

Trend Filtering with Fractional Splines

Tommaso Proietti

University of Rome Tor Vergata

Preliminary version: October 5, 2026

Abstract

The paper introduces a novel signal plus noise model, where the signal is a fractional spline (FS) process. The latter is generated by a linear combination of shocks, where the impulse response function follows a generalized harmonic sequence, governed by the memory parameter. The fractional spline process encompasses standard trend models, such as the random walk and the integrated random walk, while for fractional memory parameter it provides an alternative to standard fractional noise. We discuss two specifications, the fixed effects model, which generalizes trend filtering to the fractional case, and the random effects specification, and develop estimation and signal extraction by penalized least squares with ℓ_1 and ℓ_2 penalties in the former case, and by maximum likelihood estimation and conditional mean inference for the latter. We then characterize the time series properties of the fractional noise process in terms of memory, and, in the stationarity case, we derive the autocovariance function and spectral density, and compare them to those of the fractional noise process. The practical relevance of the FS framework is demonstrated through applications to inflation dynamics and realized volatility, illustrating its forecast performance and its ability to capture complex long-memory patterns in economic data, as well as persistent trends in tree-ring based temperature reconstructions.

Keywords: Long memory, Persistence, Prediction, Signal Extraction.

JEL codes: C14, C22, Q54.

1 Introduction

The signal plus noise model for a the univariate time series y_t , $y_t = \mu_t + \epsilon_t$, $t = 1, 2, \dots, T$, where μ_t is the persistent, or trend, component and ϵ_t is purely random and orthogonal to μ_t , provides a standard framework for signal extraction in time series analysis. Traditionally, the component μ_t has been represented by linear finite dimensional Markovian processes, either stationary or nonstationary. Canonical examples are the local level and local linear trend models. An early example is the random walk plus noise model of Muth (1960). Statistical inference for these models is carried out with the support of state space methods, such as the Kalman filter and related smoothing algorithms (Durbin and Koopman, 2012).

This literature has strong intersections with the semiparametric one on smoothing splines, see Ruppert et al. (2003), as demonstrated by Wecker and Ansley (1983), see also Proietti (2007). More recent research has considered trend extraction by penalized least squares methods with ℓ_1 penalty, see Kim et al. (2009) and Tibshirani (2014).

In all these approaches, the dynamic evolution of the trend process is specified as an integrated process of order d , denoted $\mu_t \sim I(d)$, with integer order, $d \in \mathbb{Z}$ (typically $d = 0, 1, 2$), meaning that the signal results from the d -fold cumulative sum (integration) of random shocks (in the case $d = 0$, μ_t , is modelled as a stationary autoregressive process).

The long memory literature relaxes this restriction by modelling μ_t as a fractional noise or fractionally integrated process, introduced by Granger and Joyeux (1980) and Hosking (1981), see the monographs by Palma (2007), Giraitis et al. (2012), Beran et al. (2013) and Hassler (2019) for a comprehensive treatment and further references; long memory models have received considerable attention in several disciplines, including econometrics, see, e.g., Baillie (1996), Robinson (2003), finance (Hurvich et al., 2005), and climatology, see Hartl (2025) and Proietti and Maddanu (2024).

In this paper we propose a novel specification for the signal, the fractional spline (FS) process, according to which the level of the process is subject to structural change due to shocks propagated with impulse response function represented by a generalized harmonic sequence. The resulting signal is a persistent component generated by a sequence of local interventions whose effects decay according to a power law.

When the fractional spline plus noise process is interpreted as a saturated regression model, inference is carried out by shrinking the contribution of individual changepoint to zero, according to a penalty function. Thence, the set of basis functions is a truncated power polynomial basis with continuous power parameter model is an extension of regression splines and trend filtering with ℓ_1 metric (Kim et al., 2009; Tibshirani, 2014).

The alternative assumption that the shocks are a sequence of independently and identically distributed (i.i.d.) random variables, with a known distribution, opens the way for likelihood-based inferences. The fractional spline process can be viewed as a discretized version of a generalized the Holmgren–Riemann–Liouville, or Type II fractional Brownian motion process discussed by Marinucci and Robinson (1999). The model extends the class of harmonically weighted processes considered by Hassler and Hosseinkouchack (2020) and is related to error duration by Parke (1999) and switching regimes models by Davidson and Sibbertsen (2005), which in turn have a continuous time precursor in the rainfall model by Rodriguez-Iturbe et al. (1987).

This paper contributes to the literature on trend filtering and long memory time series by (i) proposing a novel model of persistent dynamics; (ii) extending trend filtering to the $I(d)$ case, where d is a fractional parameter; (iii) characterizing the time series properties of the fractional spline process in terms of memory, stationarity, autocorrelation and spectral density; finally (iv) deriving a new aggregation result on the genesis of long memory.

After presenting and discussing the fractional spline plus noise model (Section 2), the paper develops the estimation and signal extraction theory for the fixed effects specification (Section 3)

and the random effects specification (Section 4). In the latter case, we illustrate the memory properties of the fractional spline process, by considering the rate at which the variance of the partial sums increases with time (Section 5.1). Secondly, assuming that the process started in the indefinite past, the FS model can be given a stationary distribution for values of the integration parameter in the range $0 < d < 0.5$. We derive the autocovariance function and the spectral density and compare it to that of the fractional noise process (Section 5.2). We then show that the process arises from the aggregation of heterogeneous AR(1) processes with AR parameter drawn independently from a log-gamma distribution (Section 5.3).

The practical relevance of the FS framework is demonstrated through applications to inflation dynamics and realized volatility, illustrating its forecast performance and its ability to capture complex long-memory patterns in economic data, as well as persistent trends in tree-ring based temperature reconstructions (Section 6.3).

2 Fractional Spline plus Noise Model

Let $y_t = \mu_t + \epsilon_t$, $t = 1, \dots, T$, with $\epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2)$, where WN denotes white noise, and

$$\begin{aligned}\mu_t &= \mu_1 + \sum_{j=2}^t \eta_j (t-j+1)_+^{d-1} \\ &= \mu_1 + \eta_t + 2^{d-1}\eta_{t-1} + 3^{d-1}\eta_{t-2} + \dots + (t-1)^{d-1}\eta_2,\end{aligned}\tag{1}$$

where for $d \in \mathbb{R}^+$ is the memory, or persistence, parameter, and x_+^{d-1} denotes the $d-1$ fractional truncated power function

$$x_+^{d-1} = \begin{cases} 0, & x \leq 0, \\ x^{d-1}, & x > 0. \end{cases}$$

The coefficients η_j , $j = 1, \dots, T$ represent “news” or shocks, occurring at time j , affecting the level at time $t \geq j$ with impact $(t-j+1)^{d-1}$, depending on the time distance from time t . When $d = 1$, μ_t is a random walk starting at time $t = 1$ with initial value μ_1 , and the value at time t results from adding the cumulation of the news η_s , $s = 2, \dots, t$. This can be also be viewed as a regression spline of degree zero with knots located at $t = 2, \dots, T$. More generally, for $d \in \mathbb{Z}^+$ we have a polynomial spline of degree $d-1$ in discrete time: $d = 2$ corresponds to a linear spline, $d = 4$ to a cubic spline.

For $t = 2, \dots, T$ the fractional local level (1) has the equivalent time series representation:

$$\mu_t = \mu_1 + \sum_{k=0}^{t-2} \psi_k \eta_{t-k}, \quad \psi_k = (k+1)^{d-1}.\tag{2}$$

where ψ_k represents the response of μ_t , and thus y_t , to the news η_{t-k} . Figure 1 shows the intervention signature $\{\psi_k = (k+1)^{d-1}, k = 0, 1, \dots\}$, also referred to as the impulse response function (IRF), of a shock occurring at time t after $k = 0, \dots, 15$ time units, for different values of the memory parameter d . It highlights that for $d > 1$, ψ_k is an increasing function of the lag, while for $d < 1$ the effect of past shocks decreases at an hyperbolic rate with the lag k and the process is mean reverting in the sense that $\lim_{k \rightarrow \infty} \psi_k = 0$, that is, in the absence of further news, the process will revert to μ_1 .

The Fractional Spline plus Noise (FSpN) model can be represented as a regression model as follows. Denoting $\mathbf{y} = [y_1, \dots, y_T]'$, $\boldsymbol{\mu} = [\mu_1, \dots, \mu_T]'$, $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_T]'$, $\boldsymbol{\eta} = [\eta_2, \dots, \eta_T]'$, $\mathbf{i} = [1, \dots, 1]'$, we can write $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, with $\boldsymbol{\mu} = \mu_1 \mathbf{i} + \mathbf{X}_d \boldsymbol{\eta}$, $\mathbf{X}_d = \begin{bmatrix} \mathbf{0}' \\ \mathbf{C}_d \end{bmatrix}$, where the matrix $\mathbf{C}_d$ is lower triangular with unit diagonal elements, and contains in its column the intervention

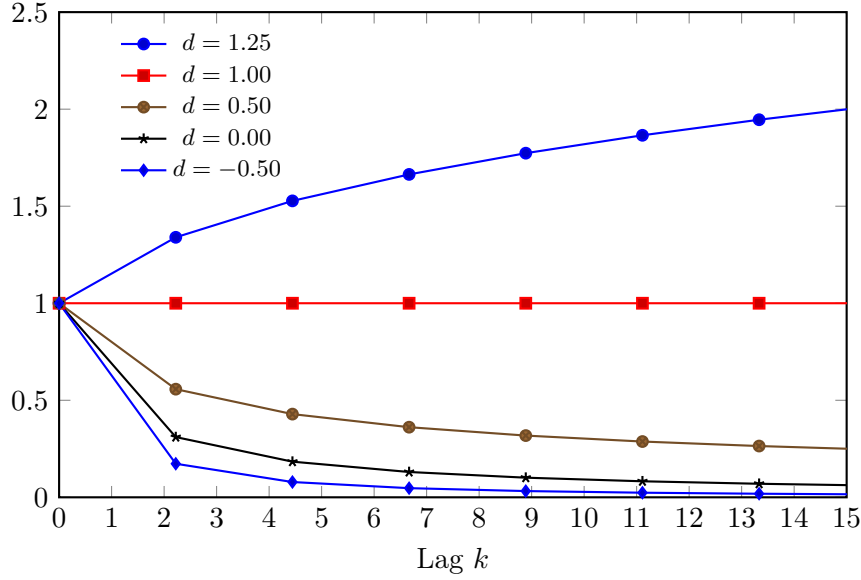


Figure 1: Intervention signature of a shock occurring at time t after k time units, $\partial\mu_{t+k}/\partial\eta_t = (k+1)^{d-1}$, for different values of the memory parameter d .

signature of the news occurring at time $j = 2, \dots, T$. Also, defining the $T \times 1$ vector $\beta = [\mu_1, \boldsymbol{\eta}']'$, and the $T \times T$ matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 2^{d-1} & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \cdots & \ddots & \ddots & \vdots & 0 & 0 \\ 1 & (T-3)^{d-1} & \cdots & \ddots & 1 & 0 & 0 \\ 1 & (T-2)^{d-1} & \cdots & \cdots & 2^{d-1} & 1 & 0 \\ 1 & (T-1)^{d-1} & \cdots & \cdots & 3^{d-1} & 2^{d-1} & 1 \end{bmatrix} = [\mathbf{i} \quad \mathbf{X}_d],$$

we can write $\mathbf{y} = \mathbf{X}\beta + \epsilon$. In the next sections we are going to discuss inferences on μ_1 and β .

3 Fixed effects specification

The parameters of the FSpN model, μ_1 and $\eta_j, j = 2, \dots, T$, are considered here as fixed regression effects. Estimation of the news η_j poses a high-dimensional inference problem, which can be dealt with by imposing a penalty on their size. The model features T parameters as many as there are observations, but we regularize the effects $\eta_j, j = 2, \dots, T$, shrinking them towards 0 according to a penalty function.

Defining the $T \times T$ matrix $\mathbf{V} = \text{diag}(0, \mathbf{I}_{T-1})$, we estimate the regression model $\mathbf{y} = \mathbf{X}\beta + \epsilon$ under the constraint $\beta' \mathbf{V} \beta = \boldsymbol{\eta}' \boldsymbol{\eta} < c$, for a positive constant c , or equivalently focus on the penalized least squares (PLS) criterion with ℓ_2 penalty:

$$S(d, \lambda) = (\mathbf{y} - \mathbf{i}\mu_1 - \mathbf{X}_d \boldsymbol{\eta})'(\mathbf{y} - \mathbf{i}\mu_1 - \mathbf{X}_d \boldsymbol{\eta}) + \lambda \boldsymbol{\eta}' \boldsymbol{\eta}. \quad (3)$$

The parameter μ_1 can be concentrated out of the PLS function; the first order conditions for

the minimiser of (3) with respect to this parameter yield

$$\hat{\mu}_1 = \frac{1}{T} \mathbf{i}'(\mathbf{y} - \mathbf{X}_d \hat{\boldsymbol{\eta}}). \quad (4)$$

Introducing the centring matrix $\mathbf{M} = \mathbf{I}_T - \mathbf{i}\mathbf{i}'/T$ and denoting $\mathbf{z} = \mathbf{M}\mathbf{y}$ and $\mathbf{U}_d = \mathbf{M}\mathbf{X}_d$, the profile PLS is written

$$S_{\mu_1}(d, \lambda) = (\mathbf{z} - \mathbf{U}_d \boldsymbol{\eta})'(\mathbf{z} - \mathbf{U}_d \boldsymbol{\eta}) + \lambda \boldsymbol{\eta}' \boldsymbol{\eta}. \quad (5)$$

Minimizing (5) with respect to $\boldsymbol{\eta}$ yields the *ridge* estimator

$$\begin{aligned} \hat{\boldsymbol{\eta}} &= (\mathbf{U}_d' \mathbf{U}_d + \lambda \mathbf{I})^{-1} \mathbf{U}_d' \mathbf{z} \\ &= (\mathbf{X}_d' \mathbf{M} \mathbf{X}_d + \lambda \mathbf{I})^{-1} \mathbf{X}_d' \mathbf{M} \mathbf{y}, \end{aligned} \quad (6)$$

where $\mathbf{I}$ is the identity matrix of order $T - 1$. Replacing into (4) gives

$$\hat{\mu}_1 = \frac{1}{T} \mathbf{i}'(\mathbf{I} - \mathbf{X}_d(\mathbf{X}_d' \mathbf{M} \mathbf{X}_d + \lambda \mathbf{I})^{-1} \mathbf{X}_d' \mathbf{M}) \mathbf{y}. \quad (7)$$

The PLS estimator of the fractional level is then

$$\hat{\boldsymbol{\mu}} = \mathbf{W}_\mu \mathbf{y}, \quad \mathbf{W}_\mu = \frac{\mathbf{i}\mathbf{i}'}{T} + \mathbf{U}_d(\mathbf{U}_d' \mathbf{U}_d + \lambda \mathbf{I})^{-1} \mathbf{U}_d';$$

this is a linear transformation of the observed vector with filter weights given by $\mathbf{W}_\mu$.

3.1 Special well known cases and extensions

When $d \in \mathbb{Z}^+$ the model can be represented in the *analysis* framework

$$\sum_t (y_t - \mu_t)^2 + \lambda \sum_{t=d+1}^T (\Delta^d \mu_t)^2$$

see Tibshirani (2014), section 7.2 for a discussion of the analysis and synthesis representation. For $d = 1$ the weight matrix $\mathbf{W}_\mu$ provides the weights of the two sided exponential smoothing filter.

In the presence of a systematic drift the fractional local linear trend model the model is amended as $y_t = \mu_t + \epsilon_t$, where

$$\mu_t = \mu_1 + \delta t + \sum_{j=0}^{t-3} (j+1)^{d-1} \eta_{t-j}.$$

For $d = 2$, $\mu_t - 2\mu_{t-1} + \mu_{t-2} = \eta_t$, and the minimiser with respect to the trend μ_t with a fixed λ (e.g. $\lambda = 1600$) is known as the Hodrick and Prescott (1997) trend.

To prove these equivalences, we first notice that when $d = 1$, $\mathbf{X}$ is a $T \times T$ lower triangular matrix with unit elements, and $\mathbf{X}^{-1}$ has 1 on the main diagonal and -1 in the lower subdiagonal, while all other elements equal to zero. As a results, $\mathbf{X}^{-1'} \mathbf{V} \mathbf{X}^{-1} = \mathbf{D}' \mathbf{D}$, where $\mathbf{D}$ is a $T - 1 \times T$ matrix with elements $D_{tt} = -1$ and $D_{t,t+1} = 1$, $t = 1, \dots, T - 1$, such that $\mathbf{D}\boldsymbol{\mu}$ is a vector with elements $\{\mu_{t+1} - \mu_t\}$. Hence, the ridge predictor is

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \mathbf{X}(\mathbf{X}' \mathbf{X} + \lambda \mathbf{V})^{-1} \mathbf{X} \mathbf{y} \\ &= (\mathbf{I} + \lambda \mathbf{X}^{-1'} \mathbf{V} \mathbf{X}^{-1})^{-1} \mathbf{y} \\ &= (\mathbf{I} + \lambda \mathbf{D}' \mathbf{D})^{-1} \mathbf{y}, \end{aligned}$$

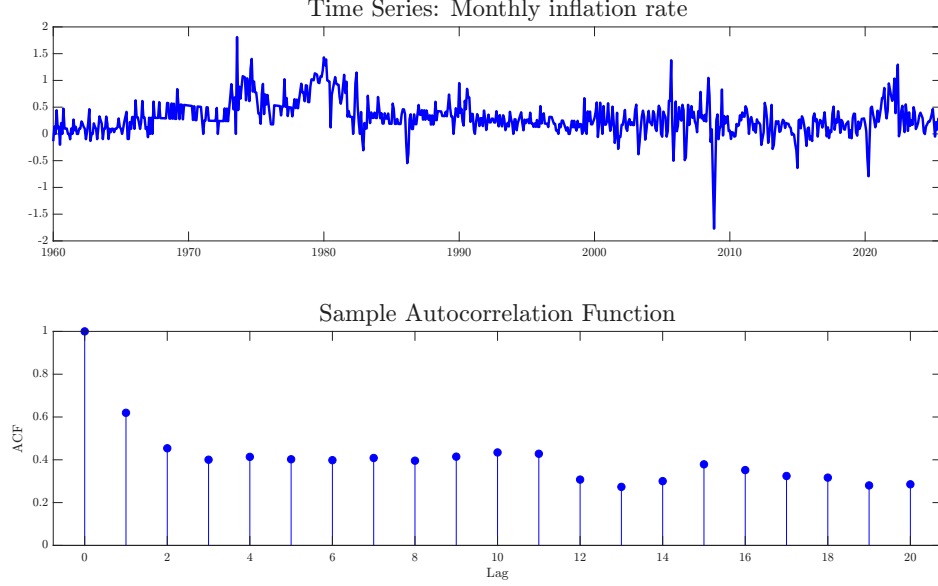


Figure 2: Monthly U.S. Inflation Rate (January 1960 - September 2025): time series (upper plot) and sample autocorrelation function.

where $\hat{\beta}$ is the minimizer of the PLS criterion $S(d, \lambda) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta'\mathbf{V}\beta$.

When $d = 2$, $\mathbf{V} = \text{diag}(\mathbf{0}_2, \mathbf{I}_{T-2})$, where $\mathbf{0}_2$ is a 2×2 matrix of zeroes and $\mathbf{X}^{-1'}\mathbf{V}\mathbf{X}^{-1})^{-1}$ equals $\mathbf{D}^*\mathbf{D}^*$, where $\mathbf{D}^*$ is a $T - 2 \times T$ matrix with elements $D_{tt}^* = 1$, $D_{t,t+1}^* = -2$, and $D_{t,t+2}^* = 1$, $t = 1, \dots, T - 2$, such that $\mathbf{D}^*\mu$ is a vector with elements $\{\mu_{t+2} - 2\mu_{t+1} + \mu_t\}$. Hence, $\hat{\mu} = (\mathbf{I} + \lambda\mathbf{D}^{*'}\mathbf{D}^*)^{-1}\mathbf{y}$, which is the HP trend estimate.

The ℓ_1 trend filtering approach was introduced by Kim et al. (2009), as a variant of the Hodrick and Prescott (1997) filter replacing the sum of squares of the second differences with the sum of their absolute values. It has been extended by Tibshirani (2014) to $d \in \mathbb{Z}^+$. In our framework we can generalize ℓ_1 trend filtering to $d \in \mathbb{R}$ by estimating the signal $\mu = \mathbf{i}\mu_1 + \mathbf{X}_d\eta$ as the minimizer of the penalized least squares criterion

$$S(d, \lambda) = (\mathbf{y} - \mathbf{i}\mu_1 - \mathbf{X}_d\eta)'(\mathbf{y} - \mathbf{i}\mu_1 - \mathbf{X}_d\eta) + \lambda \sum_{j=2}^T |\eta_j|.$$

The ℓ_1 estimator of the fractional spline trend is a lasso type estimator, such that the sparsity is a function of the penalty parameter. An elastic net penalty can alternatively be applied.

3.2 Illustration: ℓ_1 filtering for US Monthly Inflation

Figure 2 displays the U.S. monthly inflation rate, obtained as the logarithmic difference of the Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, for the period from January 1960 to December 2025, for a total of $T = 792$ observations. The original series is downloadable from the FRED database at the Federal Reserve Bank of St Louis (<https://fred.stlouisfed.org/series/CPIAUCSL>). The bottom plot is the sample autocorrelation function up to 20 lags.

ℓ_1 trend filtering can be valuable for identifying major change point: level shifts for $d = 1$ and slope change for $d = 2$. In the fractional case it serves to estimate the size of the shocks assuming a level of sparsity determined by the value of the penalty parameter λ . For all values

of d and λ sufficiently large, we obtain $y_t = \mu_1$, sometimes referred to as the null model, while for $\lambda = 0$, $y_t = \mu_t$.

Figure 3 plots the estimates of the fractional spline (FS, top plot) and of η_t (bottom plot) for $d = \{0.0, 0.5, 1.0\}$ and $\lambda = 0.03$. Figure 4 plots the estimates of the same quantities for $d = \{0.0, 0.5, 1.0\}$ and a value of the penalty parameter closer to zero, $\lambda = 0.005$.

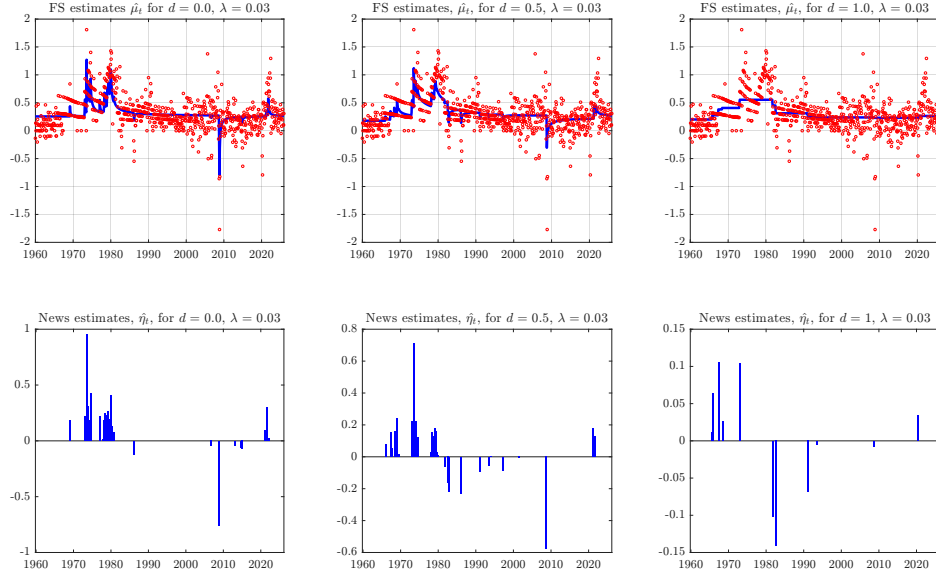


Figure 3: ℓ_1 Trend Filtering for U.S. monthly inflation, for $d = \{0.0, 0.5, 1.0\}$ and $\lambda = 0.03$. Plot of the PLS estimates of the fractional spline (top) and news (bottom).

Estimation of d and λ is carried out by grid search using time series cross-validation to estimate the test error. A grid consisting of 41 equally spaced values for d in the range from 0 to 1 with step 0.025, and 50 equally spaced values from -4.5 and -1.0 for $\log_{10} \lambda$ were considered, (hence, $\lambda \in (0, 0.1)$). Recursive training samples for $t = 1, \dots, T_0 + i$, $T_0 = 480$, $i = 0, 1, \dots, 311$ are considered: starting from $i = 0$, for each parameter combination (d, λ) , the ℓ_1 penalized least squares estimates of β are obtained by lasso, and the forecasts $\hat{y}_{T_0+i+h|T_0+i}$, $h = 1, \dots, 12$ are computed. At the end of the sample, the recursive forecasting exercise makes available $T - T_0 - h$ h -step-ahead prediction errors, which are used to estimate the mean square forecast error at horizon h , according to

$$\text{MSFE}(h) = \frac{1}{T - T_0 - h} \sum_{j=1}^{T-T_0-h} (y_{T_0+i+h} - \hat{y}_{T_0+i+h|T_0+i})^2.$$

The selected pair which minimizes $\text{MSFE}(h)$ can vary with the horizon.

The optimal combination for $h = 1$ is $d = 0$ and $\lambda = 0.0003$. Hence, for predicting one-step-ahead the estimated model features frequent small shocks with mean reverting effects. However, at longer forecast horizons, $h \geq 2$, crossvalidation select $\hat{d} = 1, \lambda = 0.0373$, which is a solution with large infrequent shocks. Table 1 displays for each forecast horizon h , the optimal combination of (d, λ) that minimizes the test sample MSFE.

Later (Section 6.1) we are going to compare the forecasting performance with that of the random effects specification. Interestingly, the results for $h \geq 1$ are consistent with the idea that apparent long memory originates from occasional level shifts (Diebold and Inoue, 2001).

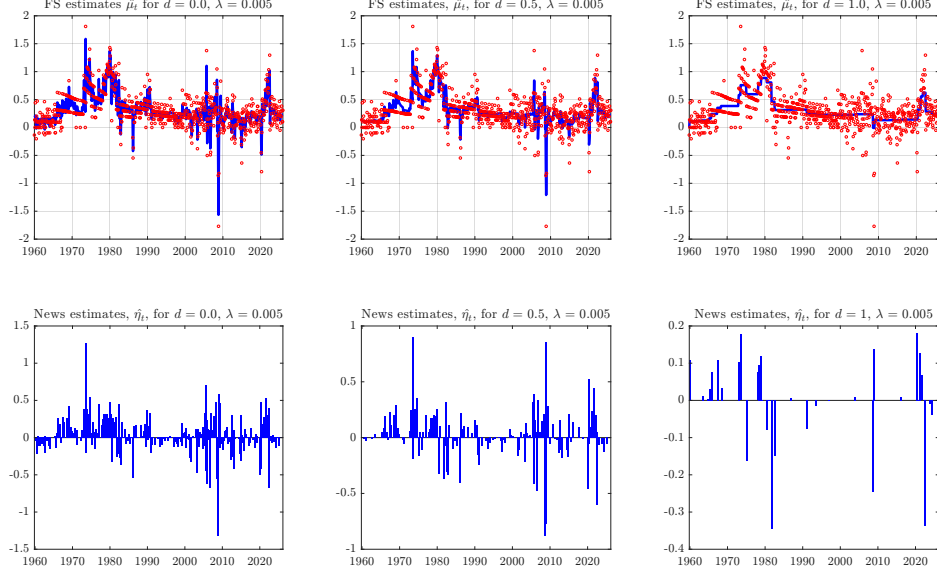


Figure 4: ℓ_1 Trend Filtering for U.S. monthly inflation, for $d = \{0.0, 0.5, 1.0\}$ and $\lambda = 0.005$. Plot of the PLS estimates of the fractional spline (top) and news (bottom).

The outcome is also related to the stochastic regimes model by Davidson and Sibbertsen (2005), where the regimes are delimited by changepoints and the duration of the regimes has probability function with tails following a power law. They consider shocks η_j occurring at random times t_j , $j = 1, \dots, K$, $1 < t_1 < t_2 < \dots < t_K \leq T$. The model is $\mu_t = \mu_1 + \sum_{j=1}^m I(t_{j-1} < t \leq t_j) \eta_j$. The durations $\tau_j = t_j - t_{j-1}$ have an invariant probability function with tails following a power law, $P(\tau_j = k) = O(k^{-\delta})$, where $\delta \in (2, 3)$.

4 Random effects specification

In the previous section we assumed that $\boldsymbol{\eta}$ is fixed. We continue to assume that μ_1 is fixed, but η_j will be assumed to be i.i.d. with Gaussian distribution centred around zero and variance σ_η^2 , $\eta_j \sim \text{i.i.d. } N(0, \sigma_\eta^2)$, while $\epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2)$, with $E(\epsilon_t \eta_j) = 0, \forall t, j$. Further, we denote the signal to noise ratio by $q_\eta = \sigma_\eta^2 / \sigma_\epsilon^2$ and define the matrix $\mathbf{Q} = \mathbf{I}_T + q_\eta \mathbf{X}_d \mathbf{X}_d'$.

Letting $\mathbf{y} = (y_1, \dots, y_t, \dots, y_T)'$, the random effects specification implies that $\mathbf{y} \sim N(\mu_1 \mathbf{i}, \boldsymbol{\Sigma}_y)$, $\boldsymbol{\Sigma}_y = \sigma_\epsilon^2 \mathbf{Q}$. Apart from a constant term, the log-likelihood is

$$\ell(d, \mu_1, q_\eta, \sigma_\epsilon^2; \mathbf{y}) = -\frac{1}{2} \left[T \ln \sigma_\epsilon^2 + \ln |\mathbf{Q}| + \frac{1}{\sigma_\epsilon^2} (\mathbf{y} - \mu_1 \mathbf{i})' \mathbf{Q}^{-1} (\mathbf{y} - \mu_1 \mathbf{i}) \right].$$

The parameters μ_1 and σ_ϵ^2 can be concentrated out of the likelihood function, giving

$$\tilde{\mu}_1 = \frac{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{y}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}}, \quad \tilde{\sigma}_\epsilon^2 = \frac{1}{T} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i})' \mathbf{Q}^{-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i}). \quad (8)$$

The profile likelihood,

$$\ell_{\mu_1, \sigma_\epsilon^2}(d, q_\eta; \mathbf{y}) = -\frac{1}{2} [T(\ln \tilde{\sigma}_\epsilon^2 + 1) + \ln |\mathbf{Q}|], \quad (9)$$

Table 1: Cross-validation estimates of the memory and ℓ_1 penalty parameter λ for different forecast horizons.

h	$\hat{d}$	$\hat{\lambda}$	MSFE(h)	$\sum_{i=2}^T I(\hat{\eta}_i > 0)$
1	0.0	0.0003	0.0905	749
2	1.0	0.0373	0.0983	8
3	1.0	0.0373	0.0987	8
4	1.0	0.0373	0.0990	8
5	1.0	0.0373	0.0989	8
6	1.0	0.0373	0.0992	8
7	1.0	0.0373	0.0995	8
8	1.0	0.0373	0.1000	8
9	1.0	0.0373	0.1000	8
10	1.0	0.0373	0.1004	8
11	1.0	0.0373	0.1008	8
12	1.0	0.0373	0.1012	8

can be maximised with respect to q_η and d .

Observing that $\text{Cov}(\boldsymbol{\eta}, \mathbf{y}) = \sigma_\eta^2 \mathbf{X}'_d$, the minimum mean square error estimator of $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\eta}} = E(\boldsymbol{\eta}|\mathbf{y})$, is

$$\begin{aligned}\tilde{\boldsymbol{\eta}} &= q_\eta \mathbf{X}'_d \mathbf{Q}^{-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i}) \\ &= \mathbf{W}_\eta \mathbf{y}, \quad \mathbf{W}_\eta = q_\eta \mathbf{X}'_d \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right).\end{aligned}\quad (10)$$

Also, from $E(\boldsymbol{\mu}) = \mu_1 \mathbf{i}$ and $\text{Cov}(\boldsymbol{\mu}, \mathbf{y}) = \sigma_\eta^2 \mathbf{X}_d \mathbf{X}'_d$, it follows that the optimal estimator of $\boldsymbol{\mu}$, $\tilde{\boldsymbol{\mu}} = E(\boldsymbol{\mu}|\mathbf{y})$ is

$$\begin{aligned}\tilde{\boldsymbol{\mu}} &= \mathbf{i} \tilde{\mu}_1 + \mathbf{X}_d \tilde{\boldsymbol{\eta}} \\ &= \mathbf{W}_\mu \mathbf{y}, \quad \mathbf{W}_\mu = \frac{\mathbf{i} \mathbf{i}'}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \mathbf{Q}^{-1} + \mathbf{X}_d \mathbf{W}_\eta.\end{aligned}$$

Note that $\mathbf{W}_\mu \mathbf{i} = \mathbf{i}$, i.e. the level extraction weights sum to unity, whereas $\mathbf{W}_\eta \mathbf{i} = \mathbf{0}$.

The following theorem deals with the conditions for the equivalence of the fixed and random effects estimators of $\boldsymbol{\beta} = [\mu_1, \boldsymbol{\eta}']'$. μ_1 and η_j and the penalized least

Theorem 1. *The fixed effects estimators of μ_1 and $\boldsymbol{\eta}$ obtained as the minimizers of $PLS(\lambda, d)$, given respectively (7) and (6), are equivalent to the random effects estimators $\tilde{\mu}_1$ and $\tilde{\boldsymbol{\eta}}$ given respectively by (7) and (6) with $q_\eta = \lambda^{-1}$.*

Proof. See Appendix A. □

4.1 Diffuse initial effects

The specification of the random effects model considers the initial value (which is also the unconditional mean of y_t) as fixed. In this section we consider μ_1 as a random effect with a diffuse prior distribution:

$$\mu_1 \sim N(0, \kappa \sigma_\eta^2), \quad \kappa \rightarrow \infty.$$

We further assume that μ_1 is distributed independently of η_j and ϵ_t at all times. The prior quantifies our uncertainty about μ_1 .

Under the above assumptions,

$$\mathbf{y} \sim N(\mathbf{0}, \sigma_\epsilon^2 (\mathbf{Q} + \kappa q_\eta \mathbf{i} \mathbf{i}')),$$

and the following results are proven in Appendix B:

Theorem 2. The diffuse likelihood, defined as $\lim_{\kappa \rightarrow \infty} \{\ell(\mathbf{y}) + 0.5 \ln(\kappa \sigma_\eta^2)\}$, is

$$\ell_\infty(d, q_\eta; \mathbf{y}) = -\frac{1}{2} [T(\ln \tilde{\sigma}_\epsilon^2 + 1) + \ln |\mathbf{Q}| + \ln(\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i})], \quad (11)$$

where

$$\tilde{\sigma}_\epsilon^2 = \frac{1}{T-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i})' \mathbf{Q}^{-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i}).$$

Moreover,

$$\begin{aligned} \lim_{\kappa \rightarrow \infty} \Sigma_y^{-1} &= \frac{1}{\sigma_\epsilon^2} \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right), \\ \lim_{\kappa \rightarrow \infty} E(\mu_1 | \mathbf{y}) &= \frac{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{y}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} = \tilde{\mu}_1, \quad \lim_{\kappa \rightarrow \infty} \text{Var}(\mu_1 | \mathbf{y}) = \frac{\sigma_\epsilon^2}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}}, \\ \lim_{\kappa \rightarrow \infty} E(\boldsymbol{\eta} | \mathbf{y}) &= \mathbf{W}_\eta \mathbf{y}, \quad \lim_{\kappa \rightarrow \infty} E(\boldsymbol{\mu} | \mathbf{y}) = \mathbf{W}_\mu \mathbf{y}, \end{aligned}$$

where the matrices $\mathbf{W}_\eta = q_\eta \mathbf{X}_d' \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right)$ and $\mathbf{W}_\mu = \frac{\mathbf{i} \mathbf{i}'}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \mathbf{Q}^{-1} + \mathbf{X}_d \mathbf{W}_\eta$. Hence, the signal extraction filters for μ_t and η_t are unchanged, conditional on the hyperparameters (d, q_η) .

Proof. See Appendix B □

The difference between $\ell_\infty(d, q_\eta; \mathbf{y})$ and (9) lies in the term $0.5 \ln(\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i})$. While the maximum likelihood estimators of μ_1 and q_η are negatively correlated, the diffuse likelihood removes this dependence, as it is based on a reduced rank transformation of the data, $(\mathbf{I}_T - (\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i})^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}) \mathbf{y}$, that eliminates dependence on μ_1 . The diffuse likelihood is less prone to the pile-up problem, i.e. the concentration of the mass of the distribution of the maximum likelihood estimator of q_η at zero, when the true value is small.

The Online Supplement reports on a Monte Carlo experiment with 1,000 replications generated from 1 with $\sigma_\epsilon^2 = 1$, $\eta_j \sim \text{i.i.d. } N(0, \sigma_\epsilon^2 q_\eta)$, $\mu_1 = 2$, aiming at comparing the mean square estimation error of the profile and diffuse MLE's of the parameters.

5 Time series properties

5.1 Nonstationarity and order of fractional integration

The stated assumptions for the random effects model imply that y_t is a nonstationary random process with $E(y_t) = \mu_1$ and $\text{Var}(y_t) = \sigma_\epsilon^2 (1 + q_\eta H_{t-1, 2(1-d)})$, where $H_{t,r} = \sum_{k=1}^t k^{-r}$ is the generalized harmonic number, such that if $r > 1$ $\lim_{t \rightarrow \infty} H_{t,r} = \zeta(r)$, the Riemann zeta function evaluated at r . Also, for $k > 0$,

$$\text{Cov}(y_t, y_{t-k}) = \sigma_\eta^2 \sum_{j=1}^{t-k-1} [j(k+j)]^{d-1}.$$

The partial sums $S_T = \sum_{t=1}^T y_t$ have variance

$$\begin{aligned} \text{Var}(S_T) &= \sigma_\epsilon^2 \left[T + q_\eta \sum_{t=1}^{T-1} \left(\sum_{j=1}^{T-t} j^{d-1} \right)^2 \right] \\ &= \sigma_\epsilon^2 \left[T + q_\eta \sum_{t=1}^{T-1} H_{t, 1-d}^2 \right]. \end{aligned}$$

The following cases arise, depending on the value of d .

- If $d < 0$, $\text{Var}(S_T) = O(T)$, as $H_{t,r} = O(t^{1-r})$, $r \neq 1$. The FSpN process can be characterized as a short memory process.
- If $d = 0$, $\text{Var}(S_T) = O(T \ln^2(T))$.
- If $d > 0$, $\sum_{t=1}^T H_{t,1-d}^2 = O(T^{2d+1})$. Hence, $\text{Var}(S_T) = O(T^{2d+1})$, and y_t has the long memory property in the rate of the partial sum sense, see Sowell (1990) and Definition (iii) in Haldrup and Valdés (2017).
- As a particular case, if $d = 1$, $H_{t,0} = t$ and $\text{Var}(S_T) = O(T^3)$.

The FS process can be viewed as a discretized version of the Holmgren–Riemann–Liouville, or Type II fractional Brownian motion process, discussed by Marinucci and Robinson (1999). Letting $B(t)$ denote the standard Wiener process, the latter is defined, for $t > 0$ and $H > 0$, by

$$\mu_H(t) = \sqrt{2\mathcal{H}} \int_0^t (t-s)^{\mathcal{H}-1/2} \sigma_\eta dB(s).$$

The parameter $\mathcal{H}$, known as the Hurst exponent, is related to d by $\mathcal{H} = d - 1/2$.

5.2 Type I fractional spline process

Assume that the process started in the indefinite past: let the initial time go to $-\infty$ in (2), set $\mu = \mu_{-\infty}$, and let $\epsilon_t = 0, \forall t$ (thus, $y_t = \mu_t$). The FS process can be characterized as a discrete time *type I* (Marinucci and Robinson (1999)) generalized harmonic process

$$\mu_t = \mu + \sum_{j=0}^{\infty} \psi_j \eta_{t-j},$$

with coefficients $\psi_j = (j+1)^{d-1}$ and $\eta_t \sim \text{WN}(0, \sigma_\eta^2)$.

For $0 < d < 0.5$, denoting the autocovariance function of μ_t by $\gamma_\mu(k) = \text{Cov}(\mu_t, \mu_{t-k})$, we have $\gamma_\mu(0) = \sigma_\eta^2 \zeta(2-2d)$ and

$$\begin{aligned} \gamma_\mu(k) &= \sigma_\eta^2 \sum_{j=1}^{\infty} j^{d-1} (k+j)^{d-1} \\ &= \sigma_\eta^2 k^{2d-1} \left\{ \frac{1}{k} \sum_{j=1}^{\infty} \left(\frac{j}{k}\right)^{d-1} \left(1 + \frac{j}{k}\right)^{d-1} \right\} \\ &\rightarrow_{k \rightarrow \infty} \sigma_\eta^2 k^{2d-1} B(d, 1-2d), \end{aligned}$$

for $k \neq 0$, where $B(d, 1-2d) = \int_0^\infty x^{d-1} (1+x)^{d-1} dx$ is the Beta function, see, e.g., Beran et al. (2013), p. 803.

The spectral density of the FS process is

$$f_\mu(\omega) = \frac{\sigma_\eta^2}{2\pi} \psi(e^{-i\omega}) \psi(e^{i\omega}), \quad \omega \in [-\pi, \pi],$$

where, applying the expansion in Erdélyi (1953, Eq. (8), page 29),

$$\psi(e^{-i\omega}) = e^{i\omega} \left\{ \Gamma(d)(i\omega)^{-d} + \sum_{r=0}^{\infty} \zeta(1-d-r) \frac{(-i\omega)^r}{r!} \right\}.$$

For $\omega \rightarrow 0$,

$$f(\omega) \sim \frac{\sigma_\eta^2}{2\pi} \Gamma^2(d) |\omega|^{-2d}.$$

Hence, for $0 < d < 0.5$, μ_t displays stationary long memory, see Giraitis et al. (2012) and Hassler (2019).

It is interesting to compare the generalized harmonic process to the fractional noise (FN) process, $y_t = \mu + \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}\xi_t$, $\xi_t \sim \text{WN}(0, \sigma^2)$. Figure 5 compares the impulse response function of the FN process to that of the FS process for $d = 0.25$ (top panels) and $d = -0.25$ (bottom panels). It should be remarked that the FS process is never characterized by negative (anti-) persistence.

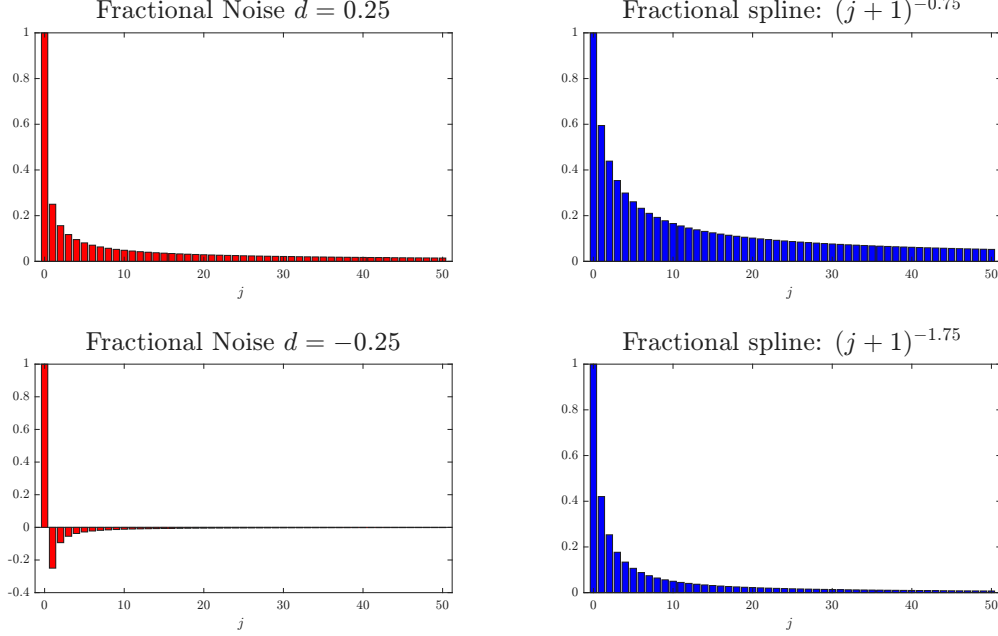


Figure 5: Comparison of the coefficients $\varphi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}$ (fractional noise, left panels) and $\psi_j = (j+1)^{d-1}$ (fractional spline, right panels) for $d = \pm 0.25$.

For $d = 0$, μ_t is the harmonically weighted process $y_t = \mu + \eta_t + \sum_{j=1}^{\infty} \frac{1}{j} \eta_{t-j}$ considered by (Palma, 2007, Problem 3.15), Demetrescu et al. (2008) and Hassler and Hosseinkouchack (2020). For $d < 0$ y_t has short memory and positive persistence (the long run variance is $\sigma_\eta^2 \zeta^2(1-d)$). For $d \geq 0.5$, y_t is non-stationary as the sequence ψ_j is not square summable.

5.3 Long memory and cross-sectional aggregation

In this section we illustrate that the FS model with memory parameter d in the range $(0, 1)$ arises from the cross-sectional aggregation of heterogeneous AR(1) processes, with idiosyncratic autoregressive coefficients. The derivation of the result will be based on the limiting infinite moving average representation of the aggregate process, as in Lemma 1 by Haldrup and Valdés (2017).

Assumption 1. *There exist n heterogeneous agents whose behaviour is characterized by the random process $x_{i,t} = \phi_i x_{i,t-1} + \nu_{it}$, assumed to have started in the indefinite past. The unit specific autoregressive parameters $\phi_i \in (0, 1)$, $i = 1, \dots, n$, are drawn randomly and independently from the log-gamma distribution with parameters α and β , denoted $\phi \sim \text{LogGamma}(\alpha, \beta)$. The unit specific processes ν_{it} are assumed to be i.i.d. with $E(\nu_{it}) = 0$ and $E(\nu_{it}^2) = \sigma_\nu^2$.*

The log-gamma (sometimes referred to as the unit-gamma) random variable has support in the unit interval $(0, 1)$; It has been introduced by Consul and Jain (1971) and Grassia (1977).

Letting $Z \sim \text{Gamma}(\alpha, \beta)$ denote the random variable with probability density function (pdf) $f(z; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} I(z > 0)$, by the variable transformation $\phi = e^{-z}$, the density of the AR coefficient is $f(\phi; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (-\ln \phi)^{\alpha-1} \phi^{\beta-1} I(0 < \phi < 1)$. The r -th moment of ϕ is $E(\phi^r) = \left(\frac{\beta}{\beta+r} \right)^\alpha$.

The assumptions that the process x_{it} has started in the indefinite past and that $\phi_i \in (0, 1)$ imply the following Wold representation:

$$x_{it} = \nu_{it} + \phi_i \nu_{i,t-1} + \phi_i^2 \nu_{i,t-2} + \cdots + \phi_i^j \nu_{i,t-j} + \cdots.$$

For simplicity, as it is done in the literature, we have assumed that the variance of ν_{it} is constant across the units.

Assumption 2. *The log-gamma distribution has parameters $\alpha = 2(1-d)$, $\beta = 2$, i.e. $\phi \sim \text{LogGamma}(2(1-d), 2)$.*

Theorem 3. *Under Assumptions 1 and 2, the aggregate process $x_t = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{it}$ has a limiting fractional spline representation for $n \rightarrow \infty$, $x_t = \sum_{j=0}^\infty \psi_j \eta_{t-j}$, with $\psi_j = (j+1)^{d-1}$ and $\eta_t \sim i.i.d. N(0, \sigma_\nu^2)$.*

Proof. See Appendix C □

Figure 6 plots the pdf of the $\text{LogGamma}(2(1-d), 2)$ random variable for different values of d ; for $0.5 \leq d < 1$ the density starts from 0 at $\phi = 0$ and has an asymptote at $\phi = 1$; for $d < 0.5$ it is a convex function of ϕ with mode at $\phi = e^{2d-1}$. As $\phi \rightarrow 1^-$, in view of $-\ln(\phi) \approx (1-\phi)$, the density $f(\phi; 2(1-d), 2)$ belongs to the family $f(\phi) \sim c_b (1-\phi)^b$ for some constant $c_b > 0$ where $b = 1 - 2d \in (-1, \infty)$ as in Zaffaroni (2004).

The FS representation can deal only with positive persistence even for negative values of d . Negative persistence is produced by the aggregation of the process $x_{it} = -\phi_i x_{i,t-1} + \nu_{it}$, where all the quantities have the same meaning as above. The impulse response function (intervention signature) becomes then $\psi_j = (-1)^j (j+1)^{d-1}$.

Other generative approaches to stationary long memory of the type considered here can be traced to the error duration model by Parke (1999) and the stochastic regimes model by Davidson and Sibbertsen (2005). A continuous time antecedent is the Poisson model with rectangular pulses by Rodriguez-Iturbe et al. (1987), which is a model for precipitations.

5.4 Approximate Inference

Estimation of the random effect model with fixed or diffuse initial condition raises a numerical issue. The direct profile or diffuse maximum likelihood approach based on (9) or (11) approach is computationally expensive, since the inversion of $\mathbf{Q}$ requires $O(T^3)$ operations; only for integer d the computational complexity is linear in T .

Moreover, likelihood evaluation fails already for T moderately large (e.g. $T > 1,000$), since the multicollinearity among the $T-1$ variables $X_{jt} = (t+j-1)^{d-1}$ makes the evaluation numerically unstable or even infeasible. This is attributable to the fact that, when $d > 0$ the condition number of $\mathbf{Q}$ is $1 + q_\eta O(T^{2d})$, as we show in the next Theorem, resulting in a non invertible $\mathbf{Q}$ matrix even for moderate T , in the presence of long memory.

Theorem 4. *The condition number of $\mathbf{Q}$, $C_T = \varrho_{\max}(\mathbf{Q})/\varrho_{\min}(\mathbf{Q})$, is such that*

$$C_T = \begin{cases} O(1), & d < 0, \\ O(\ln^2 T), & d = 0, \\ O(T^{2d}), & d > 0. \end{cases}$$

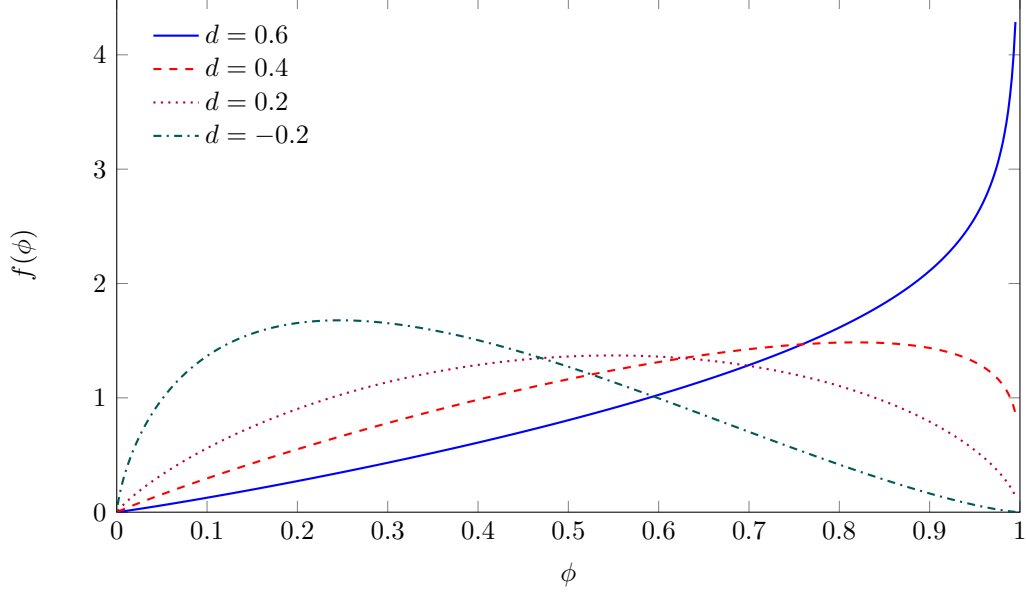


Figure 6: Density $f(\phi) = \frac{2^{2(1-d)}}{\Gamma\{2(1-d)\}}(-\log \phi)^{1-2d}\phi$, $0 < \phi < 1$, for the values of $d = 0.6, 0.4, 0.2, -0.2$.

Proof. See Appendix D. □

To circumvent the problem, we consider the truncated AR(m) representation

$$\pi_m(L)\mu_t = \pi_m(1)\mu_1 + \eta_t,$$

where, for $\psi(L) = \sum_{j=0}^{\infty} (j+1)^{d-1} L^j$, the AR polynomial satisfies $\pi_m(L)\psi(L) = 1$. The model $y_t = \mu_t + \epsilon_t$ is cast in state space form and estimated by maximizing the log-likelihood defined by the diffuse Kalman filter De Jong (1991). The truncation of the dimension of the state space has been proposed for ARFIMA processes by Chan and Palma (1998).

The AR(m) coefficients are obtained as follows. Write $\boldsymbol{\mu} = \mathbf{i}\mu_1 + \boldsymbol{\mu}^*$, where $\boldsymbol{\mu}^* = \mathbf{X}_d \boldsymbol{\eta}$. The matrix $\boldsymbol{\Phi} = (\mathbf{X}_d' \mathbf{X}_d)^{-1} \mathbf{X}_d'$ is such that $\boldsymbol{\Phi} \mathbf{X}_d = \mathbf{I}_{T-1}$ and contains in its rows the coefficients $\boldsymbol{\pi}_m' = [0, \pi_{m-2}, \dots, \pi_1, 1, \mathbf{0}_{T-1-m}]$, $m = 1, 2, \dots, T-1$; hence, $\boldsymbol{\Phi} \boldsymbol{\mu}^* = \boldsymbol{\eta}$. The state space model uses the truncated representation

$$\mu_t^* = - \sum_{j=1}^{\min(m, t-2)} \pi_j \mu_{t-j}^* + \eta_t.$$

State space representation is $y_t = \mathbf{z}' \boldsymbol{\alpha}_t + \mu_1 + \epsilon_t$, $\boldsymbol{\alpha}_t = \mathbf{T} \boldsymbol{\alpha}_{t-1} + \mathbf{z} \eta_t$, where $\mathbf{z}' = [1, 0, \dots, 0]$,

$$\mathbf{T} = \begin{bmatrix} -\pi_1 & -\pi_2 & \cdots & -\pi_{m-1} & -\pi_m \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

In the type I framework with $d \in (0, 0.5)$ pseudo maximum likelihood estimation can be performed by using the Whittle broadband approximation. In this case the evaluation of the spectral density of the FSpN model requires the truncation of formula for $|\psi(\omega)|$.

6 Illustrations

6.1 U.S. monthly Inflation

A recursive forecasting experiment is conducted that aims at evaluating the predictive performance of the FSpN model and comparing it with the fractional noise plus noise (FNpN) specification, which replaces μ_t with fractional noise process; also in this case the AR(m) approximation is considered with the same order m , which is set equal to $m = 40$ after a preliminary selection stage.

The design of the recursive forecasting exercise is the same as that in Section 3.2: starting from December 1999, h step-ahead forecasts are computed for the FSpN and FNpN models, using $m = 40$; the training sample is then moved forward by adding an observation until the end of sample is reached. We compare the predictive accuracy with three benchmarks: the AR(1) model, the ARMA(1,1) models with a constant, and the driftless ARIMA(1,1,1) model. Parameters are re-estimated at each iteration.

The recursive estimates of the parameters (not presented for brevity) show that before 2008 the point estimates of the memory parameter lie in the nonstationary region, being more affected by the high and variable inflation levels of the 70ies and 80ies and by the transition to the great moderation. Secondly, at the end of the great moderation a structural break seem to have affected all the parameter estimates.

At the end of the sample the d estimate is in the stationary region for the FSpN specification, and it is systematically lower than that of the FNpN. In particular, the estimates of the FSpN model parameters are $\hat{d} = 0.2557$, $\hat{\sigma}_\eta^2 = 0.0320$, $\hat{\sigma}_\epsilon^2 = 0.0196$ ($\hat{q}_\eta = 1.6304$) $\hat{\mu}_1 = 0.2336$. Those of the FNpN model parameters are $\hat{d} = 0.4976$, $\hat{\sigma}_\eta^2 = 0.0387$, $\hat{\sigma}_\epsilon^2 = 0.0148$ ($\hat{q}_\eta = 2.6216$) $\hat{\mu}_1 = 0.1701$. They point out that the size of the estimate of the memory parameter depends on the model that is assumed, i.e., it is a property of the process that is assumed for μ_t .

The evaluation results are based on $312 - h + 1$ h -step-ahead prediction errors. The comparison is based on the Mean Square Forecast Error (MSFE). Table 2 displays the MSFE of the FSpN model in the first column for all forecast horizons h going from 1 to 12. The following columns display the MSFE of the alternative predictors relative to that of the FSpN model, taken as a reference. Values greater than 1 signify that the latter is more accurate than the competitor. Values in bold correspond to a significant one sided 5% Diebold and Mariano (2002) test of equal forecast accuracy, based on the squared loss differential. The predictive performance of FSpN and FNpN is very similar, and the long memory specification outperforms short memory ones and the ARIMA(1,1,1) model, except at shorter horizons.

Finally, Figure 7 plots the point and approximate interval estimates of the persistent component of U.S. monthly inflation, μ_t , as represented by a fractional spline. The estimates were obtained by the Kalman filter and smoothing algorithm applied to the AR(m) approximation of the FSpN model.

6.2 Realized variance

The daily realized variance for the stock index S&P500 is obtained as the sum of squared 5-minute intradaily returns sampled at 5-minute intervals. It has been made available by the Oxford-Man Institute Realized Library, for the period from January 03, 2000, to June 03, 2021, for a total of $n = 5,369$ observations. The series has been discontinued, but it is used as a testbed for time series methods. The series is plotted in Figure 8 along with its sample autocorrelation function.

For the complete series, the estimates of the FSpN model parameters are $\hat{d} = 0.571$, $\hat{\sigma}_\eta^2 = 0.153$, $\hat{\sigma}_\epsilon^2 = 0.155$ ($\hat{q}_\eta = 0.988$) $\hat{\mu}_1 = -9.433$. Those of the FNpN model parameters are $\hat{d} = 0.647$, $\hat{\sigma}_\eta^2 = 0.197$, $\hat{\sigma}_\epsilon^2 = 0.125$ ($\hat{q}_\eta = 1.577$) $\hat{\mu}_1 = -9.408$. They confirm that the size of the

Horizon	FSpN	FNpN	AR(1)	ARMA(1,1)	ARIMA(1,1,1)
1	0.0800	0.9962	0.9379	1.0607	0.9463
2	0.1011	0.9980	1.0461	1.0381	0.9740
3	0.0995	1.0001	1.0428	1.0381	0.9907
4	0.0958	1.0034	1.0511	1.0386	1.0128
5	0.0968	1.0058	1.0633	1.0431	1.0239
6	0.0964	1.0076	1.0800	1.0447	1.0331
7	0.0950	1.0087	1.1025	1.0483	1.0421
8	0.0958	1.0095	1.1044	1.0449	1.0522
9	0.0956	1.0118	1.1123	1.0463	1.0627
10	0.0936	1.0114	1.1376	1.0506	1.0671
11	0.0968	1.0119	1.1060	1.0499	1.0671
12	0.1014	1.0135	1.0637	1.0470	1.0760

Table 2: Mean square forecast error of the FSpN model (first column) and MSFE ratios (columns 2-5) taking FSpN as the numeraire. Values in bold correspond to a significant one sided 5% DM test of equal forecast accuracy, based on the squared loss differential.

estimate of the memory parameter depends on the model that is assumed, i.e., it is a property of the process that is assumed for μ_t .

Starting from 31 December 2018, h step-ahead forecasts are computed for the FSpN and FNpN models, using $m = 80$; the training sample is then moved forward by adding an observation until the end of sample is reached. We compare the predictive accuracy with three benchmarks: the AR(1) model, the ARMA(1,1) models with a constant, and the driftless ARIMA(1,1,1) model and the Heterogeneous Autoregressive (HAR) model by Corsi (2009). The latter is a restricted AR(22) model depending on three parameters. Parameters are re-estimated at each iteration. The evaluation results are based on $601 - h + 1$ h -step-ahead prediction errors.

The FSpN model has similar performance to the FNpN and the Heterogeneous Autoregressive (HAR) model predictor by Corsi (2009) as illustrated by Table 3 (although at some forecast horizons the Diebold and Mariano (2002) test rejects the null of equal predictive accuracy according to the squared error loss function, due to the large sample size, the differences in MSFE are quantitatively small). Notably, the best performance is offered by the ARMA(1,1) predictor.

6.3 Tree ring-based European temperature reconstruction

The reconstruction of temperature anomalies in the summer months (JJA) is based on annually resolved tree (pine and larch) ring width measurement series located in Central Europe; see Büntgen et al. (2011). The widths are detrended both to take into account the growth curve of the trees (alignment by cambial age) and to isolate the high-frequency component (cubic-spline with cut-off period 20 yrs). After a standardization, they are scaled on observed temperatures in the period 1864-2006.

The series is plotted in Figure 9, along with its sample correlogram. There are $T = 2,503$ annual observations starting in the year $t_1 = -499$ B.C. and ending in $t_T = 2003$. The series displays a clear warming trend in the last part of the sample. We estimated using the sample up to the year 1983, and leave the last 20 observations for predictive evaluation.

The estimates of the FSpN model parameters are $\hat{d} = 0.349$, $\hat{\sigma}_\eta^2 = 0.232$, $\hat{\sigma}_\epsilon^2 = 0.1074$ ($\hat{q}_\eta = 2.158$) $\hat{\mu}_1 = -0.388$. Those of the FNpN model parameters are $\hat{d} = 0.558$, $\hat{\sigma}_\eta^2 = 0.284$, $\hat{\sigma}_\epsilon^2 = 0.071$ ($\hat{q}_\eta = 4.008$) $\hat{\mu}_1 = -0.246$. Again, we observe that, while the two fractional models provide a similar fit, the memory parameter estimate is larger for the FN specification.

The observations from 1987 to 2008 were left out for predictive validation. At the end of the

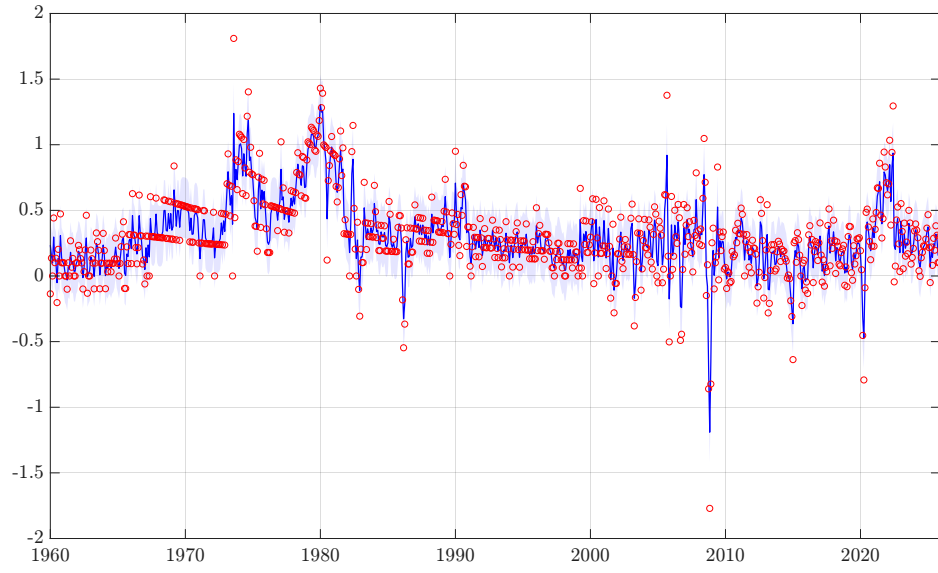


Figure 7: Approximate point and interval posterior estimates of the fractional spline μ_t .

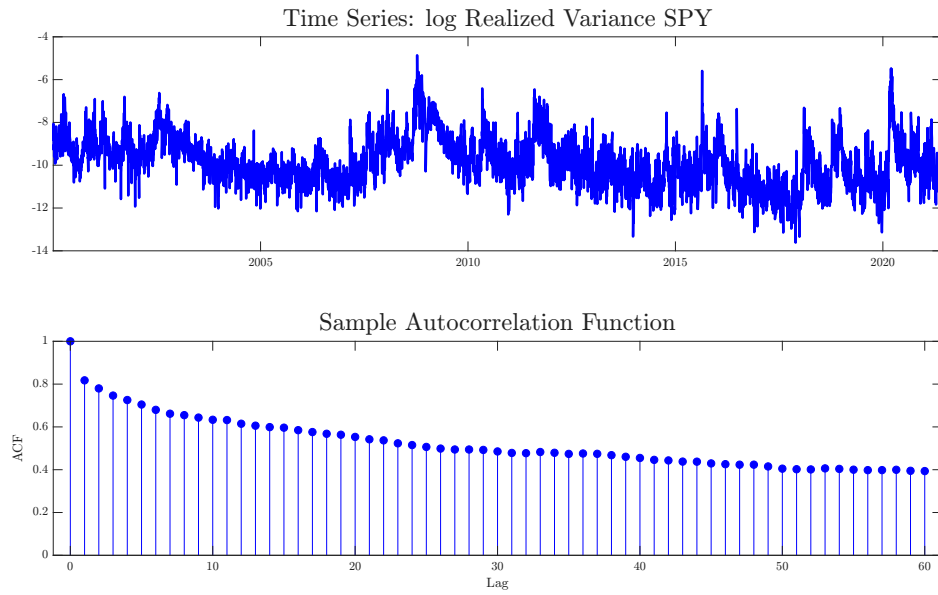


Figure 8: Daily realized variance series (logarithms) of the S&P500 index (January 03, 2000 - June 03, 2021): time series (upper plot) and sample autocorrelation function.

Horizon	FSpN	FNpN	AR(1)	ARMA(1,1)	ARIMA(1,1,1)	VAR
1	0.4799	1.0028	1.1125	0.9912	1.0192	1.0172
2	0.6046	1.0050	1.0827	0.9799	1.0331	1.0157
3	0.7144	1.0054	1.1051	0.9749	1.0339	1.0197
4	0.7989	1.0058	1.1274	0.9691	1.0369	1.0167
5	0.8688	1.0060	1.1382	0.9634	1.0400	1.0136
6	0.9452	1.0063	1.1582	0.9603	1.0468	1.0184
7	0.9938	1.0070	1.1660	0.9506	1.0529	1.0174
8	1.0451	1.0075	1.1635	0.9446	1.0596	1.0122
9	1.1024	1.0079	1.1579	0.9425	1.0679	1.0083
10	1.1493	1.0084	1.1508	0.9375	1.0753	1.0054
11	1.1928	1.0090	1.1394	0.9314	1.0819	0.9953
12	1.2349	1.0096	1.1282	0.9262	1.0901	0.9955
13	1.2665	1.0103	1.1179	0.9173	1.0959	0.9931
14	1.3027	1.0112	1.1036	0.9106	1.1055	0.9916
15	1.3365	1.0121	1.0871	0.9034	1.1148	0.9890

Table 3: Mean square forecast error of the FSpN model (first column) and MSFE ratios (columns 2-6) taking FSpN as the numeraire. Values in bold correspond to a significant one sided 5% DM test of equal forecast accuracy, based on the squared loss differential.

sample the forecasts are mean reverting towards $\hat{\mu}_1 = -0.388$, as can be seen from Figure 10, while the observations are almost systematically above the prediction. This prediction failure (a significant underprediction) provides the evidence for a potential structural break, likely due to an acceleration of global warming.

7 Conclusions

The Fractional Spline process provides an alternative representation of long-memory dynamics. Empirically, the estimated memory parameter is typically smaller than that obtained under the fractional noise model. The model provides an alternative to the fractional noise model and has interesting applications to economic and financial time series, as well as climate time series.

In the paper we have also explored the application of ℓ_1 regularization of the fractional spline parameters so as to enforce sparsity. While we think that this can be a valuable tool for the identification of structural change and large interventions, we leave the issue to future research.

Acknowledgements

I thank Niels Haldrup, Federico Carlini and Paolo Santucci de Magistris for discussions over the preliminary ideas in the paper. I also thank participants to the 2026 Econometric Models of Climate Change Conference, Aalborg University, Denmark, August 20-21, 2026, and in particular Eduardo Vera-Valdés and Charisios Grivas for their suggestions.

Figure 9: Tree ring-based reconstruction of temperatures: 480 BC - 2004 AD.

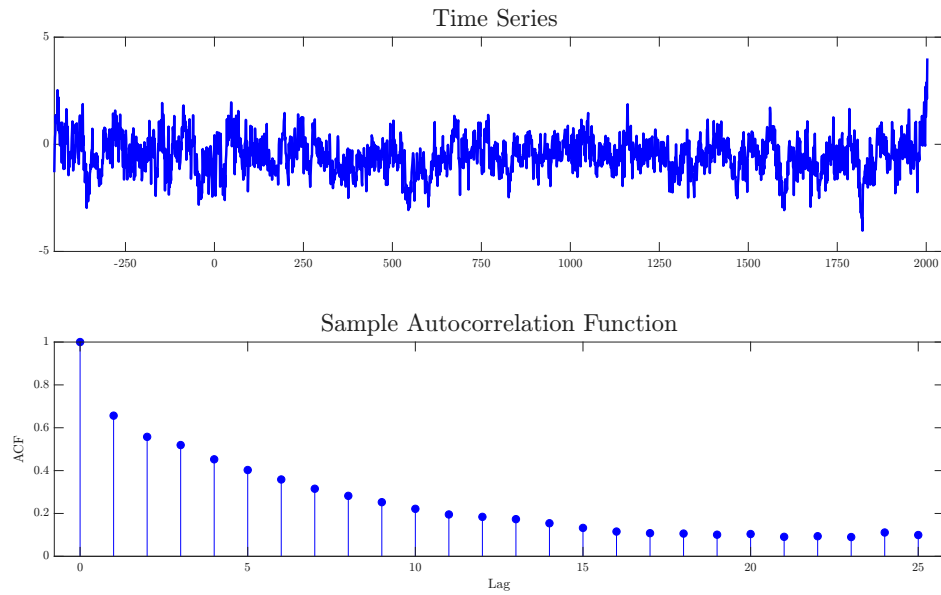
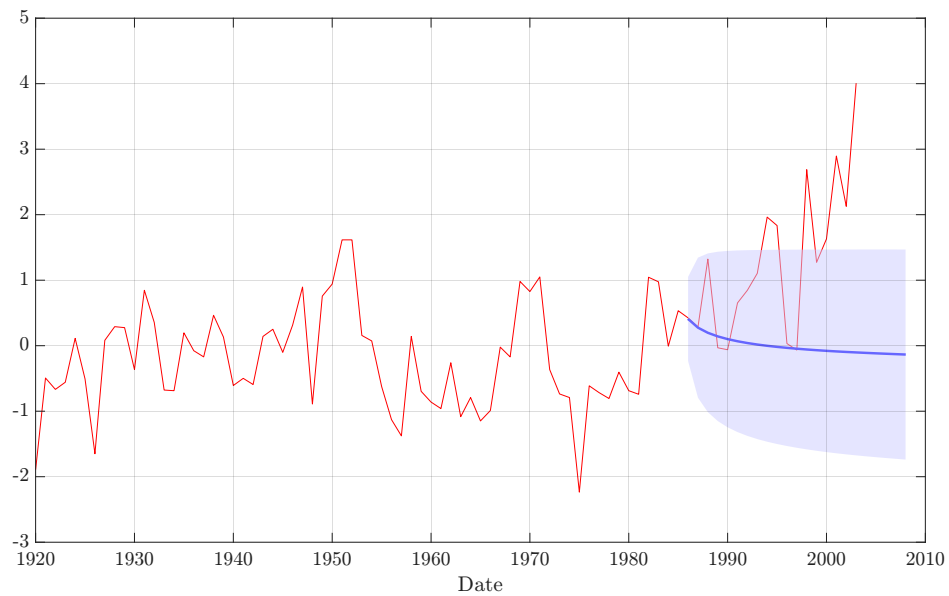


Figure 10: Out of sample predictions: evidence for a potential structural break



A Proof of Theorem 1

Proof of the equivalence of the FE and RE estimates

Recall the definition of the $T \times T$ matrix $\mathbf{X} = [\mathbf{i}, \mathbf{X}_d]$, the $T \times 1$ vector $\boldsymbol{\beta} = [\mu_1, \boldsymbol{\eta}']'$, and the $T \times T$ matrix $\mathbf{V} = \text{diag}(0, \mathbf{I}_{T-1})$. The regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is estimated under the constraint $\boldsymbol{\beta}'\mathbf{V}\boldsymbol{\beta} = \boldsymbol{\eta}'\boldsymbol{\eta} < c$, for a positive constant c , or equivalently by minimizing the penalized least squares (PLS) criterion:

$$S(d, \lambda) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\mathbf{V}\boldsymbol{\beta}. \quad (12)$$

In the expression for the “ridge” estimator, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{V})^{-1}\mathbf{X}'\mathbf{y}$, we partition

$$\mathbf{X}'\mathbf{X} + \lambda\mathbf{V} = \begin{bmatrix} T & \mathbf{a}' \\ \mathbf{a} & \mathbf{A} \end{bmatrix},$$

with $\mathbf{a} = \mathbf{X}_d'\mathbf{i}$ and $\mathbf{A} = \mathbf{X}_d'\mathbf{X}_d + \lambda\mathbf{I}$.

Consider the following two equivalent expressions for the inverse

$$(\mathbf{X}'\mathbf{X} + \lambda\mathbf{V})^{-1} = \begin{bmatrix} T^{-1}(1 + T^{-1}\mathbf{i}'\mathbf{X}_d(\mathbf{X}_d'\mathbf{M}\mathbf{X}_d + \lambda\mathbf{I})^{-1}\mathbf{X}_d'\mathbf{i}), & -T^{-1}\mathbf{i}'\mathbf{X}_d(\mathbf{X}_d'\mathbf{M}\mathbf{X}_d + \lambda\mathbf{I})^{-1} \\ -T^{-1}(\mathbf{X}_d'\mathbf{M}\mathbf{X}_d + \lambda\mathbf{I})^{-1}\mathbf{X}_d'\mathbf{i}, & (\mathbf{X}_d'\mathbf{M}\mathbf{X}_d + \lambda\mathbf{I})^{-1} \end{bmatrix}, \quad (13)$$

and

$$(\mathbf{X}'\mathbf{X} + \lambda\mathbf{V})^{-1} = \begin{bmatrix} (\mathbf{i}'\mathbf{Q}^{-1}\mathbf{i})^{-1}, & -q_\eta(\mathbf{i}'\mathbf{Q}^{-1}\mathbf{i})^{-1}\mathbf{i}'\mathbf{Q}^{-1}\mathbf{X}_d \\ -q_\eta\mathbf{X}_d'\mathbf{Q}^{-1}\mathbf{i}(\mathbf{i}'\mathbf{Q}^{-1}\mathbf{i})^{-1}, & \mathbf{A}^{-1} + q_\eta^2\mathbf{X}_d'\frac{\mathbf{Q}^{-1}\mathbf{i}\mathbf{i}'\mathbf{Q}^{-1}}{\mathbf{i}'\mathbf{Q}^{-1}\mathbf{i}}\mathbf{X}_d \end{bmatrix}. \quad (14)$$

The equivalence is shown by considering standard results on the inverse of a partitioned matrix, see e.g. (Henderson and Searle, 1981, Section 1.2), noticing that $\mathbf{A} - \frac{\mathbf{a}\mathbf{a}'}{T} = \mathbf{X}_d'\mathbf{M}\mathbf{X}_d + \lambda\mathbf{I}$ and $\mathbf{Q}^{-1} = \mathbf{I} - \mathbf{X}_d'\mathbf{A}^{-1}\mathbf{X}_d$, as it is shown below:

$$\begin{aligned} \mathbf{Q}^{-1} &= (\mathbf{I} + q_\eta\mathbf{X}_d\mathbf{X}_d')^{-1} \\ &= \mathbf{I} - q_\eta\mathbf{X}_d(\mathbf{I} + q_\eta\mathbf{X}_d'\mathbf{X}_d)^{-1}\mathbf{X}_d' \\ &= \mathbf{I} - \mathbf{X}_d(\lambda\mathbf{I} + \mathbf{X}_d'\mathbf{X}_d)^{-1}\mathbf{X}_d' \\ &= \mathbf{I} - \mathbf{X}_d'\mathbf{A}^{-1}\mathbf{X}_d. \end{aligned}$$

Similarly, by the Sherman-Woodbury-Morrison inversion formula Henderson and Searle (1981), $\mathbf{A}^{-1} = q_\eta(\mathbf{I} - q_\eta\mathbf{X}_d'\mathbf{Q}^{-1}\mathbf{X}_d)$, so that

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{X}_d &= q_\eta(\mathbf{I} - q_\eta\mathbf{X}_d'\mathbf{Q}^{-1}\mathbf{X}_d)\mathbf{X}_d' \\ &= q_\eta\mathbf{X}_d' - q_\eta^2\mathbf{X}_d'\mathbf{Q}^{-1}\mathbf{X}_d\mathbf{X}_d' \\ &= q_\eta^2\mathbf{X}_d'\mathbf{Q}^{-1}(\mathbf{Q} - q_\eta\mathbf{X}_d\mathbf{X}_d') \\ &= q_\eta^2\mathbf{X}_d'\mathbf{Q}^{-1}, \end{aligned}$$

and $T - \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} = \mathbf{i}'\mathbf{Q}^{-1}\mathbf{i}$.

Postmultiplying (13) by $\mathbf{X}'\mathbf{y} = [\mathbf{i}, \mathbf{X}_d']'\mathbf{y}$ yields $[\hat{\mu}_1, \hat{\boldsymbol{\eta}}']'$ according to (7) and (6). Postmultiplying (14) by $\mathbf{X}'\mathbf{y}$ yields $[\tilde{\mu}_1, \tilde{\boldsymbol{\eta}}']'$ according to (8) and (10), which shows the equivalence.

B Proof of Theorem 2

Suppressing dependence on unknown parameters, the log-likelihood admits the additive representation $\ell(\mathbf{y}) = \ell(\mathbf{y}|\mu_1) + \ell(\mu_1) - \ell(\mu_1|\mathbf{y})$, where, neglecting constant terms

$$\begin{aligned} \ell(\mathbf{y}|\mu_1) &= -\frac{1}{2} \left\{ T \ln \sigma_\epsilon^2 + \ln |\mathbf{Q}| + \frac{1}{\sigma_\epsilon^2} (\mathbf{y} - \mu_1\mathbf{i})'\mathbf{Q}^{-1}(\mathbf{y} - \mu_1\mathbf{i}) \right\}, \\ \ell(\mu_1) &= -\frac{1}{2} \left\{ \ln(\kappa\sigma_\eta^2) + \frac{\mu_1^2}{\kappa\sigma_\eta^2} \right\}, \\ \ell(\mu_1|\mathbf{y}) &= -\frac{1}{2} \left\{ \ln \sigma_{\mu_1|\mathbf{y}}^2 + \frac{(\mu_1 - \sigma_{\mu_1|\mathbf{y}}^2 \sigma_\epsilon^{-2} \mathbf{i}'\mathbf{Q}^{-1}\mathbf{y})^2}{\sigma_{\mu_1|\mathbf{y}}^2} \right\}, \quad \sigma_{\mu_1|\mathbf{y}}^2 = \left(\frac{1}{\kappa\sigma_\eta^2} + \frac{\mathbf{i}'\mathbf{Q}^{-1}\mathbf{i}}{\sigma_\epsilon^2} \right)^{-1}. \end{aligned}$$

Substituting into $\ell(\mathbf{y})$ and rearranging gives, after some algebra,

$$\ell(\mathbf{y}) + \frac{1}{2} \ln(\kappa \sigma_\eta^2) = -\frac{1}{2} \left\{ T \ln \sigma_\epsilon^2 + \ln |\mathbf{Q}| + \frac{\mathbf{y}' \mathbf{Q}^{-1} \mathbf{y}}{\sigma_\epsilon^2} + \ln \left(\frac{1}{\kappa \sigma_\eta^2} + \frac{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}}{\sigma_\epsilon^2} \right) - \frac{\mathbf{y}' \mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1} \mathbf{y}}{\sigma_\epsilon^4} \left(\frac{1}{\kappa \sigma_\eta^2} + \frac{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}}{\sigma_\epsilon^2} \right)^{-1} \right\}.$$

Taking the limit for $\kappa \rightarrow \infty$, gives the diffuse log-likelihood

$$\ell_\infty(\mathbf{y}) = -\frac{1}{2} \left\{ (T-1) \ln \sigma_\epsilon^2 + \ln |\mathbf{Q}| + \ln (\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}) + \frac{1}{\sigma_\epsilon^2} \mathbf{y}' \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right) \mathbf{y} \right\}.$$

Differentiating with respect to σ_ϵ^2 and setting the derivative equal to zero yields the diffuse MLE

$$\begin{aligned} \tilde{\sigma}_\epsilon^2 &= \frac{1}{T-1} \mathbf{y}' \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right) \mathbf{y} \\ &= \frac{1}{T-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i})' \mathbf{Q}^{-1} (\mathbf{y} - \tilde{\mu}_1 \mathbf{i}). \end{aligned}$$

Replacing into $\ell_\infty(\mathbf{y})$ yields the diffuse profile likelihood (11).

By application of the Woodbury-Sherman-Morrison inversion lemma,

$$\begin{aligned} \Sigma_y^{-1} &= \frac{1}{\sigma_\epsilon^2} (\mathbf{Q} + \mathbf{i} \mathbf{i}' \kappa q_\eta)^{-1} \\ &= \frac{1}{\sigma_\epsilon^2} \left(\mathbf{Q}^{-1} - \frac{\mathbf{Q}^{-1} \mathbf{i} \mathbf{i}' \mathbf{Q}^{-1}}{\frac{1}{\kappa q_\eta} + \mathbf{i}' \mathbf{Q}^{-1} \mathbf{i}} \right), \end{aligned}$$

which gives the limiting expression in Section 4.1. The limiting results for the posterior mean and variance of μ_1 , as well as $\lim_{\kappa \rightarrow \infty} \mathbb{E}(\boldsymbol{\eta}|\mathbf{y}) = \mathbf{W}_\eta \mathbf{y}$, $\lim_{\kappa \rightarrow \infty} \mathbb{E}(\boldsymbol{\mu}|\mathbf{y}) = \mathbf{W}_\mu \mathbf{y}$, follow immediately.

C Proof of Theorem 3

Consider the cross-sectional aggregation of the series x_{it} , defined as in Granger (1980),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_{it} = \sum_{j=0}^{\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i^j \nu_{i,t-j}.$$

For j given, the random variables $\phi_i^j \nu_{i,t-j}$ are i.i.d. with mean zero and variance $\sigma_\nu^2 \mathbb{E}(\phi_i^{2j})$, with $\mathbb{E}(\phi_i^{2j}) = (j+1)^{2(d-1)}$. Hence, applying the Lindeberg-Lévy central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_i^j \nu_{i,t-j} \rightarrow_d \mathcal{N} \left(0, \sigma_\nu^2 (j+1)^{2(d-1)} \right),$$

and $x_t = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{it}$ can be expressed as

$$x_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j},$$

where $\psi_j = (1+j)^{d-1}$ and $\eta_{t-j} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\nu^2)$.

D Proof of Theorem 4

The eigenvalues of $\mathbf{Q} = \mathbf{I}_T + q_\eta \mathbf{X}_d \mathbf{X}_d'$, are $\varrho(\mathbf{Q}) = 1 + q_\eta \varrho(\mathbf{X}_d \mathbf{X}_d')$. Now, $\varrho(\mathbf{X}_d \mathbf{X}_d') = \varrho(\mathbf{X}_d' \mathbf{X}_d)$, and since $\mathbf{X}_d' \mathbf{X}_d = \text{diag}(0, \mathbf{C}_d' \mathbf{C}_d)$, $\varrho_{\min}(\mathbf{X}_d \mathbf{X}_d') = 0$, which implies $\varrho_{\min}(\mathbf{Q}) = 1$. As a result, $C_T = 1 + \varrho_{\max}(\mathbf{C}_d' \mathbf{C}_d)$.

We now derive an upper and lower bound for $\varrho_{\max}(\mathbf{C}'_d \mathbf{C}_d)$. By Corollary 2.3.2 in Golub and Van Loan (2013),

$$\begin{aligned}\varrho_{\max}(\mathbf{C}'_d \mathbf{C}_d) &\leq \left\{ \sum_{k=1}^{T-1} k^{d-1} \right\}^2 \\ &= H_{T-1,1-d}^2.\end{aligned}$$

The lower bound can be obtained by the matrix norm inequality (2.3.12) in Golub and Van Loan (2013), according to which

$$\varrho_{\max}(\mathbf{C}'_d \mathbf{C}_d) \geq \left(\frac{1}{\sqrt{T-1}} \sum_{k=1}^{T-1} k^{d-1} \right)^2 = \frac{1}{T-1} H_{T-1,1-d}^2.$$

This bound is conservative and can be made sharper by considering a test eigenvector, e.g. $\frac{1}{\sqrt{T-1}} \mathbf{i}_{T-1}$, giving

$$\begin{aligned}\varrho_{\max}(\mathbf{C}'_d \mathbf{C}_d) &\geq \frac{1}{T-1} \mathbf{i}'_{T-1} \mathbf{C}'_d \mathbf{C}_d \mathbf{i}_{T-1} \\ &= \frac{1}{T-1} \left\{ \sum_{k=1}^{T-1} H_{k,d-1}^2 \right\}^2 \\ &= O(H_{T-1,1-d}^2).\end{aligned}$$

Hence, C_T is as stated in Theorem 4.

References

- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1):5–59.
- Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer Science & Business Media.
- Büntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., Kaplan, J. O., Herzig, F., Heussner, K.-U., Wanner, H., et al. (2011). 2500 years of european climate variability and human susceptibility. *science*, 331(6017):578–582.
- Chan, N. H. and Palma, W. (1998). State space modeling of long-memory processes. *The Annals of Statistics*, 26(2):719–740.
- Consul, P. C. and Jain, G. C. (1971). On the log-gamma distribution and its properties. *Statistische Hefte*, 12(2):100–106.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of financial econometrics*, 7(2):174–196.
- Davidson, J. and Sibbertsen, P. (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of econometrics*, 128(2):253–282.
- De Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics*, pages 1073–1083.
- Demetrescu, M., Kuzin, V., and Hassler, U. (2008). Long memory testing in the time domain. *Econometric Theory*, 24(1):176–215.
- Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of econometrics*, 105(1):131–159.
- Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.
- Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press (UK).
- Erdélyi, A., editor (1953). *Higher Transcendental Functions*. Bateman Manuscript Project. McGraw-Hill, New York.
- Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large sample inference for long memory processes*. Imperial College Press.
- Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. The Johns Hopkins University Press.
- Granger, C. W. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227–238.
- Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29.
- Grassia, A. (1977). On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions. *Australian Journal of Statistics*, 19(2):108–114.

- Haldrup, N. and Valdés, J. E. V. (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics*, 199(1):1–11.
- Hartl, T. (2025). Robust trend estimation for strongly persistent data with unobserved memory. *Journal of Business & Economic Statistics*, (just-accepted):1–23.
- Hassler, U. (2019). *Time series analysis with long memory in view*. John Wiley & Sons.
- Hassler, U. and Hosseinkouchack, M. (2020). Harmonically weighted processes. *Journal of Time Series Analysis*, 41(1):41–66.
- Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM review*, 23(1):53–60.
- Hodrick, R. J. and Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, pages 1–16.
- Hosking, J. R. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.
- Hurvich, C. M., Moulines, E., and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica*, 73(4):1283–1328.
- Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009). ℓ_1 trend filtering. *SIAM review*, 51(2):339–360.
- Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, 80(1-2):111–122.
- Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the american statistical association*, 55(290):299–306.
- Palma, W. (2007). *Long-memory time series: theory and methods*. John Wiley & Sons.
- Parke, W. R. (1999). What is fractional integration? *Review of Economics and Statistics*, 81(4):632–638.
- Proietti, T. (2007). Signal extraction and filtering by linear semiparametric methods. *Computational statistics & data analysis*, 52(2):935–958.
- Proietti, T. and Maddanu, F. (2024). Modelling cycles in climate series: The fractional sinusoidal waveform process. *Journal of Econometrics*, 239(1):105299.
- Robinson, P. M. (2003). *Time series with long memory*. Oxford University Press, Advanced Texts in Econometrics.
- Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 269–288.
- Ruppert, D., Wand, M. P., Carroll, R. J., and Ruppert, D. (2003). *Semiparametric regression*, volume 12. Cambridge university press Cambridge.
- Sowell, F. (1990). The fractional unit root distribution. *Econometrica: Journal of the Econometric Society*, pages 495–505.

- Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323.
- Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381):81–89.
- Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120(1):75–102.