the time-invariant part. Section 3 summarizes relevant aspects of wavelet theory. Section 4 discusses our approach to the de-noising problem for the time-varying parameter. Certain folklore facts regarding the classical nonparametric model is recast in a way that we believe makes clear the superiority of the wavelet approach relative to other estimators. In a model with white noise error terms, results in Donoho (1995a) provides a wavelet minimax estimator that preserves the smoothness of the true parameter, although a technical lemma is needed in bridging to our slightly different formulation. We also show that the model with short range dependent errors decomposes asymptotically into white noise models across different scales. This suggest naturally a scale-dependent wavelet estimator that is minimax. Section 5 contains simulation results. We conclude afterwards.

2 The model and estimation of time invariant parameters

We consider the linear model with time series data

$$Y_t = \sum_{i=0}^{m} \beta_t x_t + \gamma(t) x_{m+1,t} + \epsilon_t.$$

where

1. The parameters $\beta_0, \cdots, \beta_m$ are time-invariant.

2. $\{(x_{1,t}, \cdots, x_{m+1,t}, \epsilon_t)\}$ is strictly stationary α-mixing.

3. The time series $\{(x_{1,t}, \cdots, x_{m+1,t}, Y_t)\}$ is sampled at the rate $\frac{1}{n}$ on the interval $[0, 1]$, i.e. at times $\frac{1}{n}, \frac{2}{n}, \cdots, 1$.

4. The distribution of the regressor x_{m+1} has support bounded away from zero.

5. The time-varying parameter $\gamma(t)$ is Riemann integrable and lies in a Besov space $B^{\eta}_{p,q}$ (defined precisely below). The restriction to the unit interval is for technical convenience and without loss of generality.

Assumptions 1 and 2 are for consistent estimation of time-invariant part of the model, with consistency in the sense of in-fill asymptotics as Assumption 3 indicates. Assumption 4 is required for minimax estimation of the time-varying parameter γ. Assumption 5 specifies the type of function space that contains γ.

At any randomly chosen $s_0 \in (0,1)$, γ is continuous. Therefore $(\beta_1, \cdots, \beta_m, \gamma(s_0)) \in \mathbb{R}^{m+1}$ can be consistently estimated by a Nadaraya-Watson estimator. Let $K(\cdot)$ be a kernel function, h_n a bandwidth sequence satisfying $h_n \to 0$ and $nh_n \to \infty$. Then the Nadaraya-Watson estimate is

$$(\hat{\beta}_1, \hat{\beta}_m, \hat{\gamma}(s_0)) = \arg\min_{\theta \in \mathbb{R}^{m+1}} \sum_{t=1}^{n} (y_t - \mathbf{x}_t'\theta)^2 K(\frac{\frac{t}{n} - s_0}{h_n}),$$

where $\mathbf{x}_t' = ((x_{1,t}, \cdots, x_{m+1,t},)$. The closed form solution is

$$(\hat{\beta}_1, \hat{\beta}_m, \hat{\gamma}(s_0))' = (\sum_{t=1}^{n} K(\frac{\frac{t}{n} - s_0}{h_n})\mathbf{x}_t\mathbf{x}_t')^{-1}(\sum_{t=1}^{n} K(\frac{\frac{t}{n} - s_0}{h_n})\mathbf{x}_t y_t).$$

We recall the following standard consistency result for the Nadaraya-Watson estimator (see, for example, Theorem 1 in Cai (2007)):

Theorem 2.1. *If the kernel K is symmetric, has compact support, and is differentiable with continuous derivative, then under Assumptions 1, 2, and 5, the Nadaraya-Watson estimator is consistent.*

3 Wavelets and function spaces

We summarize in this section relevant aspects of wavelet theory. Let ψ be a Daubechies mother wavelet of compact support having r vanishing moments, r continuous derivatives and unit L^2-norm (see Daubechies (1988)). An orthonormal basis of the Hilbert space $L^2(\mathbb{R})$ is generated using ψ by integer translations and dyadic dilations by defining ($\mathbb{Z}$ denotes the integers):

$$\psi_{jk} = 2^{\frac{j}{2}}\psi(2^j x - k), \quad j, k \in \mathbb{Z}.$$

For a function in $L^2(\mathbb{R})$, its wavelet decomposition is $f(x) = \sum_{j,k} \theta_{jk}\psi_{jk}(x)$ where $\theta_{jk} = \int f(x)\psi_{jk}(x)dx$. For $L^2[0,1]$, an appropriate modification can be made on a subset of $\mathbb{Z} \times \mathbb{Z}$ so that the resulting $\{\psi_{jk}\}$ form an orthonormal basis of $L^2[0,1]$(Cohen *et al.* (1993)).

Wavelet bases are unconditional bases for a wide variety of functional spaces which are convenient in modelling spatially inhomogeneous signals. Empirically, this means that wavelets provide parsimonious representation of functions whose smoothness can, for example, vary between possible discontinuities. We work with *Besov spaces* and refer to Meyer (1995) and Triebel (1978) for more details on functional analytic properties of wavelets. Unlike the classical Sobolev or Hölder spaces, Besov spaces go beyond continuity and allow one to quantify smoothness of, for instance, cádlág functions. Allowing γ to lie in a Besov space means that our model accommodates structural breaks of arbitrary type.

The *Paley-Littlewood* definition of Besov space is as follows (Meyer (1995)). Let $\mathcal{S}'$ be the space of tempered distributions, i.e. the topological vector space dual of the Schwartz space $\mathcal{S}$ of C^∞-test functions (Gelfand and Vilenkin (1964)). Fix $\Psi, \{\Phi_n\}_{n\geq 0} \subset \mathcal{S}$ such that their Fourier transforms $\hat{\Psi}, \{\hat{\Phi}_n\}$ form a partition of unity subordinate to the open cover $A_0 = (-1, 1)$, $A_n = \{2^{n-1} < |\xi| < 2^{n+1}\}$. So an element $f \in \mathcal{S}'$ can be written as $f = \Psi * f + \sum_{n\geq 0} \Phi_n * f$. f is said to lie in the inhomogeneous Besov space $B^\alpha_{p,q}$ if

$$\|\Psi * f\|_{L^p} + (\sum_{n\geq 0}(2^{n\alpha}\|\Phi_n * f\|_{L^p})^q)^{\frac{1}{q}} < \infty.$$

As a corollary of (sufficiently smooth) wavelets forming an unconditional basis for $B^\eta_{p,q}$, f lies in the inhomogeneous Besov space $B^\eta_{p,q}$ if

$$\sum_{j\geq j_0}\left(2^{jq(\eta+\frac{1}{2}-\frac{1}{p})}\|\theta_{jk}\|_{l^p}\right)^q < \infty.$$

Example In the model being considered, let the time-varying parameter $\gamma(t) = 1_{[0,\frac{1}{2}]}(t)$, the indicator function on $[0, \frac{1}{2}]$. This is a parameter that has an abrupt structural break at $t = \frac{1}{2}$ but constant otherwise, a Chow alternative hypothesis. Let $\mathcal{F}(\gamma)$ denote the Fourier transform of γ, then

$$\mathcal{F}(\gamma)(\xi) = \frac{1}{4}e^{-2\pi i\cdot\frac{1}{4}\xi}\cdot\frac{\sin\frac{\pi\xi}{4}}{\frac{\pi\xi}{4}}.$$

Choose $p = 2$, then in the above notation, the L^2-norm of the n-th Paley-Littlewood piece is, up to a multiplicative constant independent of n,

$$\|\Phi_n * g\|_{L^2} \approx 2^{-\frac{n}{2}}.$$

Therefore $\gamma \in B^\eta_{2,q}$ if and only if

$$\sum_n 2^{n(\eta-\frac{1}{2})q} < \infty,$$

For $q \in [1,\infty]$, this is true whenever $\eta < \frac{1}{2}$. For $\eta = \frac{1}{2}$, one must have $q = \infty$.

More generally, almost all sample paths of a Lévy process, which can be, for example, a càdlàg function where each continuous piece is a Brownian sample path, lie in a Besov space (Herren (1997)). Almost all sample paths of a Brownian motion belong to the suitable Besov spaces $B^{\eta}_{p,q}$ with $1 \leq p, q \leq \infty$, $\frac{1}{p} < \eta \leq \frac{1}{2}$. Such sample paths have the property, for example, of crossing 0 infinitely many times on the time interval $(0,\epsilon)$ for ϵ arbitrarily small. From a practical perspective, this descriptive power of Besov spaces means that, our model places no restrictions on the time-varying parameter γ, allowing for features such as smooth structural changes, abrupt jumps, or combinations thereof.

The family of Besov spaces contains both both L^2-Sobolev spaces (the cases $p = q = 2$) and Hölder spaces (the case $p = q = \infty$). The assumption of twice-differentiability that is common in the time-varying coefficient literature correspond to the case $\eta = 2$ and $p = q = \infty$.

4 Estimation of the time-varying parameter

Plugging in the consistent estimates of $\beta_1, \cdots, \beta_m$ and re-writing the model gives, for $t = 1, \cdots, n$,

$$\begin{aligned}
\frac{Y_t - \sum_i \hat{\beta}_i x_{i,t}}{x_{m+1,t}} &= \frac{Y_t - \sum_i \beta_i x_{i,t}}{x_{m+1,t}} + \frac{\sum_i (\beta_i - \hat{\beta}_i) x_{i,t}}{x_{m+1,t}} \\
&= \underbrace{\gamma(t) + \frac{\epsilon_t}{x_{m+1,t}}}_{\text{De-noising problem}} + \underbrace{\frac{\sum_i (\beta_i - \hat{\beta}_i) x_{i,t}}{x_{m+1,t}}}_{\text{Estimation error of } \beta_1, \cdots, \beta_m} .
\end{aligned}$$

The above expression suggests that the problem of estimating γ is a denoising problem conditional on the true time-invariant parameters with additional additive error, the estimation error of the time-invariant parameters. We first consider the de-noising problem, showing the faithfulness of the wavelet estimator with respect to the smoothness of γ and establishing its minimax rates of convergence, before incorporating estimation error.

4.1 The De-noising Problem

The term $\gamma(t) + \frac{\epsilon_t}{x_{m+1,t}}$ can be viewed as noisy observations of γ. Thus estimation of γ is now a de-noising problem. In more compact notation, rewrite as

$$Z_t = \gamma(t) + u_t.$$

Allowing for jumps necessarily means that one must abandon the notion of pointwise consistency. Instead, we measure the loss by the squared norm $\| \cdot \|^2_{L^2}$ on $L^2[0,1]$. The corresponding risk is the mean integrated square error (MISE). For a given estimator $\hat{\gamma}$, the MISE is

$$\mathbb{E}[\|\hat{\gamma} - \gamma\|_2^2] = \mathbb{E}[\int_0^1 |\hat{\gamma}(t) - \gamma(t)|^2 dt].$$

The benchmark is the *minimax risk*. The (asymptotic) minimax risk over a subset $\mathcal{F}$ (e.g. a subspace such as $B^{\alpha}_{2,q}$ or a family of subspaces) is defined by

$$\lim_{n\to\infty} \inf_{T} \inf \sup_{\gamma \in \mathcal{F}} \mathbb{E}[\|\hat{\gamma} - \gamma\|_2^2],$$

where $\inf_T$ denotes infimum over all $\mathcal{F}$-valued maps measurable with respect to data.

4.1.1 White noise $\{\epsilon_t\}$

If the error terms $\{\epsilon_t\}$ in the model is homoskedastic white noise, Assumption 4 on the distribution of x_{m+1} implies that $\{u_t\}$ is heteoskedastic white noise with uniformly bounded variance $\leq \sigma^2$. We treat the model

$$Z_t = \gamma(t) + u_t$$

as one where u_t is homoskedastic white noise with variance σ^2. In practice, error variance is estimated from the highest level wavelet coefficients, the most noisy part of the noisy observations of γ.

The de-nosing problem with homoskedastic white noise and the infinite dimensional Gaussian sequence model are shown formally to be statistically equivalent by Brown *et al.* (1996). We offer a development of this equivalence that explicates the utility of wavelets, first as the basis used in an orthogonal series estimator and second in facilitating the extension to $\{\epsilon_t\}$ with long-range dependence.

From de-noising to filtering model: For a given n, define a stochastic process on $[0,1]$ as follows:

$$F_t^{(n)} = \frac{1}{n}\sum_{i=1}^{\lceil nt \rceil} Z_i = \frac{1}{n}\sum_{i=1}^{\lceil nt \rceil} \gamma(t) + \frac{1}{n}\sum_{i=1}^{\lceil nt \rceil} u_t.$$

The drift term $\frac{1}{n}\sum_{i=1}^{\lceil nt \rceil}\gamma(t)$ converges to $\int_0^t f(t)dt$. By the classical Functional Central Limit Theorem (see, e.g. Davidson (1994)), the process

$$t \mapsto \frac{1}{n}\sum_{i=1}^{\lceil nt \rceil} u_t$$

converges to $\frac{\sigma}{\sqrt{n}}B_t$, where B_t is the standard Brownian motion on $[0,1]$ weakly.[1] Therefore the sequence of processes $\{F_t^{(n)}\}$ converges to the Itô process

$$dF_t = \gamma(t)dt + \sigma\frac{1}{\sqrt{n}}dB_t,\ t \in [0,1].$$

From filtering model to Gaussian sequence model: We make use of the Lévy-Ciesielski-Itô construction (Lévy (1954), Ciesielski (1961), Itô *et al.* (1968)) of the Brownian motion B_t.[2] Let $\{\psi_j\}$ of $L^2[0,1]$ be an arbitrary orthonormal basis, then the standard Brownian motion on $[0,1]$ can be expressed as

$$dB_t = \sum_j w_j\psi_j(t)dt$$

where $\{w_j\}$ is standard Gaussian white noise and the series converges in the mean square sense. Therefore

$$\int_0^1 \psi_j(t)dF_t = \int_0^1 \psi_j(t)\gamma(t)dt + \int_0^1 \psi_j(t)\cdot\sigma\frac{1}{\sqrt{n}}dB_t$$

where $\int_0^1 \psi_j(t)\cdot\sigma\frac{1}{\sqrt{n}}dB_t = \sigma\frac{1}{\sqrt{n}}w_j$ by orthonormality of $\{\psi_j\}$. In a stochastic sense, the Fourier transform of the filtering model, taken with respect to $\{\psi_j\}$, is the Gaussian sequence model

$$X_j = \theta_j + e_j,$$

where $(\theta_j) \in l^2(\mathbb{N})$, $e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n}),\ j = 1, 2, \cdots$.[3]

[1] In the sense of probability measures on the space of càdlàg functions on $[0,1]$.

[2] We recall the details of this construction in the Appendix.

[3] $l^2(\mathbb{N})$ denotes the Hilbert space of square summable sequences.

We summarize the above in a theorem:

Theorem 4.1. *The following three models are statistically equivalent as $n \to \infty$:*

(i)

$$Z_t = \gamma(t) + u_t$$

where $\{u_t\}$ is homoskedastic white noise with variance σ^2, $\gamma \in L^2[0,1]$ is Riemann-integrable, and Z_t is sampled at the rate $\frac{1}{n}$.

(ii) The filtering model where the drift of the Itô process

$$dF_t = \gamma(t)dt + \sigma \frac{1}{\sqrt{n}} dB_t$$

is to be estimated.

(iii) The Gaussian sequence model

$$X_j = \theta_j + e_j,$$

where $(\theta_j = \int_0^1 \gamma\psi_j) \in l^2(\mathbb{N})$ for some orthonormal basis $\{\psi_j\}$, $e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n})$, $j = 1, 2, \cdots$.

Remark 4.2. *The above development, with passage through the intermediate filtering model, generalizes to the case of long range dependent $\{\epsilon_t\}$. In the white noise case, model equivalence holds for any basis, with the means of the resulting Gaussian sequence model being the expansion coefficients with respect the the chosen basis. The effectiveness of wavelet lies in that a wide variety of functions can be parsimoniously encoded by their wavelet coefficients. The long range dependent case, however, the additional property of wavelets to de-correlate serial dependence is required to arrive at a Gaussian sequence model.*

The sequence (θ_j) are the expansion coefficients of γ with respect to the chosen basis ψ_j whose empirical counterpart is (X_j). The seminal estimation result in this setting is due to Pinsker, who used the classical Fourier basis and showed that minimax risk can be achieved over Sobolev ellipsoids by shrinking the empirical Fourier coefficients (Pinsker (1980)). While a shrinkage estimator is natural in this setting, it is impossible to extend the minimax result beyond Sobolev spaces.[4] Wavelets, however, allows one to extend beyond

[4]We have the following fact from Fourier analysis: For any $g \in L^2[0,1]$, there exists a continuous h on $[0,1]$ such that all the Fourier coefficients of h are larger than those of g (Katznelson (2004)). Therefore shrinking the Fourier coefficients does not in general preserve smoothness.

continuous functions. Define the *soft threshold function* η_λ with threshold λ by

$$\eta_\lambda(x) = \text{sgn}(x)(|x| - \lambda)_+.$$

The *wavelet universal threshold estimator* $\hat{\gamma}_n$, where n is sample size, applies $\eta_\lambda(\cdot)$ to each empirical wavelet coefficient with threshold $\lambda = \hat{\sigma}\sqrt{\frac{2\log n}{n}}$, where $\hat{\sigma}$ is the median absolute deviation estimate of σ from the highest level of wavelet coefficients. The following theorem, proved in Donoho (1995a), says that not only does wavelet universal threshold estimator achieve minimax risk over a Besov scale (up to a log factor), the estimated function $\hat{\gamma}$ is as smooth as γ with probability approaching 1.[5]

Theorem 4.3. *(Donoho 1995) Let ψ, the mother wavelet that generates the wavelet basis used, have r vanishing moments and r continuous derivatives, where $r > \max\{1, \eta\}$. In the de-noising model with white noise $\{\epsilon_t\}$, let $R_n(p, q, \eta, L)$ denote the minimax risk over the Besov ball $B^\eta_{p,q}(L)$. Then*

(i)

$$\lim_{n\to\infty} \frac{\sup_{\gamma \in B^\eta_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n - \gamma\|^2]}{R_n(p, q, \eta, L)} = 2\log n + 1$$

for all $1 \leq p, q \leq \infty$, $0 < L < \infty$, and $\eta_0 < \eta < r$ where

$$\eta_0 = \max\{\frac{1}{p}, 2(\frac{1}{p} - \frac{1}{2})_+\}.$$

(ii) There exists a constant C such that

$$Prob\{\|\hat{\gamma}_n\|_{B^\eta_{p,q}} \leq C\| \ \gamma\|_{B^\eta_{p,q}}\} \to 1.$$

Even as we abandon the notion of pointwise estimation, conclusion (ii) of Theorem 4.3 guarantees that the estimate $\hat{\gamma}_n$ is as smooth as the true γ, as measured by the Besov norm $\|\cdot\|_{B^\alpha_{p,q}}$, with probability approaching 1. Therefore empirically there are no spurious jumps in the estimate.

[5]Strictly speaking, the statements from Donoho (1995a) applies to the sequence of truncated finite dimensional Gaussian sequence models. The gap with our formulation is bridged with a technical lemma, given in the Appendix.

4.1.2 Short range dependent $\{\epsilon_t\}$

The covariance-stationary time series $\{\epsilon_t\}$ is said to have *short range dependence* if its autocorrelation function ρ is absolutely summable: $\sum_{-\infty}^{+\infty} |\rho(h)| < \infty$ and *long range dependence* if ρ has sub-hyperbolic decay: $\rho(h) \approx \frac{A}{|h|^\alpha}$ for some $0 < \alpha < 1$ and $A > 0$.

The model equivalence for the de-noising problem in fact generalizes to short range dependent $\{\epsilon_t\}$. The error partial sum process

$$\frac{1}{n} \sum_{i=1}^{\lceil nt \rceil} u_t$$

still converges to a Brownian motion $\frac{\tau}{\sqrt{n}} dB_t$ where $\tau^2 = \sum_{-\infty}^{+\infty} |\rho(h)|$ (De Jong and Davidson (2000)). Therefore the de-noising problem is still equivalent to estimating the means of a sequence of independent Gaussian random variables. In simulations, however, we found that the presence of short range dependence causes the estimates of γ be be noisier than the white noise case. This indicates that serial correlation dissipates slowly relative to sample size and motivates us to consider the situation where the de-noising problem, the second step in our procedure, features long range dependent errors.

Unlike the short range dependent case, the whitening effect of wavelets, not shared by other basis, is now required to establish model equivalence. In this more general setting, the limit process is now the fractional Brownian motion dB_t^H is with Hurst index H, which is determined by $H = 1 - \frac{\alpha}{2} \in (\frac{1}{2}, 1)$. dB_t^H is a mean-zero Gaussian process that behaves like $(\Delta t)^H$ for small time increments Δt, and self-similar in the sense that $B^H(ct)$ and $c^H B^H(t)$ have the same distribution. For $H = \frac{1}{2}$, dB_t^H is the Brownian motion. Unlike the Brownian motion, dB_t^H does not have independent increments in general nor is it a semi-martingale. (The Appendix contains a precise definition of dB_t^H.)

By a version of Functional Central Limit Theorem for long range dependent processes (Taqqu (1975)), the observation and error partial sum processes satisfy

$$n^{1-H}(F_t^{(n)} - \int_0^t \gamma(s) ds) \to \tau B_t^H$$

where the asymptotic variance $\tau^2 = \frac{2A}{(1-\alpha)(2-\alpha)}$. This gives the long-range filtering model

$$dF_t = \gamma(t)dt + \frac{\tau}{(\sqrt{n})^\alpha} dB_t^H.$$

Unlike the case of Brownian motion, where the Iτ̂o isometry yields i.i.d. Gaussian sequence for any basis of $L^2[0,1]$, the long memory situation require unique properties of wavelets to decorrelate the fractional Brownian motion. By performing a principal component analysis on the reproducing kernel of dB_t^H using wavelets, we obtain a Lévy-Ciesielski-Iτ̂o type representation (details provided in Appendix):

$$dB_t^H = \sum_{jk} w_{jk} v_{jk}(t) dt,$$

where v_{jk} is approximately orthogonal to $\{\psi_{jk}\}$. The stochastic integrals

$$\int \psi_{jk} dF_t = \int \psi_{jk} \gamma(t) dt + \frac{\tau}{(\sqrt{n})^\alpha} \psi_{jk} dB_t^H$$

give a Gaussian sequence model whose error terms are approximately white noise, with level-dependent error variance.

Theorem 4.4. *Denote the empirical wavelet coefficient by $X_{jk} = \int \psi_{jk} dF_t$. Then*

(i)

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^\alpha} \sigma_j e_{jk},$$

where, at each level j, $\sigma_j = 2^{-j(1-\alpha)}$.

(ii) The random variables $e_{jk} = \frac{1}{\sigma_j} \int_0^1 \psi_{jk} dB_t^H$ have mean zero, variance 1, and are approximately uncorrelated in the sense that $0 < c_0 \leq \; Var(e_{jk} | e_{j'k'}, \; (j',k') \neq (j,k)) \leq 1$.

It was shown in Donoho (1995b) that a Gaussian sequence model whose noise satisfies the approximate uncorrelated condition (ii) has the same asymptotic minimax risk as a model with independent noise. In other words, asymptotically the model is equivalent to

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^\alpha} \sigma_j e'_{jk}$$

where e'_{jk}'s are i.i.d. standard normal. Since for a fixed resolution level j, the equivalent model is a Gaussian sequence model with homoskedastic independent noise for which the unversal wavelet threshold estimator achieves minimax risk, as a corollary of Theorem 4.3 we arrive at a level-dependent thresholding estimator for the long memory case.

Theorem 4.5. *Let ψ, the mother wavelet that generates the wavelet basis used, have r vanishing moments and r continuous derivatives, where $r > \max\{1, \eta\}$. In the de-noising* *problem with long memory error $\{\epsilon_t\}$, let $R_n(p,q,\eta,L)$ denote the minimax risk over the Besov ball $B^{\eta}_{p,q}(L)$ and $\hat{\gamma}_n$ be the estimate obtained by applying the universal wavelet threshold estimator to each level j. Then*

(i)

$$\lim_{n\to\infty} \frac{\sup_{\gamma\in B^{\eta}_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n-\gamma\|^2]}{R_n(p,q,\eta,L)} = 2\log n + 1,$$

for all $1\le p,q\le\infty$, $0<L<\infty$, and $\eta_0<\eta<r$ where

$$\eta_0 = \max\{\frac{1}{p}, 2(\frac{1}{p}-\frac{1}{2})_+\}.$$

(ii) There exists a constant C such that

$$Prob\{\|\hat{\gamma}_n\|_{B^{\eta}_{p,q}} \le C\|\gamma\|_{B^{\eta}_{p,q}}\} \to 1.$$

Also by asymptotic equivalence, the (non-adaptive) minimax rate of convergence of the wavelet threshold estimator can is therefore the same as that obtained in Donoho *et al.* (1998) for the Gaussian sequence model with level-dependent noise

$$X_{jk} = \theta_{jk} + \frac{\tau}{(\sqrt{n})^{\alpha}}\sigma_j e'_{jk},$$

under the assumption that the time-varying parameter γ is sufficiently regular relative to the correlation structure of the limit process dB^H_t.

Theorem 4.6. *Suppose $\eta+\frac{1}{2}-\frac{1}{p} > \frac{\alpha(2-p)}{2p}$, then the wavelet threshold estimator is rate-optimal and its minimax risk satisfies*

$$\liminf_{n\to\infty} \sup_{\gamma\in B^{\eta}_{p,q}(L)} \mathbb{E}[\|\hat{\gamma}_n-\gamma\|^2] = O(n^{-r})$$

where $r = 2\cdot\frac{(\eta+\frac{1}{2}-\frac{1}{p})\cdot\frac{\alpha}{2}}{\eta+\frac{1}{2}-\frac{1}{p}+\frac{\alpha}{2}}$.

4.2 Estimation Error

From the time-invariant part of the model, the mean square error $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^m(\hat{\beta}_j-\beta_j)^2$, with optimal bandwidth selection, is known to be of order $O_p(n^{-\frac{4}{5}})$Bierens (1987). Also, since m is finite, the norms $\|\boldsymbol{\beta}\|_2$ and $\|\boldsymbol{\beta}\|_\infty = \max\limits_{1\le j\le m}\beta_j$ on $\mathbb{R}^m$ are equivalent. Therefore $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_\infty$ is of order $O_p(n^{-\frac{4}{5}})$.

Making the additional assumption that distributions of the regressors x_j, $j = 1, \cdots, m$, have compact support, the estimation error term[6]

$$\left|\frac{\sum_i(\beta_i - \hat{\beta}_i)x_{i,t}}{x_{m+1,t}}\right| \leq const \cdot \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_\infty = O_p(n^{-\frac{2}{5}}).$$

Therefore

$$\frac{Y_t - \sum_i \hat{\beta}_i x_{i,t}}{x_{m+1,t}}$$

is asymptotically equivalent to

$$\gamma(t) + \frac{\epsilon_t}{x_{m+1,t}} + O_p(n^{-\frac{2}{5}}). \tag{1}$$

With $\hat{\gamma}_n$ denoting the wavelet estimate from the de-noising problem, let $\hat{\hat{\gamma}}_n$ be the wavelet threshold estimator applied to (1). Because the wavelet threshold operator is identity on the one dimensional subspace spanned by constants, we have

$$\hat{\hat{\gamma}}_n = \hat{\gamma}_n + O_p(n^{-\frac{2}{5}}),$$

which immediately gives the following:

Theorem 4.7. *Suppose the distributions of the regressors x_j, $j = 1, \cdots, m$, have compact support and the time-varying parameter is not constant. If $\eta + \frac{1}{2} - \frac{1}{p} > \frac{\alpha(2-p)}{2p}$, then the wavelet threshold estimator $\hat{\hat{\gamma}}$ for time-varying parameter γ in the full model stated in Section 2 has minimax risk satisfying*

$$\liminf_{n\to\infty} \sup_{\gamma \in B^\eta_{p,q}(L)} \mathbb{E}[\|\hat{\hat{\gamma}}_n - \gamma\|^2] = O(\max\{n^{-r}, n^{-\frac{4}{5}}\})$$

where $r = 2 \cdot \frac{(\eta + \frac{1}{2} - \frac{1}{p}) \cdot \frac{\alpha}{2}}{\eta + \frac{1}{2} - \frac{1}{p} + \frac{\alpha}{2}}$.

In the full model, one achieves the worst of the two rates from the time-invariant and time varying parts. Similarly, since $\|\hat{\hat{\gamma}}_n\|_{B^\eta_{p,q}} = \|\hat{\gamma}_n\|_{B^\eta_{p,q}} + O_p(n^{-\frac{2}{5}})$, the faithfulness result of Theorem 4.5(ii) extends to the full model:

$$\text{Prob}\{\|\hat{\hat{\gamma}}_n\|_{B^\eta_{p,q}} \leq const \cdot (\|\gamma\|_{B^\eta_{p,q}} + n^{-\frac{2}{5}})\} \to 1.$$

In summary, our proposed estimator consists of the following two steps:

[6]Recall that $x_{m+1,t}$ is assumed to be bounded away from zero.

1. Estimate the time-invariant part of the model consistently by local smoothing kernel methods at a randomly chosen $s_0 \in (0,1)$.

2. Apply a non-linear wavelet orthogonal series estimator to obtain $\hat{\hat{\gamma}}$.

5 Monte Carlo simulations

5.1 White noise $\{\epsilon_t\}$

We simulate the following model

$$Y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \gamma(t) x_{4,t} + \epsilon_t, \tag{2}$$

where

- The value of the time-invariant parameters are $\alpha = 7, \beta_1 = 5, \beta_2 = -6, \beta_3 = 2$.
- The functional parameter $\gamma : [0,1] \to \mathbb{R}$ is defined by

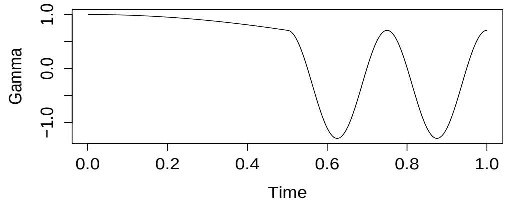
$$\gamma(t) = \begin{cases} \cos \frac{\pi}{2} t & \text{if } t \in [0, \frac{1}{2}] \\ \cos 8\pi t + (\cos \frac{\pi}{4} - \cos 4\pi) & \text{if } t \in [\frac{1}{2}, 1] \end{cases}.$$

  γ is a continuous but not C^1, being not differentiable at $t = \frac{1}{2}$. Figure 1(a) contains a plot of γ. This simulated data generating process has a structural change from low frequency to high frequency oscillation at $t = \frac{1}{2}$.
- The regressors and errors have the following independent distributions:
  - $x_{1,t} \sim$ i.i.d. χ^2_2.
  - $x_{2,t} \sim$ i.i.d. $\mathcal{N}(0,1)$.
  - $x_{3,t}$ is a realization of a ARMA$(1,2)$ time series with AR parameters 0.1 and MA parameters $(1,-1)$. The underlying innovation is standard normal white noise.
  - $x_{4,t}$ has i.i.d. distributions which is a χ^2_2 distribution shifted to the right by 1, making it bounded away from zero.
  - $\epsilon_t \sim$ i.i.d. $\mathcal{N}(0,4)$.

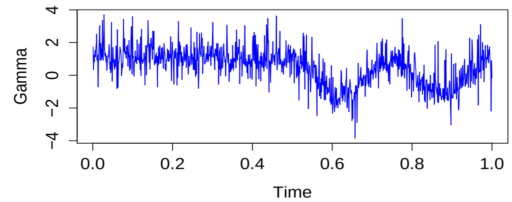
The sample size is $1,024$. We now describe a typical estimate. Using an Epanechnikov kernel, the estimates of time-invariant parameters are

$$\hat{\alpha} = 6.640, \hat{\beta}_1 = 5.058, \hat{\beta}_2 = -6.082, \hat{\beta}_3 = 2.023.$$

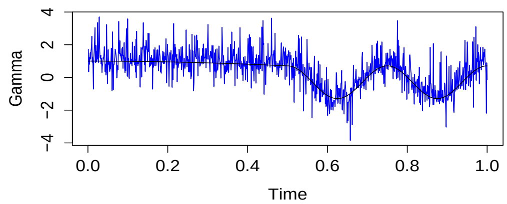
Plugging-in gives the noisy version of γ to be wavelet-thresholded, shown in Figure 1(c),



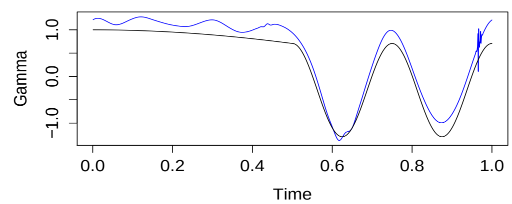
(a) The functional parameter γ



(b) Z_t obtained by plugging in consistent estimates of time-invariant parameters.



(c) Z_t in comparison with true γ.



(d) Estimate $\hat{\gamma}$ given by the universal wavelet threshold estimator

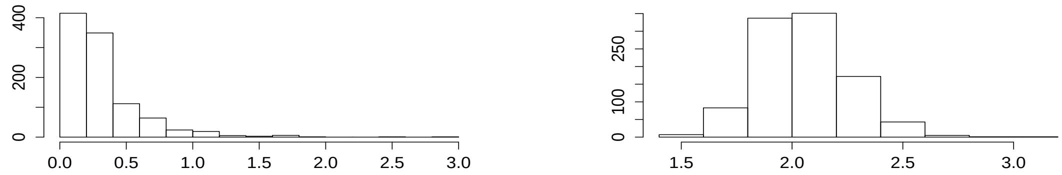
Figure 1: Estimate of γ by universal wavelet thresholding from a typical realization of data generated by (2). The errors $\{\epsilon_t\}$ in the simulated DGP is i.i.d. $\mathcal{N}(0,4)$ white noise.

$$Z_t = \frac{Y_t - \hat{\alpha} - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \hat{\beta}_3 x_{3,t}}{x_{4,t}}.$$

Even though Z_t is very noisy compare to γ, wavelet thresholding removes most of the noise, yielding estimate shown in Figure 1(d). The average mean square error (AMSE), the empirical counterpart to mean integrated square error (MISE), in this case is

$$\frac{1}{n}\sum_t (\hat{\hat{\gamma}}(t) - \gamma(t))^2 = 0.02996225.$$

On the other hand, the total energy contain in γ is $\frac{1}{n}\sum_t \gamma(t)^2 = 0.7018039$, making the relative error approximately 4.2%. This is in spite of a very large noise-to-signal ratio of 2.041884, as measure by the ratio of empirical variances of (Z_t) over that of $\gamma(t)$. Relative errors from a Monte Carlo simulations of 1000 repetitions are shown in Figure 2, with mean of approximately 6% and standard deviation 0.05. From the same 1000 simulations, the noise-to-signal variance ratios have a relatively large mean 2.001118 and standard deviation 0.09640994.

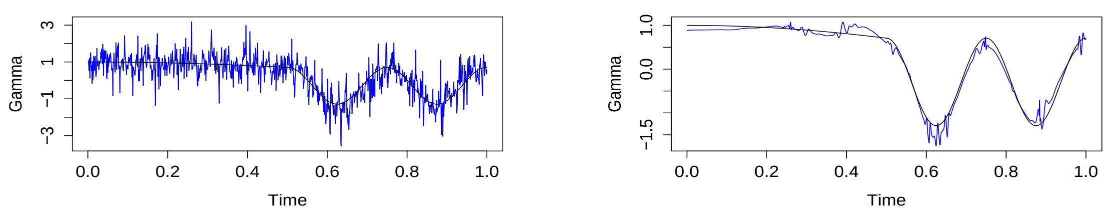


(a) Relative errors measured by MISE from 1000 simulations.

(b) Noise-to-signal ratio from 1000 simulations.

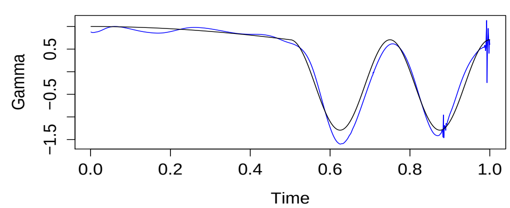
Figure 2: Monte Carlo simulation of 1000 repetitions. The average relative error is approximately 6%, which is small considering the amount of noise faced by the estimator.

5.2 Short-range dependent $\{\epsilon_t\}$



(a) Noisy γ in comparison with true γ.

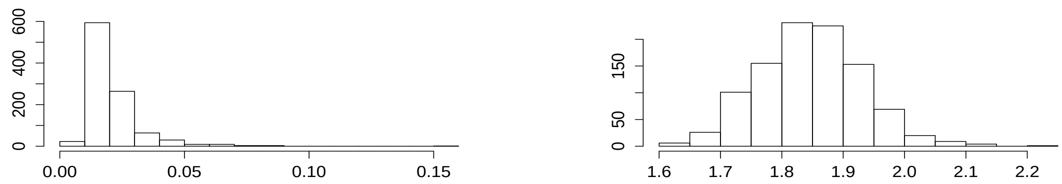
(b) Universal threshold estimate of γ when $\{\epsilon_t\}$ has short-range dependence.



(c) Level-dependent threshold estimate of γ

Figure 3: The wavelet universal and level-dependent threshold estimators applied to data with short-range dependent errors.

As stated in Section 4.2, in theory asymptotic minimaxity of the universal threshold estimator extends to models with such short range dependent error $\{\epsilon_t\}$. However, in simulations we found that the universal threshold estimator often gives somewhat noisy estimates of γ when $\{\epsilon_t\}$ has short range dependence. Figure 3(b) shows one such estimate. The simulated DGP satisfies all previous conditions except $\{\epsilon_t\}$ is now an AR(1) time series with AR parameter 0.5. Noticeable more noise survives universal thresholding than in the white



(a) Relative errors, mean $= 0.02083235 \times 100\%$, sd $= 0.01107202$.

(b) Noise-to-signal ratio, mean $= 1.846658$, sd $= 0.08510039$.

Figure 4: Monte-Carlo simulation of 1000 repetitions with short range dependent errors.

noise case. This is empirical evidence that while the observation processes still converge to the Brownian motion, serial dependence slows down the speed of convergence considerably. In comparison, applying the level-dependent threshold estimator specified in Theorem 4.5 to the same set of data gives an improved estimate, which is shown in Figure 3(c). Figure 4 shows a results from 1000 simulations using the level dependent threshold estimator; 4(a) shows the distribution of relative errors and 4(b) shows the distribution of the noise-to-signal ratio faced by the wavelet estimator in the de-noising problem.

5.3 Small sample

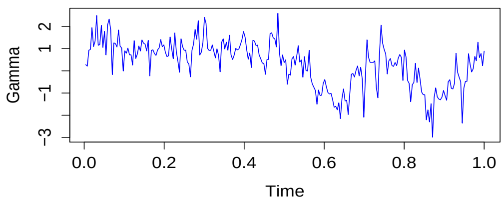
Figure 5 contains simulation result of 1000 repetitions with sample size reduced from $1,024$ to 256. For comparison, 5(b) shows the level-dependent threshold estimate and 5(c) universal threshold estimate for the same realization in 5(a). The level-dependent threshold estimator is used, as specified by our methodology. As expected, the quality of estimates worsens with smaller sample size but it is still acceptable in our view.

5.4 Discontinuous γ

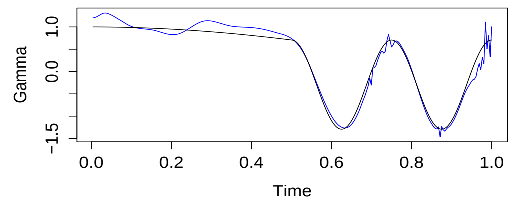
Consider now a γ that undergoes simultaneously smooth structural change and abrupt structural break at $t = \frac{1}{2}$, plotted in Figure 6(a):

$$\gamma(t) = \begin{cases} \cos \frac{\pi}{2} t & \text{if } t \in [0, \frac{1}{2}) \\ \cos 8\pi t + 3 & \text{if } t \in [\frac{1}{2}, 1] \end{cases}.$$

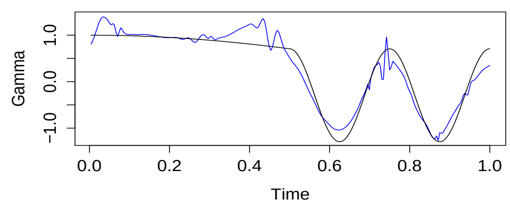
With a sample size of $1,024$ and short range dependent errors, typical estimates using the level-dependent threshold estimator and the universal threshold estimator are shown in Figures 6(c) and 6(d) respectively. Both capture well the true jump at $t = \frac{1}{2}$, although the



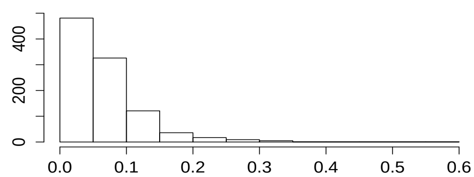
(a) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



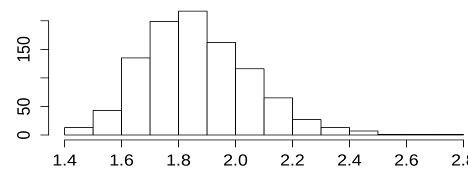
(b) Corresponding level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(c) Corresponding universal threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



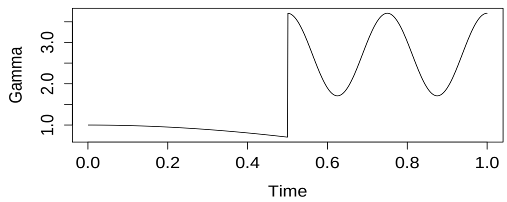
(d) Relative errors, mean $= 0.07042005 \times 100\%$, sd $= 0.057314$.



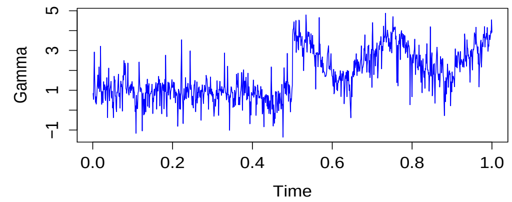
(e) Noise-to-signal ratio, mean $= 1.868884$, sd $= 0.1902459$

Figure 5: Monte-Carlo simulation of 1000 repetitions with short range dependent errors. Sample size reduced from $1,024$ to 256.

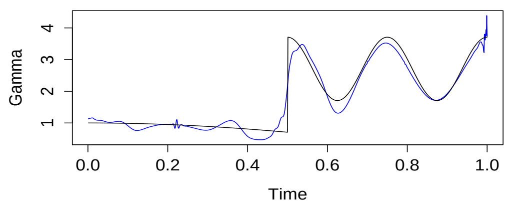
universal threshold estimate is visually more noisy. It is level-dependent estimator that is used in the 1000 repetitions. The average relative error of approximately 2.13% is approximately the same as the 2.08% obtained for a γ that is continuous, in Figure 4(c). The noise-to-signal ratio is approximately the same in the two cases. Thus the performance of the wavelet estimator is unaffected by possible discontinuities in γ. Whereas a pointwise estimator obtained by local smoothing would breakdown in the presence of discontinuity, the wavelet estimator can actually exploit the additional spatial inhomogeneity in γ. With a smaller sample size of 256, the wavelet estimate still retains the essential features of true γ as shown by results in Figure 8. Again comparing to the case where γ is continuous with the same sample size of 256, discontinuity in γ causes no discernible degradation on the quality



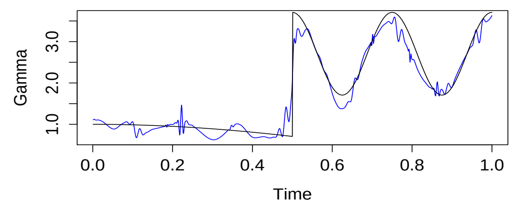
(a) γ with discontinuity.



(b) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



(c) Level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



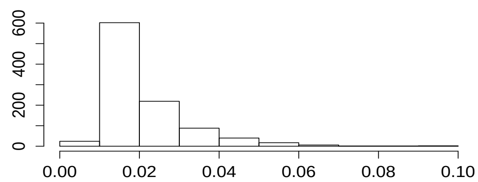
(d) Universal threshold estimate of γ.

Figure 6: Discontinuous γ with short range dependent errors.

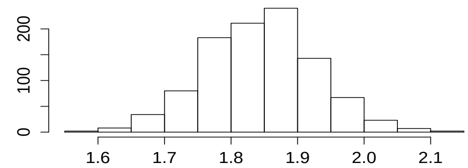
of the estimate, with the noise level being comparable.

6 Conclusion

We provide a framework for functional estimation of a time-varying parameter in a linear regression model. Leaving the paradigm of pointwise estimation, we estimate the parameter over the entire observed period simultaneously. This new perspective allows one to incorporate the machinery of wavelets. Wavelets stand out as the only family of basis functions that can efficiently encode spatial inhomogeneity and whiten serial correlated time series. Both properties are exploited in our methodology. By transforming the estimation problem into one of estimating wavelet coefficients, we make use of the fact that the wavelet decomposition of spatially inhomogeneous functions are concentrated at a few relatively large coefficients. Furthermore, the estimation problems are statistically independent across resolution levels even when the error terms feature serial correlation. The allowance for jumps, abrupt breaks, and smooth trends of any type is in contrast with local smoothing methods, which must impose smoothness assumptions. Wavelets also also computational advantages. A Monte Carlo simulation of 1000 repetitions of the estimator, as one performed in Section 5, takes approximately 25 seconds. In this more general environment, important characteristics of the parameter such as the locations of possible discontinuities and different types



(a) Relative error, mean $= 0.02125429 \times 100\%$, sd $= 0.01088821$



(b) Noise-to-signal ratio, mean $= 1.84448$, sd $= 0.08350869$

Figure 7: Monte-Carlo simulation of 1000 repetitions with short range dependent errors and γ with discontinuity.

of time trending behavior are captured by the wavelet estimate. It is of use for economics, finance as well as other disciplines where similar linear regressions in a time series context are relevant.

7 Appendix

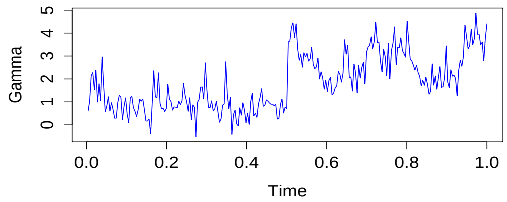
Technical lemma for Theorem 4.3: For $B^{\alpha}_{p,q}(L)$ with space of wavelet coefficients $\Theta \subset l^2(\mathbb{N})$, Donoho (1995a) considers a sequence of truncated Gaussian sequence models $\mathcal{M}_n$

$$X_j = \theta_j + e_j, \text{ where } e_j \sim \mathcal{N}(0, \frac{\sigma^2}{n}), \ 1 \leq j \leq n,$$

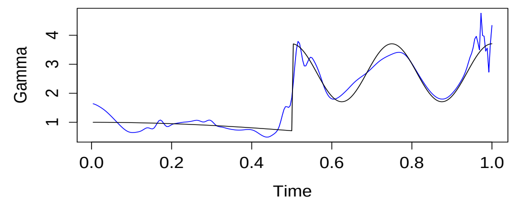
where for each n, $(\theta_j) \in \Theta \cap \mathbb{R}^n$ (as it is irrelevant for the argument in this case, we suppress the double indices for wavelet coefficients). $\hat{\gamma}_n$ is constructed by estimating the n coefficients. We show that this is without loss of generality in our global L^2-formulation. As any $\gamma \in L^2[0,1]$ can be approximated by its truncations γ_n, it is without loss of generality to estimate γ_n. Furthermore, estimating γ_n from its first n coefficients does not increase risk: In the filtering model, let $\mathbf{P}_\gamma$ and $\mathbf{P}_0$ be the probability measures on $C[0,1]$ that gives the distribution of $\mathbf{F} = dF_t$ and dB_t respectively. By Girsanov's theorem (Girsanov (1960)), the likelihood ratio is

$$\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0} = e^{\frac{1}{n}\sum_{j=1}^n \theta_j X_j - \frac{1}{2n}\sum_{j=1}^n \theta_j^2}.$$

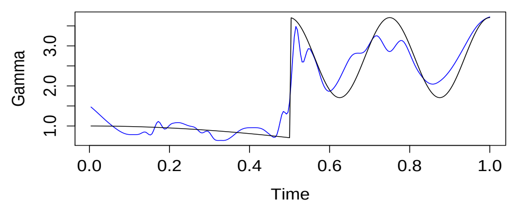
By Jensen's inequality, for any estimator $\hat{\theta}_j(\mathbf{F})$ where $1 \leq j \leq n$,



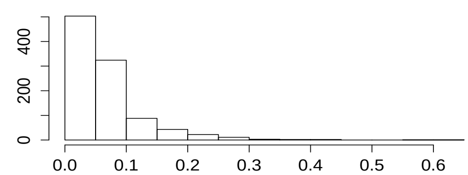
(a) One realization of noisy γ with AR(1) $\{\epsilon_t\}$.



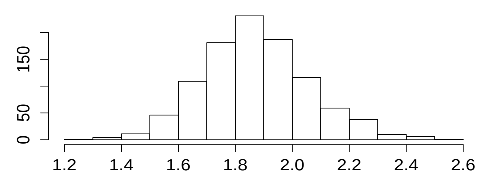
(b) Corresponding level-dependent threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(c) Corresponding universal threshold estimate of γ with AR(1) $\{\epsilon_t\}$.



(d) Relative errors, mean $= 0.07014388 \times 100\%$, sd $= 0.05950948$



(e) Noise-to-signal ratio, mean $= 1.87364$, sd $= 0.1876208$.

Figure 8: Monte-Carlo simulation of 1000 repetitions with short range dependent errors with discontinuous γ. Sample size reduced from 1024 to 256.

$$\begin{aligned}
\mathbb{E}_\gamma[(\hat{\theta}_j(\mathbf{F}) - \theta_j)^2] &= \mathbb{E}_0[\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}(\mathbf{F})(\hat{\theta}_j(\mathbf{F}) - \theta_j)^2] \\
&= \mathbb{E}_0[\mathbb{E}_0[(\hat{\theta}_j(\mathbf{F}) - \theta_j)^2 | X_1, \cdots, X_n]\frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}] \\
&\geq \mathbb{E}_0[(\bar{\theta}_j(X_1, \cdots, X_n) - \theta_j)^2 \frac{d\mathbf{P}_\gamma}{d\mathbf{P}_0}] \\
&= \mathbb{E}_\gamma[(\bar{\theta}_j(X_1, \cdots, X_n) - \theta_j)^2],
\end{aligned}$$

where $\bar{\theta}_j(X_1, \cdots, X_n) = \mathbb{E}_0[\hat{\theta}_j(\mathbf{F})|X_1, \cdots, X_n]$. Therefore Theorem 1.1 and Theorem 1.2 in Donoho (1995a) imply Theorem 4.3.

The Lévy-Ciesielski-Itô construction of Brownian motion: This construction was used in passing from the filtering model to the Gaussian sequence model in Section 4.1 and generalized to fractional Brownian motion en route to proving Theorem 4.4. Let $\epsilon'_j \sim \mathcal{N}(0,1)$ be i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{H}$ be the Hilbert subspace of $L^2(\Omega, \mathcal{F}, P)$ generated by $\{\epsilon'_j\}$. Define an Itô isometry by

$$\epsilon'_j \in \mathcal{H} \overset{\Psi}{\mapsto} \psi_j \in L^2[0,1].$$

The stochastic process $t \mapsto \Psi(1_{[0,t]})$ is a standard Brownian motion by the following properties:

1. The increment $\Psi(1_{[0,t]}) - \Psi(1_{[0,s]}) = \Psi(1_{[s,t]})$ for any $0 \leq s < t \leq 1$ is distributed $\mathcal{N}(0, t-s)$, being the mean square limit of normal random variables.

2. Two increments $\Psi(1_{[s,t]})$ and $\Psi(1_{[s',t']})$ are uncorrelated, therefore independent by normality.

Definition of fractional Brownian motion: We recall here the precise definition of the fractional Brownian motion, which is the limit process for the Functional Central Limit Theorem used in Section 4.2. The fractional Brownian motion with Hurst exponent H $B_H(t), t \in \mathbb{R}$, is a zero-mean Gaussian process with covariance function

$$\gamma(s,t) = \frac{V_H}{2}(|s|^{2H} + |t|^{2H} - |t-s|^{2H})$$

where

$$V_H = var(B_H(1)) = \frac{-\Gamma(2-2H)\cos(\pi H)}{\pi H(2H-1)},$$

and $\Gamma(\cdot)$ is the gamma function.

Proof of Theorem 4.4: Fix a wavelet basis $\{\psi_{jk}\}$ that is continuously differentiable up to $r > H + \frac{3}{2}$ times. Let $\Delta = \frac{d^2}{dx^2}$ be the Laplace operator on $[0,1]$. For Hurst exponent $H \in (\frac{1}{2}, 1)$, the operator $K_H = (-\Delta)^{H+\frac{1}{2}}$ is the reproducing kernel of the reproducing kernel Hilbert space of dB_t^H.[7] The functions $K_H^{-\frac{1}{2}}\psi_{jk}$ diagonalizes K_H, which gives a Karhunen-

[7]The reproducing kernel Hilbert space of dB_t^H consists of $f \in L^2[0,1]$ for which

Loéve decomposition of fractional Brownian motion

$$B_t^H = \sum_{jk} w_{jk} K_H^{-\frac{1}{2}} \psi_{jk}$$

where $\{w_{jk}\}$ is a Gaussian white noise. Define

$$v_{jk} = \frac{d}{dt} K_H^{-\frac{1}{2}} \psi_{jk} = (-\Delta)^{\frac{1}{4}-\frac{H}{2}} \psi_{jk}.$$

Then we have a representation:

$$dB_t^H = \sum_{jk} w_{jk} v_{jk}(t) dt.$$

The random variables e_{jk} in the statement of the theorem can be expressed by

$$e_{jk} = \frac{1}{\sigma_j} \int \psi_{jk} dB_t^H = \frac{1}{\sigma_j} \sum_{j'k'} w_{j'k'} \int \psi_{jk} v_{j'k'}(t) dt.$$

The near-independence property of e_{jk} can now be shown using the time-scale localization property of wavelets: By normality, the conditional mean $\widehat{e}_{jk} = \mathbb{E}[e_{jk} | e_{j'k'},\ (j',k') \neq (j,k)]$ must lie in the l^2-span of $\{e_{j'k'},\ (j',k') \neq (j,k)\}$: $\widehat{e}_{jk} = \sum_{(j',k') \neq (j,k)} a_{j'k'} e_{j'k'}$. So

$$\begin{aligned} e_{jk} - \widehat{e}_{jk} &= -\sum_{j',k'} [\int v_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''} \psi_{j''k''} dt] w_{j'k'} && (3) \\ &= -\sum_{j',k'} [\int \psi_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''} v_{j''k''} dt] w_{j'k'}, && (4) \end{aligned}$$

where the second equality follows from the symmetry of the operator $(-\Delta)^{\frac{1}{4}-\frac{H}{2}}$. Using the fact that $\{\psi_{jk}\}$ is an orthonormal basis,

---

$$\int_0^1 K_H f(t) f(t) dt < \infty.$$

For standard Brownian motion, the case $H = \frac{1}{2}$, this space is the Cameron-Martin space of Brownian motion: the Sobolev space of absolutely continuous functions f with $f(0) = 0$ and $f' \in L^2[0,1]$.

$$
\begin{aligned}
\mathrm{Var}(e_{jk} - \widehat{e}_{jk}) &= \sum_{j',k'} [\int \psi_{j'k'}(t) \sum_{j'',k''} \frac{1}{\sigma_{j''}} a_{j''k''} v_{j''k''} dt]^2 && (5)\\
&= \| \sum_{j'',k''} a_{j''k''} \cdot \frac{1}{\sigma_{j''}} v_{j''k''} \|_{L^2}^2 && (6)\\
&= c_0 \| \sum_{j'',k''} a_{j''k''}^2 \| && (7)\\
&> 0. && (8)
\end{aligned}
$$

This proves Theorem 4.4.

Proof of Theorem 4.5: Let $f^{(j)}$ be the L^2-projection of f onto the j-th resolution detail subspace, and $R_n^{(j)}(p, q, \zeta, L)$ be the minimax risk of the corresponding Gaussian sequence model with independent noise. By Parseval's equality

$$
\mathbb{E}[\|\hat{f}_n - f\|^2] = \sum_j \mathbb{E}[\|\hat{f}_n^{(j)} - f^{(j)}\|^2],
$$

and

$$
R_n(p, q, \zeta, L) = \sum_j R_n^{(j)}(p, q, \zeta, L).
$$

Therefore Theorem 4.5 follows immediately from Theorem 4.3.

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61**(4), 821–856.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, **20**(3), 651–707.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**(1), 47–78.

Besov, O. V., Ilin, V. P., and Nikolskiĭ, S. M. (1978). *Integral Representations of Functions and Embedding Theorems*. Halsted Press.

Bierens, H. J. (1987). Kernel estimators of regression functions. In *Advances in Econometrics: Fifth World Congress, Volume 1.*

Bodurtha, J. N. and Mark, N. C. (1991). Testing the CAPM with time-varying risks and returns. *Journal of Finance*, **46**(4), 1485–1505.

Brown, J. P., Song, H., and McGillivray, A. (1997). Forecasting UK house prices: A time varying coefficient approach. *Economic Modelling*, **14**(4), 529–548.

Brown, L. D., Low, M. G., *et al.* (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, **24**(6), 2384–2398.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, **136**(1), 163–188.

Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, **148**(2), 101–113.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, **80**(3), 1157–1183.

Chen, S.-N. and Keown, A. J. (1981). Risk decomposition and portfolio diversification when beta is nonstationary: a note. *Journal of Finance*, **36**(4), 941–947.

Ciesielski, Z. (1961). Hölder conditions for realizations of Gaussian processes. *Transactions of the American Mathematical Society*, **99**(3), 403–413.

Cochrane, J. H. (1999). New facts in finance. Technical report, National Bureau of Economic Research.

Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, **1**(1), 54–81.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, **106**(1), 157–181.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, **41**(7), 909–996.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

De Jong, R. M. and Davidson, J. (2000). The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory*, **16**(05), 621–642.

Donoho, D. L. (1995a). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, **41**(3), 613–627.

Donoho, D. L. (1995b). Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and Computational Harmonic Analysis*, **2**(2), 101–126.

Donoho, D. L., Johnstone, I. M., *et al.* (1998). Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, **26**(3), 879–921.

Evans, M. D. (1994). Expected returns, time-varying risk, and risk premia. *Journal of Finance*, **49**(2), 655–679.

Fan, Y. and Gençay, R. (2010). Unit root tests with wavelets. *Econometric Theory*, **26**(05), 1305–1331.

Gelfand, I. M. and Vilenkin, N. Y. (1964). *Generalized functions*. Academic press.

Gençay, R. and Signori, D. (2015). Multi-scale tests for serial correlation. *Journal of Econometrics*, **184**(1), 62–80.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *Journal of Finance*, **53**(2), 549–573.

Girsanov, I. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, **5**(3), 285–301.

Guirguis, H. S., Giannikos, C. I., and Anderson, R. I. (2005). The US housing market: Asset pricing forecasts using time varying coefficients. *The Journal of Real Estate Finance and Economics*, **30**(1), 33–53.

Herren, V. (1997). Lévy-type processes and Besov spaces. *Potential Analysis*, **7**(3), 689–704.

Hong, Y. and Kao, C. (2004). Wavelet-based testing for serial correlation of unknown form in panel models. *Econometrica*, **72**(5), 1519–1563.

Itô, K., Nisio, M., *et al.* (1968). On the convergence of sums of independent Banach space valued random variables. *Osaka Journal of Mathematics*, **5**(1), 35–48.

Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*. Cambridge University Press.

Kilian, L. and Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, **60**(1), 85–107.

Kristensen, D. (2012). Non-parametric detection and estimation of structural change. *The Econometrics Journal*, **15**(3), 420–461.

Lee, J. and Hong, Y. (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory*, **17**(02), 386–423.

Lévy, P. (1954). Le mouvement brownien. *Mémorial des sciences mathématiques*, **126**, 1–84.

Lucas Jr, R. E. (1976). Econometric policy evaluation: A critique. In *Carnegie-Rochester Conference Series on Public Policy*. Elsevier.

Meyer, Y. (1995). *Wavelets and Operators*. Cambridge University Press.

Nellis, J. G. and Longbottom, J. A. (1981). An empirical analysis of the determination of house prices in the United Kingdom. *Urban Studies*, **18**(1), 9–21.

Perron, P. and Yamamoto, Y. (2014). A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory*, **30**(02), 491–507.

Pinsker, M. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Prob. Info. Transmission*, **16**(2), 120–133.

Rapach, D. E. and Wohar, M. E. (2005). Valuation ratios and long-horizon stock price predictability. *Journal of Applied Econometrics*, **20**(3), 327–344.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, **56**(4), 931–954.

Taqqu, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probability Theory and Related Fields*, **31**(4), 287–302.

Triebel, H. (1978). *Interpolation Theory, Function spaces, Differential Operators*. North Holland.