Discussion Paper 1/2021

Guatemalan Econometric Study Group

Universidad Francisco Marroquin

January 24, 2021

Optimal signal extraction for score-driven models

Astrid Ayala, Szabolcs Blazsek*, and Adrian Licht

School of Business, Universidad Francisco Marroquín, Guatemala City, Guatemala

Abstract: In this paper, a novel approach of signal smoothing for score-driven models is suggested, by using results from the literature on minimum mean squared error (MSE) signals. The new smoothing procedure can be applied to more general score-driven models than the state space smoothing recursions procedures from the literature. Score-driven location, trend, and seasonality models with constant and score-driven scale parameters for macroeconomic variables are used. The two-step smoothing procedure is computationally fast, and it uses closed-form formulas for smoothed signals. In the first step, the score-driven models are estimated by using the maximum likelihood (ML) method. In the second step, the ML estimates of the score functions are substituted into the minimum MSE signal extraction filter. Applications for monthly data of the seasonally adjusted and the not seasonally adjusted (NSA) United States (US) inflation rate variables for the period of 1948 to 2020 are presented.

Keywords: Signal extraction, Minimum mean squared error (MSE) signals, Dynamic conditional score (DCS), Generalized autoregressive score (GAS)

*School of Business, Universidad Francisco Marroquín, Calle Manuel F. Ayau, Guatemala City, 01010, Guatemala.

E-mails: aayala@ufm.edu (Ayala), sblazsek@ufm.edu (Blazsek), adrian.licht@ufm.edu (Licht)

1. Introduction

In this paper, we apply results from the literature on signal extraction, and we present a new approach of minimum mean squared error (MSE) signal extraction for score-driven state space models of location, trend, seasonality, and scale for macroeconomic time series variables.

Signal extraction from economic or financial variables y_t is important for statistical analyses that support the decisions of economic agents (Erceg and Lewin 2003), or policymakers (Ghysels 1987; Bryan and Pike 1991; Cristadoro et al. 2005). From the literature on signal extraction, some relevant works are Bell (1984), Bell and Hillmer (1988), McElroy (2008), and McElroy and Maravall (2014), in which minimum MSE signal smoothing formulas are presented for signal plus noise models. Those works use a variety of models, for which signal or noise or both is non-stationary, and the error terms in the signal and noise equations are heteroskedastic, non-Gaussian, and correlated. As an empirical contribution we apply the results of McElroy and Maravall (2014) to signal smoothing for score-driven models, and as a theoretical contribution we prove that the assumptions of those results are satisfied for the score-driven models.

The class of score-driven models are observation-driven (Cox 1981) state space models (Harvey 1989; Durbin and Koopman 2012), which are introduced in the works of Creal et al. (2008), and Harvey and Chakravarty (2008). Those papers have started an extensive literature that includes more than 200 publications until the date of the present paper. Score-driven models represent one of the most important developments in the field of time series econometrics in the last decade. The filters of score-driven models are updated by using the scaled conditional score of the log-likelihood (LL) with respect to a time-varying parameter, and score-driven models are estimated by using the maximum likelihood (ML) method (e.g. Harvey 2013; Creal et al. 2013; Blasques et al. 2017). Score-driven models are robust to outliers and missing observations, and the filters that update score-driven models are generalizations of the updating mechanisms of classical time series models. In many cases, score-driven models have superior in-sample statistical and out-of-sample predictive performances than classical models. This suggests that the practical use of score-driven models for macroeconomic decisions will be important in the future, and there will be a practical need for straightforward smoothing methods for those models.

In the work of Harvey (2013, pp. 83-89), it is shown that the filtered signal $E(s_t|y_1, \ldots, y_{t-1})$ for $t = 1, \ldots, T$ and the smoothed signal $E(s_t|y_1, \ldots, y_T)$ for $t = 1, \ldots, T$ differ for the score-driven models. The work of Harvey (2013) is relevant to our paper, because it presents an application of a state space smoothing recursions procedure (Koopman and Harvey 2003) to first-order score-driven location models. According to Harvey (2013, p. 87), the generalization of the state space smoothing recursions procedure to score-driven models with several lags is not straightforward. The smoothing procedure that is suggested in the present paper can be applied to score-driven models with several lags in a straightforward way.

Another paper from the literature that is relevant to our paper is the recent work of Buccheri et al. (2019), in which a state space smoothing recursions procedure is suggested for the score-driven location, scale, or duration model. The authors note that, as score-driven models are observation-driven models, the time-varying parameters are one-step ahead predictable, hence, the estimate of the filtered signal is improved in score-driven models, by using information from contemporaneous and future observations in the smoothed signal. This is the main motivation for the development of the smoothing method for score-driven models in our paper.

In the signal extraction procedure of this paper, the ML estimates of parameters of score-driven models are obtained in the first step by using numerical maximization of the LL function, and the minimum MSE signals are estimated in the second step by using a closed-form signal smoothing formula. We suggest a smoothing procedure for score-driven models, which are more complex than the score-driven models of Harvery (2013, p. 87) and Buccheri et al. (2019). In the score-driven models of the present paper, score-driven location, trend, and seasonality filters with constant and score-driven scale parameters are included. The smoothing procedure is easy to apply in practice, because the minimum MSE signal extraction formula for score-driven models is available in closed form (i.e., it is not determined by smoothing recursions).

In the signal plus noise models for score-driven data, the updating terms of the signal and noise components are correlated time series. For the theoretical contribution, we present that the assumptions of the minimum MSE signal extraction filter of the work of McElroy and Maravall (2014) are satisfied for the following models: (i) Score-driven location model with constant scale; for this model, the statistical properties of the score functions are presented in the work of Harvey (2013, p. 61). (ii) Score-driven location model with score-driven scale; for this model, estimation results for United States (US) inflation rate data are presented in the work of Harvey (2013, p. 140). The statistical properties of the score functions of the score-driven location model with score-driven scale are shown in the present paper. (iii) Score-driven trend plus score-driven seasonality model with constant scale. (iv) Score-driven trend plus score-driven seasonality model with score-driven scale. Models (iii) and (iv) include a new score-driven seasonality specification, which is applied to the minimum MSE signal extraction filters.

For the empirical contribution, we use US inflation rate data for the period of January 1948 to May 2020, for which signal smoothing can be motivated by the following points:

First, the series derived from the inflation smoothing procedure could allow the private sector to infer the present and future stance of monetary policy in a better way than the official inflation series (which would be understood to contain more noise and less signal than the smoothed inflation series). In relation to the use of smoothed inflation in the private sector, we refer to the work of Erceg and Lewin (2003). Second, in central banks, it is common to monitor one or more measures of core inflation (Rich and Steindel 2005). The core inflation measure corrects the excessive volatility that the official inflation usually shows, since the latter is affected by some relative price changes that are not of direct interest to the monetary policy. There is a significant literature on alternative measures of core inflation, from which we refer to the works of Bryan and Pike (1991) and Cristadoro et al. (2005). In the present paper, a novel procedure to generate a measure of core inflation is presented.

The remainder of this paper is organized as follows: Section 2 reviews the literature. Section 3 presents the minimum MSE signal for correlated signal and noise. Section 4 presents the method of signal smoothing for score-driven location models for constant and score-driven scales. Section 5 presents the method of signal smoothing score-driven trend plus score-driven seasonality models for constant and score-driven scales. Section 6 presents the empirical application for the US inflation rate. Section 7 concludes. Supplementary Material presents technical details.

2. Review of the literature

2.1. Signal extraction

In the works of Bell (1984), and Bell and Hillmer (1988), minimum MSE signal extraction filters are presented, for which either signal or noise or both is non-stationary, and the signal and noise components are uncorrelated. In those works, signal extraction formulas are presented by using the transformation approach (Ansley and Kohn 1985), which eliminates the effects of the non-stationary initial conditions. In relation to this, we also refer to the works of Bell and Hillmer (1984), and Bell (2004). In the work of Bell (1984, p. 662), it is shown that minimum MSE signal extraction filters can be applied to non-Gaussian observations.

In the work of McElroy (2008), minimum MSE signal extraction filters are presented for finitely-sampled non-stationary ARIMA (autoregressive integrated moving average) processes. Signal smoothing formulas are provided, for which the updating terms of signal and noise, $\{u_t\}_{t=1}^T$ and $\{v_t\}_{t=1}^T$, respectively, are uncorrelated. It is shown that the minimum MSE signal extraction filter can be used for non-Gaussian observations (McElroy 2008, p. 991).

Several works perform signal smoothing for correlated signal and noise, for example, Beveridge and Nelson (1981), Snyder (1985), Ghysels (1987), Ord et al. (1997), Hyndman et al. (2002), Proietti (2006), and McElroy and Maravall (2014). From these works, McElroy and Maravall (2014) is an extension of the aforementioned works of Bell (1984), Bell and Hillmer (1988), and McElroy (2008). In the work of McElroy and Maravall (2014), minimum MSE signal extraction filters for correlated updating terms of signal and noise are presented, for which either signal or noise or both is non-stationary, with heteroskedastic and non-Gaussian error terms. The work of McElroy and Maravall (2014) is relevant to the present paper, because signal and noise are correlated, and the updating terms are non-Gaussian for the score-driven models. For all score-driven models of this paper, we show that u_t and v_t are contemporaneously uncorrelated, but for some lags or leads u_t and v_t are correlated.

2.2. *Score-driven time series models*

Score-driven models are named generalized autoregressive score (GAS) models (Creal et al. 2008), or dynamic conditional score (DCS) models (Harvey and Chakravarty 2008; Harvey 2013). Score-driven models are observation-driven time series models (Cox 1981), which are alternatives to several classical observation-driven time series models. For several score-driven models, the sufficient conditions of consistency and asymptotic normality of the ML estimates are known (Blasques et al. 2017). Score-driven models are applied to $I(0)$, co-integrated $I(1)$, and fractionally integrated variables. An advantage of score-driven models is that they are more robust to outliers and missing observations than classical time series models (Harvey 2013).

An example of score-driven models is the quasi-AR (QAR) location model (Harvey 2013), which is an alternative to the ARMA model (Box and Jenkins 1970). Another example of score-driven models is the Beta-t-EGARCH (exponential generalized autoregressive conditional heteroskedasticity) model (Harvey and Chakravarty 2008; Harvey 2013), which is an alternative to the GARCH (Engle 1982; Bollerslev 1986), and EGARCH (Nelson 1991) models.

Univariate score-driven models, such as Beta-t-EGARCH, implement an optimal filtering mechanism, according to the Kullback–Leibler divergence in favor of the true data-generating process. In the work of Blasques et al. (2015), it is shown that, asymptotically, a score-driven update of the time series model reduces the distance between the true conditional density and the conditional density that is implied by the score-driven model, in expectation and at every step, even for misspecified score-driven models. The authors show that only score-driven updates have this property, by providing an information-theoretic support for the use of score-driven models. The results of Blasques et al. (2015) are asymptotic results. Nevertheless, the recent work of Blasques et al. (2020) supports the information-theoretic effective filtering mechanism of score-driven volatility models for finite samples in practically significant cases.

3. Minimum MSE signal extraction filter for correlated signal and noise

In this section, we summarize the results of McElroy and Maravall (2014) for minimum MSE signal extraction filters for correlated signal and noise, which we use for score-driven models.

For variable y_t with $t = 1 \ldots, T$, the signal plus noise model is $y_t = s_t + n_t$, for which the updating terms are defined as follows:

$$u_t = \alpha_p(L)s_t = (1 - \alpha_1 L - \cdots - \alpha_p L^p)s_t \tag{3.1}$$

$$v_t = \beta_m(L)n_t = (1 - \beta_1 L - \cdots - \beta_m L^m)n_t \tag{3.2}$$

where the vectors $(\alpha_1, \ldots, \alpha_p)$ and $(\beta_1, \ldots, \beta_m)$ include time-invariant parameters, s_t represents the signal component, and n_t represents the noise component.

For the signal plus noise model of equations (3.1) and (3.2), the minimum MSE signal (or smoothed signal) is $\hat{s}_t = E(s_t|Y)$, where $Y \equiv (y_1, \ldots, y_T)'$. In the work of McElroy (2008), it is shown that $\hat{s}_t$ is also the minimum MSE signal when observations are generated from a distribution other than the normal distribution. That is the case of the score-driven models of our paper, in which the Student's t-distribution is used.

The Student's t-distribution generalizes the standard normal distribution, for which equations (3.1) and (3.2) are a linear Gaussian state space model (Harvey 1989), and it also provides robustness to outliers for the score-driven models. Furthermore, the Student's t-distribution is useful, because some expected value formulas in this paper, which are used for the computation of the minimum MSE signal, can be directly expressed from the model parameters.

First, we introduce notation in order to apply McElroy and Maravall (2014, Theorem 1) to signal smoothing of score-driven models. In matrix notation, u_t is $U = \Delta_S S$, where

$$\begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_{T-p} \end{bmatrix} = \begin{bmatrix} -\alpha_p & \cdots & \cdots & -\alpha_1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -\alpha_p & \ddots & \ddots & -\alpha_1 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -\alpha_p & \cdots & \cdots & -\alpha_1 & 1 \end{bmatrix} \begin{bmatrix} s_{-p+1} \\ \vdots \\ s_0 \\ s_1 \\ \vdots \\ s_{T-p} \end{bmatrix} \tag{3.3}$$

where U is a $(T-p)\times 1$ vector, Δ_S is a $(T-p)\times T$ matrix, and S is a $T\times 1$ vector (Bell and Hillmer, 1988). Moreover, in matrix notation (Bell and Hillmer 1988), v_t is $V=\Delta_N N$, where

$$\begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_{T-m} \end{bmatrix} = \begin{bmatrix} -\beta_m & \cdots & \cdots & -\beta_1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -\beta_m & \cdots & \cdots & -\beta_1 & 1 & 0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & -\beta_m & \cdots & \cdots & -\beta_1 & 1 \end{bmatrix} \begin{bmatrix} n_{-m+1} \\ \vdots \\ n_0 \\ n_1 \\ \vdots \\ n_{T-m} \end{bmatrix} \tag{3.4}$$

where V is a $(T-m)\times 1$ vector, Δ_N is a $(T-m)\times T$ matrix, and N is a $T\times 1$ vector.

Matrix $\underline{\Delta}_S$ with dimensions $(T-d)\times(T-m)$ is also used, where $d=p+m$, which is defined by the first $(T-d)$ rows and the first $(T-m)$ columns of Δ_S (McElroy and Maravall 2014). Matrix $\underline{\Delta}_N$ with dimensions $(T-d)\times(T-p)$ is also used, which is defined by the first $(T-d)$ rows and the first $(T-p)$ columns of Δ_N (McElroy and Maravall 2014). Furthermore, matrix $\Delta=\underline{\Delta}_S\Delta_N$ is also defined (McElroy and Maravall 2014).

The covariance matrices of U and V are C_U with dimensions $(T-p)\times(T-p)$, and C_V with dimensions $(T-m)\times(T-m)$, respectively. The covariance matrix of the elements of U and V, with dimensions $(T-p)\times(T-m)$ is $C_{U,V}$, and $C_{V,U}=C'_{U,V}$. Furthermore, the following matrix is included in the minimum MSE signal extraction formula (McElroy and Maravall 2014): $C_W\equiv\underline{\Delta}_S C_V \underline{\Delta}'_S+\underline{\Delta}_N C_U \underline{\Delta}'_N+\underline{\Delta}_N C_{U,V}\underline{\Delta}'_S+\underline{\Delta}_S C_{V,U}\underline{\Delta}'_N$. Covariance matrices C_U, C_V, C_W, $C_{V,U}$, and $C_{U,V}$ are available in closed form for all score-driven models of this paper.

Second, in the work of McElroy and Maravall (2014, Theorem 1) it is assumed that: (A1) $Y^*\equiv(y_1,\ldots,y_d)'$ are uncorrelated with u_t and v_t, for $t>d$. (A2) $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ are relatively prime polynomials, i.e. polynomials $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ share no common zeros. (A3) $\{u_t\}$ and $\{v_t\}$ have mean zero, are covariance stationary, are correlated, and are purely nondeterministic. (A4) C_U, C_V, and C_W are invertible. In Sections 4 and 5, we show that these assumptions are satisfied for the score-driven location, trend, seasonality, and scale models. Under assumptions

(A1) to (A4), the minimum MSE signal is $\hat{S} = FY$, where the $T \times T$ matrix F is:

$$F = M^{-1}(\Delta_N' C_V^{-1} \Delta_N + P C_W^{-1} \Delta) \tag{3.5}$$

$$M = \Delta_S' C_U^{-1} \Delta_S + \Delta_N' C_V^{-1} \Delta_N \tag{3.6}$$

$$P = \Delta_S' C_U^{-1} C_{U,V} \underline{\Delta}_S' - \Delta_N' C_V^{-1} C_{V,U} \underline{\Delta}_N' \tag{3.7}$$

All score-driven models of this paper can be written according to the representation of the dynamic equations (3.1) and (3.2). Therefore, if (A1) to (A4) hold, then the signal smoothing results of McElroy and Maravall (2014) can be applied to the score-driven models.

4. Score-driven location

4.1. Score-driven location with constant scale

In this section, we present minimum MSE signal extraction for the score-driven location model (Harvey 2013), where the error terms of the signal and noise components are correlated, the signal component s_t may be non-stationary, and the noise component n_t is independent and identically distributed (i.i.d.). The score-driven location model for $\{y_t\}_{t=1}^T$ is:

$$y_t = s_t + n_t = s_t + v_t \tag{4.1}$$

where $n_t = v_t = \exp(\lambda)\epsilon_t$, and $\epsilon_t \sim t(\nu)$ is an i.i.d. error term. We assume that $\nu > 2$ (hence, the second moment of ϵ_t exists). In this model, $\beta_m(L)$ is normalized to one. The log conditional density of $y_t|\mathcal{F}_{t-1} \equiv y_t|(y_1, \ldots, y_{t-1})$ is:

$$\ln f(y_t|\mathcal{F}_{t-1}, \Theta) = \tag{4.2}$$

$$\ln\Gamma\left(\frac{\nu+1}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda - \frac{\nu+1}{2}\ln\left\{1 + \frac{(y_t - s_t)^2}{\nu\exp(2\lambda)}\right\}$$

where $\ln(x)$ is the natural logarithm function, Θ represents the time-invariant parameters, $\Gamma(x)$ is the gamma function, and $\exp(x)$ is the exponential function.

For the $I(0)$ case, the signal is specified as a QAR(p) model (Harvey 2013, p. 63) as follows:

$$s_t = \alpha_1 s_{t-1} + \ldots + \alpha_p s_{t-p} + u_t = \alpha_1 s_{t-1} + \ldots + \alpha_p s_{t-p} + \psi_1 l_{t-1} \tag{4.3}$$

where the roots of $(1 - \alpha_1 z - \ldots - \alpha_p z^p) = 0$ lie outside the unit circle, and $E(s_t) = 0$. Alternatively, for the $I(1)$ case, the signal is specified as follows (Harvey 2013, p. 76):

$$s_t = s_{t-1} + u_t = s_{t-1} + \psi_1 l_{t-1} \tag{4.4}$$

where $E(s_t) = 0$; $\alpha_1 = 1$ and $\alpha_2, \ldots, \alpha_p$ are zeros. For all score-driven models of this paper, the signal is equal to the filtered signal, i.e. $s_t = E(s_t | y_1, \ldots, y_{t-1})$, because the signal is determined by the history of past observations of the dependent variable. Equations (4.3) and (4.4) indicate the elements of matrices Δ_S and $\underline{\Delta}_S$. The score function with respect to s_t is:

$$\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial s_t} = \frac{\nu + 1}{\nu \exp(2\lambda)} \times l_t = k \times l_t = k \times \left[1 + \frac{v_t^2}{\nu \exp(2\lambda)}\right]^{-1} v_t \tag{4.5}$$

$$= k \times \frac{\nu \exp(\lambda) \epsilon_t}{\nu + \epsilon_t^2}$$

where k is the scaling factor, and l_t is the scaled score function (Harvey 2013). Harvey (2013) shows that l_t is i.i.d. with $E(l_t) = 0$, and variance $\mathrm{Var}(l_t) = \nu^2 \exp(2\lambda)/[(\nu + 3)(\nu + 1)] < \infty$.

Score-driven models are estimated by using the ML method. The conditions of Harvey (2013), Creal et al. (2013), and Blasques et al. (2017, 2018) can be applied to the score-driven signal plus noise models with constant scale. We also refer to the paper of Blazsek et al. (2020), in which ML conditions are proven for score-driven location plus score-driven scale models, which can be applied to the score-driven signal plus noise models with score-driven scale.

4.2. Signal smoothing for score-driven location with constant scale

In this section, we prove that the assumptions of McElroy and Maravall (2014, Theorem 1) hold for the score-driven location with constant scale model: (A1) $d = p + m = p$. Initial values $Y^* = (y_1, \ldots, y_p)'$ are uncorrelated with u_t and v_t for $t > p$, because $(s_1, \ldots, s_p)'$ and $(l_1, \ldots, l_p)'$ are set to zero vectors. (A2) $m = 0$; therefore, polynomials $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ share no common zeros. (A3) $u_t = \psi_1 l_{t-1}$ and $v_t = \exp(\lambda)\epsilon_t$ have mean zero, are covariance stationary, because both l_t, and v_t are i.i.d. with zero mean; u_t and v_t are purely nondeterminstic due to the model formulation. (A4) Matrices $C_U = \{[\psi_1^2 \nu^2 \exp(2\lambda)]/[(\nu+3)(\nu+1)]\} \times I_{T-p}$ and $C_V = [\exp(2\lambda)\nu/(\nu-2)] \times I_T$ are invertible (Harvey 2013). C_W and M are invertible for the score-driven location model with constant scale. Matrix $C_{U,V}$ with dimensions $(T-p) \times T$ is:

$$C_{U,V} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ C & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & C & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C & 0 & \cdots & 0 \end{bmatrix} \tag{4.6}$$

where $C = \text{Cov}(\psi_1 l_t, v_t) = \psi_1 \nu \exp(2\lambda)/(\nu+1)$ (Harvey 2013). The location of C in the first column of $C_{U,V}$ is the second row. Time series $\{l_t\}$ and $\{v_t\}$ are both independent. Hence, all lags and leads of l_t and v_t are independent, and the remaining elements of $C_{U,V}$ are zero.

4.3. Score-driven location and score-driven scale

In this section, the score-driven location model (Harvey 2013) and the score-driven scale model (Harvey 2013) are combined into a score-driven location model with score-driven scale, for which we present the statistical properties of the score functions, and the minimum MSE signal extraction filter. The updating terms of signal and noise are correlated, the signal component may be non-stationary, and the zero mean noise component is conditionally heteroskedastic. The score-driven location model with score-driven scale for $\{y_t\}_{t=1}^T$ is:

$$y_t = s_t + n_t = s_t + v_t = s_t + \exp(\lambda_t)\epsilon_t \tag{4.7}$$

where s_t and n_t are signal and noise, respectively. In this model, $\beta_m(L)$ is normalized to one. We assume that $n_t|\mathcal{F}_{t-1} = v_t|\mathcal{F}_{t-1} \sim t[0, \exp(\lambda_t), \nu]$ is heteroskedastic with $\nu > 2$, where $\mathcal{F}_{t-1} = (y_1, \ldots, y_{t-1})$. Therefore, the standardized error term $\epsilon_t \sim t(\nu)$ is i.i.d., and it has the Student's t-distribution. The log conditional density of $y_t|\mathcal{F}_{t-1}$ is:

$$\ln f(y_t|\mathcal{F}_{t-1}, \Theta) = \tag{4.8}$$

$$\ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda_t - \frac{\nu+1}{2}\ln\left\{1 + \frac{(y_t - s_t)^2}{\nu \exp(2\lambda_t)}\right\}$$

The signal s_t is specified as in (4.3) or (4.4), and the score function is:

$$\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial s_t} = \frac{\nu+1}{\nu \exp(2\lambda_t)} \times l_t = k_t \times l_t = k_t \times \left[1 + \frac{v_t^2}{\nu \exp(2\lambda_t)}\right]^{-1} v_t \tag{4.9}$$

$$= k_t \times \frac{\nu \exp(\lambda_t)\epsilon_t}{\nu + \epsilon_t^2}$$

where k_t is the dynamic scaling factor, and l_t is the scaled score function. The difference between the score functions of (4.5) and (4.9) is that k_t and λ_t are constant in (4.5), but they are dynamic in (4.9). In Supplementary Material A, we show that l_t is white noise with zero mean.

The time-varying log-scale λ_t is specified as follows:

$$\lambda_t = a + b\lambda_{t-1} + cz_{t-1} = \frac{a}{1-b} + \sum_{j=0}^{\infty} cb^j z_{t-1-j} = \frac{a}{1-b} + \sum_{j=1}^{\infty} cb^{j-1} z_{t-j} \tag{4.10}$$

where $|b| < 1$, $E(\lambda_t) = a/(1-b)$, which is the Beta-t-EGARCH(1,1) model. The updating term z_{t-1}, named score function with respect to log-scale λ_t is:

$$z_t = \frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial \lambda_t} = \frac{(\nu+1)v_t^2}{\nu \exp(2\lambda_t) + v_t^2} - 1 = \frac{(\nu+1)\epsilon_t^2}{\nu + \epsilon_t^2} - 1 \tag{4.11}$$

where the scaling factor is normalized to one, as in the work of Harvey (2013, p. 99). In Supplementary Material A, we show that z_t is i.i.d. with zero mean.

4.4. Signal smoothing for score-driven location and score-driven scale

In this section, we prove that the assumptions of McElroy and Maravall (2014, Theorem 1) hold for the score-driven location and score-driven scale model: (A1) $d = p + m = p$. Initial values $Y^* = (y_1, \ldots, y_p)'$ is uncorrelated with u_t and v_t for $t > p$, because $(s_1, \ldots, s_p)'$, $(l_1, \ldots, l_p)'$, and $(z_1, \ldots, z_p)'$ are set to zero vectors, and all elements of $(\lambda_1, \ldots, \lambda_p)'$ are set to $E(\lambda_t) = a/(1-b)$. (A2) $m = 0$; thus, polynomials $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ share no common zeros. (A3) $u_t = \psi_1 l_{t-1}$ and $v_t = \exp(\lambda_t)\epsilon_t$ have mean zero, are covariance stationary, and are purely nondeterministic. Variables u_t and v_t have zero mean and are covariance stationary, due to the properties of the score functions; u_t and v_t are purely nondeterminstic due to the model formulation. (A4) $C_V = \sigma_V^2 \times I_T$ and $C_U = \psi_1^2\sigma_L^2 \times I_{T-p}$ are invertible, where σ_V^2 and σ_L^2, respectively, are:

(i) The conditional variance of $v_t|\mathcal{F}_{t-1}$ is

$$\text{Var}(v_t|\mathcal{F}_{t-1}) = E(v_t^2|\mathcal{F}_{t-1}) = \exp(2\lambda_t) \times \frac{\nu}{\nu - 2} \tag{4.12}$$

By using the law of iterated expectations,

$$\sigma_V^2 = \text{Var}(v_t) = E[\exp(2\lambda_t)] \times \frac{\nu}{\nu - 2} \tag{4.13}$$

where (Harvey 2013, p. 102):

$$E[\exp(2\lambda_t)] = \exp\left(\frac{2a}{1-b}\right) \prod_{j=1}^{\infty} \exp(-2cb^{j-1})\tilde{\beta}_\nu(2cb^{j-1}) \tag{4.14}$$

$$\tilde{\beta}_\nu(x) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{1+2r}{\nu+1+2r}\right) \frac{x^k(\nu+1)^k}{k!} \tag{4.15}$$

In (4.15), function $\tilde{\beta}_\nu(x)$ is named Kummer's confluent hypergeometric function.

(ii) The conditional variance of $l_t|\mathcal{F}_{t-1}$ is:

$$\text{Var}(l_t|\mathcal{F}_{t-1}) = E(l_t^2|\mathcal{F}_{t-1}) = \exp(2\lambda_t) \times \frac{\nu^2}{(\nu+3)(\nu+1)} \tag{4.16}$$

By using the law of iterated expectations,

$$\sigma_L^2 = \text{Var}(l_t) = E[\exp(2\lambda_t)] \times \frac{\nu^2}{(\nu+3)(\nu+1)} \tag{4.17}$$

where the expectation is given by (4.14) and (4.15). C_W and M are invertible for the score-driven location model with score-driven scale.

Variables $\{u_t\}$ and $\{v_t\}$ are correlated. Matrix $C_{U,V}$ with dimensions $(T-p)\times T$ is:

$$C_{U,V} = \begin{bmatrix} B & D_2 & \cdots & \cdots & \cdots & \cdots & D_T \\ C_0 & B & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_1 & C_0 & B & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_J & \cdots & C_1 & C_0 & B & D_2 & \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ C_0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & C_0 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C_0 & 0 & \cdots & 0 \end{bmatrix} \tag{4.18}$$

where $J = (T-p-2)$. Element $C_0 = \text{Cov}(\psi_1 l_t, v_t) = \psi_1 \nu E[\exp(2\lambda_t)]/(\nu+1)$, where (4.14) and (4.15) are used for the computation of $E[\exp(2\lambda_t)]$. In Supplementary Material B, C_0 is presented, and it is also shown that B, C_i for $i = 1, \ldots, J$, and D_j for $j = 2, \ldots, T$, are zeros.

5. Score-driven trend plus score-driven seasonality

5.1. Score-driven trend and seasonality with constant scale

For the score-driven location model, $\beta_m(L)$ is normalized to unity; hence, seasonality is not modeled in the noise component n_t. We extend the score-driven location model to the score-driven trend plus score-driven seasonality model, for which we present minimum MSE signal extraction in this paper.

In the score-driven trend plus score-driven seasonality model with constant scale for $\{y_t\}_{t=1}^T$, the error terms of signal and noise are correlated, and signal and noise may be non-stationary:

$$y_t = s_t + n_t = s_t + \rho_t + I_t \tag{5.1}$$

where s_t is the trend component, ρ_t is the seasonality component, and $I_t = \exp(\lambda)\epsilon_t$ is the irregular component. We assume that $\epsilon_t \sim t(\nu)$ is an i.i.d. error term with $\nu > 2$. Notation s_t is used for the trend component, because the objective of this model is to extract the trend from the observed seasonal data series. This implies that noise is the sum of the seasonal and irregular components, as suggested in the work of McElroy (2008).

The log conditional density of $y_t|\mathcal{F}_{t-1} = y_t|(y_1, \ldots, y_{t-1})$ is:

$$\ln f(y_t|\mathcal{F}_{t-1}, \Theta) = \tag{5.2}$$

$$\ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln(\pi\nu) - \lambda - \frac{\nu+1}{2}\ln\left\{1 + \frac{(y_t - s_t - \rho_t)^2}{\nu\exp(2\lambda)}\right\}$$

The signal is specified as in (4.3) or (4.4). In the work of McElroy (2008, p. 990), the methodology is described regarding how the signal plus noise model, $y_t = s_t + n_t$, can be applied to models with trend s_t plus seasonal ρ_t plus irregular I_t components. If the interest is to extract the trend component from the observed data, then n_t is defined as $n_t = \rho_t + I_t$.

We have studied the possibility of applying a score-driven seasonal component for ρ_t, as defined in the works of Harvey (2013, Section 3.6), and Harvey and Luati (2014), to the minimum MSE signal methods of McElroy (2008), and McElroy and Maravall (2014). However, the application proves to be not straightforward. Therefore, in this paper, we use a seasonality specification, which is simpler than the models of Harvey (2013), and Harvey and Luati (2014), but its application is more straightforward to minimum MSE signals.

Seasonality ρ_t is specified as a QAR(m) model with restricted parameters:

$$\rho_t = \beta_m \rho_{t-m} + \Psi_m l_{t-m} \tag{5.3}$$

where the period of seasonality is $m > 1$, $\beta_1, \ldots, \beta_{m-1}$ are zeros, $|\beta_m| < 1$, and $E(\rho_t) = 0$. Equation (5.3) indicates the elements of Δ_N and $\underline{\Delta}_N$. Condition $|\beta_m| < 1$, i.e. the covariance stationarity of ρ_t, is important for (A2) of McElroy and Maravall (2014, Theorem 1). The score function with respect to $(s_t + \rho_t)$ is:

$$\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial(s_t + \rho_t)} = \frac{\nu + 1}{\nu \exp(2\lambda)} \times l_t = k \times l_t = k \times \left[1 + \frac{I_t^2}{\nu \exp(2\lambda)}\right]^{-1} I_t \tag{5.4}$$

$$= k \times \frac{\nu \exp(\lambda)\epsilon_t}{\nu + \epsilon_t^2}$$

where k is the scaling factor, and l_t is the scaled score function. The difference between the score functions of (4.5) and (5.4) is that the derivative of the log-density in (4.5) is with respect to s_t and the derivative of the log-density in (5.4) is with respect to $(s_t + \rho_t)$. Scaled score function l_t is i.i.d. with $E(l_t) = 0$, and variance $\text{Var}(l_t) = \nu^2 \exp(2\lambda)/[(\nu + 3)(\nu + 1)] < \infty$.

5.2. Signal smoothing for score-driven trend and seasonality with constant scale

In this section, we prove that the assumptions of McElroy and Maravall (2014, Theorem 1) hold for the score-driven trend and seasonality with constant scale model: (A1) $d = p + m$. Initial values $Y^* = (y_1, \ldots, y_{p+m})'$ is uncorrelated with u_t and v_t for $t > p+m$, because $(s_1, \ldots, s_{p+m})'$, $(l_1, \ldots, l_{p+m})'$, and $(\rho_1, \ldots, \rho_{p+m})'$ are set to zero vectors. (A2) Since $|\beta_m| < 1$, polynomials $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ share no common zeros. (A3) The noise component is:

$$n_t = \beta_m n_{t-m} + (I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \beta_m n_{t-m} + v_t \tag{5.5}$$

Variables $u_t = \psi_1 l_{t-1}$ and $v_t = I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}$ have mean zero, are covariance stationary, and are purely nondeterministic. (A4) Matrix $C_U = \{[\psi_1^2 \nu^2 \exp(2\lambda)]/[(\nu + 3)(\nu + 1)]\} \times I_{T-p}$ is invertible (Harvey 2013). Matrix C_V with dimensions $(T - m) \times (T - m)$ is invertible, because C_V is a full-rank square matrix with elements:

$$\text{Var}(v_t) = \text{Var}(I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \frac{\nu \exp(2\lambda)}{\nu - 2} + \frac{\beta_m^2 \nu \exp(2\lambda)}{\nu - 2} + \frac{\Psi_m^2 \nu^2 \exp(2\lambda)}{(\nu + 3)(\nu + 1)} \tag{5.6}$$

$$\text{Cov}(v_t, v_{t-m}) = \text{Cov}(-\beta_m I_{t-m} + \Psi_m l_{t-m}, I_{t-m}) = -\frac{\beta_m \nu \exp(2\lambda)}{\nu - 2} + \frac{\Psi_m \nu \exp(2\lambda)}{\nu + 1} \tag{5.7}$$

where we use the following result from Harvey (2013, p. 62): $\text{Cov}(I_t, l_t) = \nu \exp(2\lambda)/(\nu + 1)$.

The remaining elements of C_V are zeros. In addition, C_W and M are invertible.

Matrix $C_{U,V}$ with dimensions $(T-p) \times (T-m)$ is:

$$C_{U,V} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & D_m & 0 & \cdots & 0 \\ C_0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & D_m \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & C_0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{5.8}$$

where $C_0 = \text{Cov}(\psi_1 l_t, v_t) = \psi_1 \nu \exp(2\lambda)/(\nu+1)$, and

$$D_m = \text{Cov}(\psi_1 l_{t-m}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) \tag{5.9}$$

$$= -\psi_1 \beta_m \frac{\nu \exp(2\lambda)}{\nu+1} + \psi_1 \Psi_m \frac{\nu^2 \exp(2\lambda)}{(\nu+3)(\nu+1)}$$

The location of C_0 in the first column of $C_{U,V}$ is the second row, and the location of D_m in the first row of $C_{U,V}$ is the m-th column. For the formulations of C_0 and D_m, results from the work of Harvey (2013) are used. The remaining elements of $C_{U,V}$ are zero, because l_t and I_t are independent time series, and l_t is a continuous function of only I_t.

5.3. Score-driven trend, seasonality, and scale

The score-driven trend plus score-driven seasonality model with score-driven scale is:

$$y_t = s_t + n_t = s_t + \rho_t + I_t \tag{5.10}$$

where $I_t = \exp(\lambda_t)\epsilon_t$ is the irregular component, and $\epsilon_t \sim t(\nu)$ is i.i.d. with $\nu > 2$. Filter s_t is

(4.3) or (4.4), ρ_t is (5.3), and the score function with respect to $(s_t + \rho_t)$ is:

$$\frac{\partial \ln f(y_t|\mathcal{F}_{t-1}, \Theta)}{\partial(s_t+\rho_t)} = \frac{\nu+1}{\nu \exp(2\lambda_t)} \times l_t = k_t \times l_t = k_t \times \left[1 + \frac{I_t^2}{\nu \exp(2\lambda_t)}\right]^{-1} I_t \tag{5.11}$$

$$= k_t \times \frac{\nu \exp(\lambda_t)\epsilon_t}{\nu + \epsilon_t^2}$$

where $\mathcal{F}_{t-1} = (y_1, \ldots, y_{t-1})$, k_t is the scaling factor, and l_t is the scaled score function l_t. The difference between (4.9) and (5.11) is that the derivative of the log-density in (4.9) is with respect to s_t and the derivative of the log-density in (5.11) is with respect to $(s_t + \rho_t)$. Scaled score function l_t is white noise with zero mean. Variables λ_t and z_t are defined as in (4.10) and (4.11), respectively. Hence, z_t is i.i.d. with zero mean.

5.4. Signal smoothing for score-driven trend, seasonality, and scale

In this section, we prove that the assumptions of McElroy and Maravall (2014, Theorem 1) hold for the score-driven trend, seasonality, and scale model: (A1) $d = p + m$. Initial values $Y^* = (y_1, \ldots, y_{p+m})'$ is uncorrelated with u_t and v_t for $t > p+m$, because $(s_1, \ldots, s_{p+m})'$, $(l_1, \ldots, l_{p+m})'$, and $(\rho_1, \ldots, \rho_{p+m})'$ are set to zero vectors, and each element of $(\lambda_1, \ldots, \lambda_{p+m})'$ is set to $E(\lambda_t) = a/(1-b)$. (A2) Since $|\beta_m| < 1$, polynomials $\alpha_p(\cdot)$ and $\beta_m(\cdot)$ share no common zeros. (A3) $u_t = \psi_1 l_{t-1}$ and $v_t = I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}$ have mean zero, are covariance stationary, and are purely nondeterministic. Variables u_t and v_t have mean zero and are covariance stationary due to the properties of the score functions. (A4) Matrix $C_U = \{\{\psi_1^2 \nu^2 E[\exp(2\lambda_t)]\}/[(\nu+3)(\nu+1)]\} \times I_{T-p}$ is invertible (Harvey 2013). Matrix C_V with dimensions $(T-m) \times (T-m)$ is invertible, because it is a full-rank square matrix, with elements (Harvey 2013):

$$\text{Var}(v_t) = \text{Var}(I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) \tag{5.12}$$

$$= \frac{\nu E[\exp(2\lambda_t)]}{\nu-2} + \frac{\beta_m^2 \nu E[\exp(2\lambda_t)]}{\nu-2} + \frac{\Psi_m^2 \nu^2 E[\exp(2\lambda_t)]}{(\nu+3)(\nu+1)}$$

$$\text{Cov}(v_t, v_{t-m}) = \text{Cov}(-\beta_m I_{t-m} + \Psi_m l_{t-m}, I_{t-m}) \tag{5.13}$$

$$= -\frac{\beta_m \nu E[\exp(2\lambda_t)]}{\nu - 2} + \frac{\Psi_m \nu E[\exp(2\lambda_t)]}{\nu + 1}$$

where $E[\exp(2\lambda_t)]$ is given by (4.14) and (4.15). Matrices C_W and M are invertible for the score-driven trend, seasonality, and scale model. In Supplementary Material C, we show that $C_{U,V}$ with dimensions $(T-p) \times (T-m)$ is:

$$C_{U,V} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & D_m & 0 & \cdots & 0 \\ C_0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & D_m \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & C_0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{5.14}$$

where $C_0 = \text{Cov}(\psi_1 l_t, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \psi_1 \nu E[\exp(2\lambda_t)]/(\nu + 1)$, and

$$D_m = \text{Cov}(\psi_1 l_{t-m}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) \tag{5.15}$$

$$= -\psi_1 \beta_m \frac{\nu E[\exp(2\lambda)]}{\nu + 1} + \psi_1 \Psi_m \frac{\nu^2 E[\exp(2\lambda)]}{(\nu + 3)(\nu + 1)}$$

The location of C_0 in the first column of $C_{U,V}$ is the second row, and the location of D_m in the first row of $C_{U,V}$ is the m-th column.

6. Empirical application

6.1. Trend extraction from seasonally adjusted US inflation rate

Monthly data from the US inflation rate $100 \times \ln(\text{CPI}_t/\text{CPI}_{t-1})$ (consumer price index, CPI) are used for the period of January 1948 to May 2020, and the pre-sample period is December 1947 that we use for CPI_0. The source of the seasonally adjusted CPI data is Federal Research Economic Data (FRED) (ticker: CPIAUCSL). Smoothed signal estimation for the US inflation rate is relevant for policymakers at the Federal Reserve.

The score-driven models are estimated for the seasonally adjusted US inflation rate minus its sample mean, by using the model $y_t = s_t + n_t$, which ensures that $E(y_t) = 0$ under the assumption that US inflation rate is $I(0)$. In practice, monthly US inflation rate signals can be obtained after the smoothing procedures, by adding the sample average of monthly US inflation rate to the estimates of s_t and $\hat{s}_t$.

In Panel (a) of Table 1, the descriptive statistics of US inflation rate are presented. As data are seasonally adjusted, we assume that the information that is needed for a policy decision is the trend component of the inflation rate, not the seasonal component.

In Panel (b) of Table 1, the ML parameter estimates of the QAR(1), QAR(2), and QAR(3) score-driven location specifications with constant and score-driven scales are presented. We use different values of p, in order to support a correct model specification (Supplementary Material A). In the work of Harvey (2013, p. 75), the use of the likelihood-based model selection criteria is suggested for score-driven models. Therefore, we compare statistical performances by using the Bayesian information criterion (BIC).

The score-driven location specifications with score-driven scale are superior to the score-driven location specifications with constant scale (Table 1). Moreover, the BIC metric suggests the use of the QAR(2) specification for each location model. The use of QAR(2) is also motivated by the facts that α_2 is not significant for the score-driven QAR(3) location model, and α_2 and α_3 are not significant for the score-driven QAR(3) location plus score-driven scale model.

In Figure 1, for the QAR(2) location model with constant scale, we present the evolution of the seasonally adjusted US inflation rate y_t, its filtered signal s_t, and its smoothed signal $\hat{s}_t$. In Figure 2, for the QAR(2) location model with score-driven scale, we present the evolution of the seasonally adjusted US inflation rate y_t, its filtered signal s_t, its smoothed signal $\hat{s}_t$, and the time-varying log-scale parameter λ_t. The figures indicate that the smoothed signal estimates are similar for the score-driven location models with constant and score-driven scale parameters, which shows the robustness of the two-step signal smoothing procedure.

[APPROXIMATE LOCATION OF TABLE 1 AND FIGURES 1 AND 2]

6.2. Trend extraction from not seasonally adjusted (NSA) US inflation rate

For the score-driven location model $\beta_m(L) = 1$, hence, seasonality is not modeled in the noise component of the score-driven location models. Nevertheless, it may be the case in practice that trend extraction from a NSA macroeconomic variable is needed for an economic decision. We present an application of the score-driven trend plus score-driven seasonality model to the NSA US inflation rate (FRED ticker: CPIAUCNS) for the period of January 1948 to May 2020.

Similarly to the application on the seasonally adjusted US inflation rate, signal smoothing is performed for the NSA US inflation rate minus its sample mean, which defines y_t in this section. We assume that the period of seasonality is $m = 12$, which is supported by the local maximum points of the sample periodogram for the NSA US inflation rate (Figure 3).

In Panel (a) of Table 2, the descriptive statistics are presented, from which a relevant result is that NSA US inflation rate is $I(0)$. Motivated by this result, different covariance stationary QAR(p) specifications (equation (4.3)) are used.

In Panel (b) of Table 2, the ML estimates are presented for the score-driven trend plus score-driven seasonality model with constant scale and score-driven scale for the QAR(1) plus QAR(12) and QAR(2) plus QAR(12) alternatives. QAR lag selection is done using the BIC metric, which suggests the use of QAR(1) plus QAR(12). The use of the optimal p supports the assumption on the dynamically complete density function.

In Figure 4, the evolution of the NSA US inflation rate y_t, its filtered signal s_t, its smoothed signal $\hat{s}_t$, and the seasonality component ρ_t, for $t = 1, \ldots, T$, for the score-driven trend and seasonality model with constant scale, QAR(1) plus QAR(12), is presented. In Figure 5, the evolution of the NSA US inflation rate y_t, its filtered signal s_t, its smoothed signal $\hat{s}_t$, the seasonality component ρ_t, and the time-varying log-scale parameter λ_t, for $t = 1, \ldots, T$, for the score-driven trend and seasonality model with score-driven scale, QAR(1) plus QAR(12), is presented. The annual seasonality component of the US inflation rate is significant with a time-varying amplitude (Figures 4(d) and 5(d)). The smoothed signal for the NSA US inflation rate (Figures 4(c) and 5(c)) is similar to the smoothed signal for the seasonally adjusted US inflation rate (see Figures 1(c) and 2(c)). This indicates the robustness of the signal smoothing procedure for score-driven models as applied to seasonal observable variables.

[APPROXIMATE LOCATION OF TABLE 2 AND FIGURES 3 TO 5]

7. Conclusions

Signal smoothing is important in statistical applications preceding economic decisions. Motivated by this, in recent decades minimum MSE signal extraction formulas have been developed for a great variety of signal plus noise models in the literature. We have applied results from the literature on signal extraction to score-driven models, to perform signal smoothing for score-driven location, trend, and seasonality models with constant and score-driven scale parameters. We have presented the methodology regarding how minimum MSE signal extraction is applied in a straightforward way to score-driven models with higher than first-order location, trend, and seasonality components with constant and score-driven scale parameters.

Real dataset-based results have indicated the robustness of the signal smoothing procedure for seasonally adjusted and NSA US inflation rate time series for the period of January 1948 to May 2020. Our results have suggested the practical use of the following two-step signal smoothing procedure for score-driven models: (i) The score-driven model are estimated by using the ML method. (ii) The ML estimates are substituted into the minimum MSE signal extraction filter. The computer codes are available for practitioners from the authors upon request. The results have indicated that the novel signal smoothing procedure of this paper can be easily applied to complex score-driven signal plus noise models, in order to perform signal smoothing before economic decisions.

Acknowledgments

The authors are thankful for the helpful comments of Juan Carlos Castañeda Fuentes, Matthew Copley, and Serena Ng. Funding from Universidad Francisco Marroquín is acknowledged.

References

Ansley, Craig F. and Robert Kohn (1985) Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics* 13, 1286–1316.

Bell, William (1984) Signal extraction for nonstationary time series. *Annals of Statistics* 12, 646–664.

Bell, William (2004) On RegCOMPONENT time series models and their applications. In Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard (eds.), *State Space and Unobserved Component Models: Theory and Applications*, pp. 248–283. Cambridge: Cambridge University Press.

Bell, William R. and Steven C. Hillmer (1984) Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economics Statistics* 2, 291–320.

Bell, William and Steven Hillmer (1988) A matrix approach to likelihood evaluation and signal extraction for ARIMA component time series models. SRD Research Report No. RR-88/22, Bureau of the Census.

Beveridge, Stephen and Charles R. Nelson (1981) A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the 'business cycle'. *Journal of Monetary Economics* 7, 151–174.

Blasques, Francisco, Paolo Gorgi, Siem Jan Koopman, and Olivier Wintenberger (2018) Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12, 1019–1052.

Blasques, Francisco, Siem Jan Koopman, and Andre Lucas (2015) Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* 102, 325–343.

Blasques, Francisco, Siem Jan Koopman, and Andre Lucas (2017) Maximum likelihood estimation for score-driven models. TI 2014-029/III Tinbergen Institute Discussion Paper.

Blasques, Francisco, Andre Lucas, and Andries C. van Vlodrop (2020) Finite sample optimality of score-driven volatility models: some Monte Carlo evidence. *Econometrics and Statistics*.

Blazsek, Szabolcs, Alvaro Escribano, and Adrian Licht (2020) Prediction accuracy of bivariate score-driven risk premium and volatility filters: an illustration for the Dow Jones. Working Paper 2020-10, University Carlos III of Madrid, Department of Economics.

Bollerslev, Tim (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Box, George E. P. and Gwilym M. Jenkins (1970) *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Bryan, Michael F. and Christopher J. Pike (1991) Median price changes: an alternative approach to measuring current monetary inflation. *Economic Commentary*, 12.01.91, Reseach Department, Federal Reserve Bank of Cleveland.

Buccheri, Giuseppe, Giacomo Bormetti, Fulvio Corsi, and Fabrizio Lillo (2019) Filtering and smoothing with score-driven models. SSRN Electronic Journal.

Cox, David R. (1981) Statistical analysis of time series: some recent developments (with discussion and reply). *Scandinavian Journal of Statistics* 8, 93–115.

Creal, Drew, Siem Jan Koopman, and Andre Lucas (2008) A general framework for observation driven time-varying parameter models. Tinbergen Institute Discussion Paper 08-108/4.

Creal, Drew, Siem Jan Koopman, and Andre Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777–795.

Cristadoro, Riccardo, Mario Forni, Lucrezia Reichlin, and Giovanni Veronese (2005) A core inflation indicator for the euro area. *Journal of Money, Credit and Banking* 37, 539–560.

Dickey, David A. and Wayne A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Durbin, J., Koopman, S.J., 2012. *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.

Engle, Robert F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Erceg, Chritopher J. and Andrew T. Levin (2003) Imperfect credibility and inflation persistence. *Journal of Monetary Economics* 50, 915–944.

Ghysels, Eric (1987) Seasonal extraction in the presence of feedback. *Journal of Business and Economics Statistics* 5, 191–194.

Harvey, A.C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge.

Harvey, Andrew C. (2013) *Dynamic Models for Volatility and Heavy Tails.* Cambridge: Cambridge University Press.

Harvey, Andrew C. and Tirthankar Chakravarty (2008) Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge.

Harvey, Andrew and Alessandra Luati (2014) Filtering with heavy tails. *Journal of the American Statistical Association* 109, 1112–1122.

Hyndman, Rob J., Anne B. Koehler, Ralph D. Snyder, and Simone Grose (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18, 439–454.

Koopman, Siem Jan and Andrew C. Harvey (2003) Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control* 27, 1317–1333.

McElroy, Tucker (2008) Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* 24, 1–22.

McElroy, Tucker S. and Agustin Maravall (2014) Optimal signal extraction with correlated components. *Journal of Time Series Econometrics* 6, 237–273.

Nelson, Daniel B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.

Ord, Keith, Anne B. Koehler, and Ralph D. Snyder (1997) Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* 92, 1621–1629.

Proietti, Tommaso (2006) Trend-cycle decompositions with correlated components. *Econometric Reviews* 25, 61–84.

Rich, Robert W. and Charles Steindler (2005) A review of core inflation and an evaluation of its measures. FRB of New York Staff Report No. 236.

Shapiro, Samuel S. and Martin B. Wilk (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Snyder, Ralph D. (1985) Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society* 47, 272–276.

Table 1: Descriptive statistics of seasonally adjusted US inflation rate, and ML estimates for the score-driven location models.

(a). Descriptive statistics of monthly US inflation rate, $100 \times \ln(\text{CPI}_t/\text{CPI}_{t-1})$, seasonally adjusted			
Start Date	January 1948	Standard deviation	0.3327
End date	May 2020	Skewness	0.3929
T	869	Excess kurtosis	3.8868
Mean	0.2752	Shapiro–Wilk test statistic (p-value)	0.9370***(0.0000)
Median	0.2413	ADF test statistic, constant (p-value)	−4.3487***(0.0004)
Minimum	−1.7864	ADF test statistic, trend (p-value)	−4.4208***(0.0020)
Maximum	1.7938	ARCH test statistic (p-value)	208.5340***(0.0000)
(b). Parameter estimates (y_t is monthly US inflation rate minus sample mean)			
Score-driven location with constant scale		Score-driven location with score-driven scale	
QAR(1) specification		QAR(1) plus Beta-t-EGARCH(1,1) specification	
α_1	0.9677***(0.0140)	α_1	0.9761***(0.0121)
ψ	0.4296***(0.0736)	ψ	0.3035***(0.0530)
λ	−1.7474***(0.0426)	a	−0.1609**(0.0783)
ν	3.4709***(0.4228)	b	0.9027***(0.0453)
		c	0.1209***(0.0274)
		ν	5.3380***(0.9249)
BIC	−0.0167	BIC	−0.1413
Score-driven location with constant scale		**Score-driven location with score-driven scale**	
QAR(2) specification		**QAR(2) plus Beta-t-EGARCH(1,1) specification**	
α_1	0.5726***(0.0922)	α_1	0.4129**(0.1749)
α_2	0.3858***(0.0921)	α_2	0.5549***(0.1738)
ψ	0.5641***(0.0710)	ψ	0.4399***(0.0558)
λ	−1.7477***(0.0435)	a	−0.1527**(0.0742)
ν	3.5561***(0.4552)	b	0.9077***(0.0429)
		c	0.1217***(0.0277)
		ν	5.4228***(0.9909)
BIC	−0.0247	BIC	−0.1530
Score-driven location with constant scale		Score-driven location with score-driven scale	
QAR(3) specification		QAR(3) plus Beta-t-EGARCH(1,1) specification	
α_1	0.6450***(0.1812)	α_1	0.4649**(0.2207)
α_2	0.0292(0.2909)	α_2	0.3482(0.3801)
α_3	0.2820**(0.1404)	α_3	0.1545(0.1839)
ψ	0.5858***(0.0690)	ψ	0.4524***(0.0514)
λ	−1.7285***(0.0444)	a	−0.1384*(0.0773)
ν	3.8151***(0.5472)	b	0.9146***(0.0449)
		c	0.1145***(0.0289)
		ν	5.6994***(1.1649)
BIC	−0.0203	BIC	−0.1437

Notes: United States (US); maximum likelihood (ML); consumer price index (CPI); augmented Dickey–Fuller (ADF); autoregressive conditional heteroskedasticity (ARCH); Bayesian information criterion (BIC); quasi-autoregression (QAR); exponential generalized ARCH (EGARCH). The null hypothesis of the Shapiro–Wilk test (Shapiro and Wilk 1965) is normal distribution. For the ADF tests with constant and with constant plus linear time trend, BIC-based optimal lag selection is used. For the ARCH test, 5 lags are used. The Shapiro–Wilk test rejects normal distribution for US inflation rate. The ADF test (Dickey and Fuller 1979) indicates that US inflation rate is $I(0)$, motivating the use of (4.3). The ARCH test (Engle 1982) suggests significant volatility dynamics. For the parameter estimates, robust standard errors are reported in parentheses. The standard errors of ML parameters are estimated by using the Huber Sandwich Estimator (Blasques et al. 2017). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The best specification, according to BIC, is indicated by bold letters.

Table 2: Descriptive statistics of NSA US inflation rate, and ML estimates for the score-driven trend and seasonality models.

(a). Descriptive statistics of monthly US inflation rate, $100 \times \ln(\text{CPI}_t/\text{CPI}_{t-1})$, not seasonally adjusted			
Start Date	January 1948	Standard deviation	0.3770
End date	May 2020	Skewness	0.0437
T	869	Excess kurtosis	2.5195
Mean	0.2755	Shapiro–Wilk test statistic (p-value)	0.9650(0.0000)
Median	0.2792	ADF test statistic, constant (p-value)	−3.8284(0.0026)
Minimum	−1.9339	ADF test statistic, trend (p-value)	−3.8757(0.0130)
Maximum	1.7898	ARCH test statistic (p-value)	191.2120(0.0000)
(b). Parameter estimates and model diagnostics (Y_t is monthly US inflation rate minus sample mean)			
Score-driven trend plus seasonality with constant scale		**Score-driven trend plus seasonality with score-driven scale**	
QAR(1)-QAR(12)		**QAR(1)-QAR(12)-Beta-*t*-EGARCH(1,1)**	
α_1	0.9714***(0.0125)	α_1	0.9729***(0.0106)
ψ_1	0.3173***(0.0570)	ψ_1	0.2222***(0.0325)
β_{12}	0.9794***(0.0131)	β_{12}	0.9696***(0.0145)
Ψ_{12}	0.2086***(0.0388)	Ψ_{12}	0.1596***(0.0253)
λ	−1.5473***(0.0404)	a	−0.0012(0.0014)
ν	3.7753***(0.4619)	b	0.9967***(0.0027)
		c	0.0712***(0.0117)
		ν	7.4370***(1.2212)
BIC	0.3482	BIC	0.2111
Score-driven trend plus seasonality with constant scale		Score-driven trend plus seasonality with score-driven scale	
QAR(2)-QAR(12)		QAR(2)-QAR(12)-Beta-*t*-EGARCH(1,1)	
α_1	0.7598***(0.0840)	α_1	0.6629***(0.0764)
α_2	0.2436***(0.0788)	α_2	0.3048***(0.0777)
ψ_1	0.3855***(0.0616)	ψ_1	0.2849***(0.0467)
β_{12}	0.9800***(0.0127)	β_{12}	0.9694***(0.0137)
Ψ_{12}	0.2066***(0.0389)	Ψ_{12}	0.1601***(0.0240)
λ	−1.5478***(0.0403)	a	−0.0012(0.0016)
ν	3.8024***(0.4670)	b	0.9966***(0.0036)
		c	0.0726***(0.0178)
		ν	7.6265***(1.8434)
BIC	0.3509	BIC	0.2116

Notes: Not seasonally adjusted (NSA); United States (US); maximum likelihood (ML); consumer price index (CPI); augmented Dickey–Fuller (ADF); autoregressive conditional heteroskedasticity (ARCH); Bayesian information criterion (BIC); exponential generalized ARCH (EGARCH); quasi-autoregression (QAR). The null hypothesis of the Shapiro–Wilk test is normal distribution. For the ADF tests with constant and with constant plus linear time trend, BIC-based optimal lag selection is used. For the ARCH test, 5 lags are used. For the parameter estimates, robust standard errors are reported in parentheses. The standard errors of ML parameters are estimated by using the Huber Sandwich Estimator. ** and *** indicate significance at the 5% and 1% levels, respectively. The best specification, according to BIC, is indicated by bold letters.

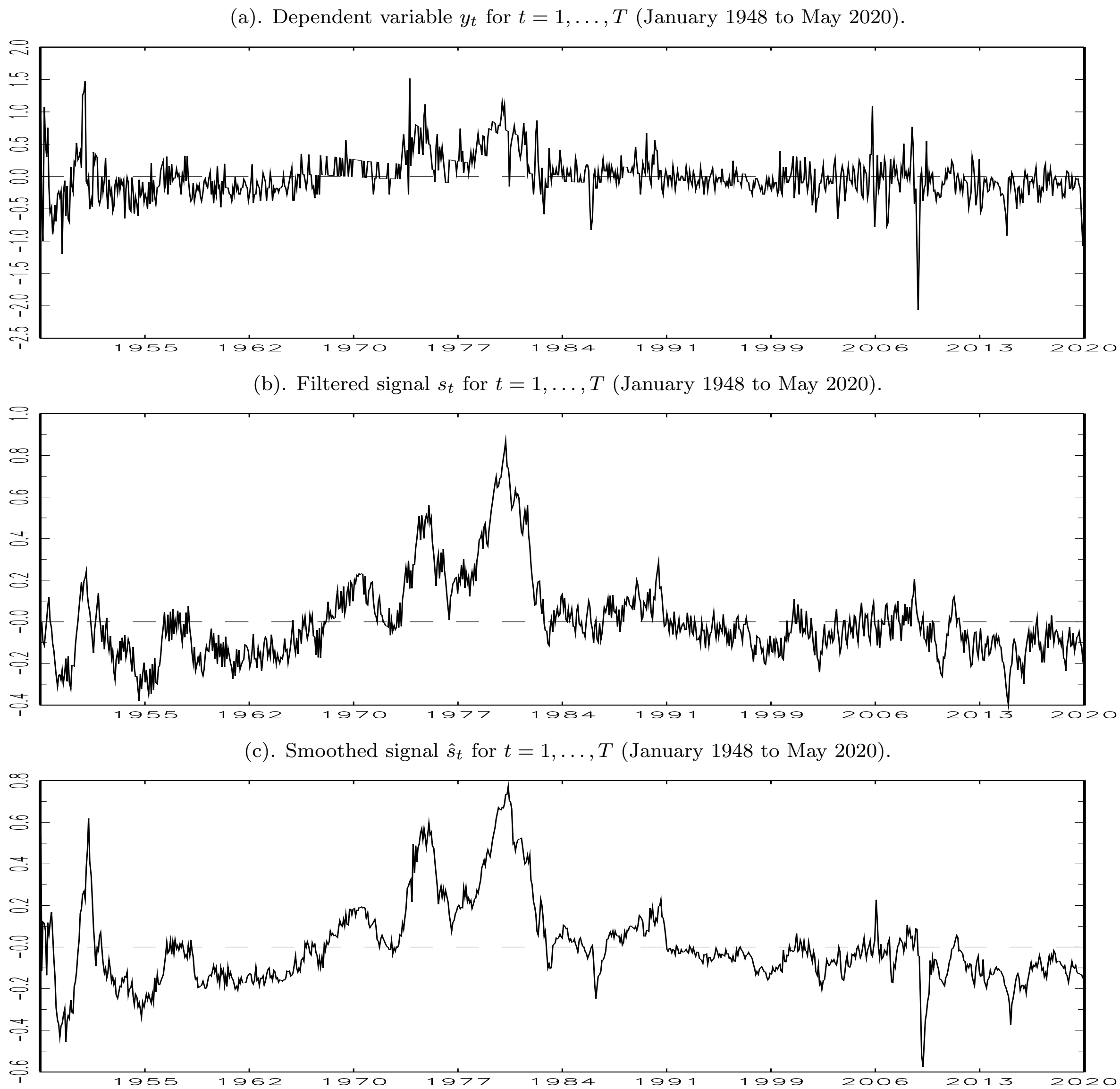


Figure 1: Seasonally adjusted US inflation for score-driven location (constant scale), QAR(2).

Notes: y_t is monthly seasonally adjusted US inflation rate minus its sample mean.

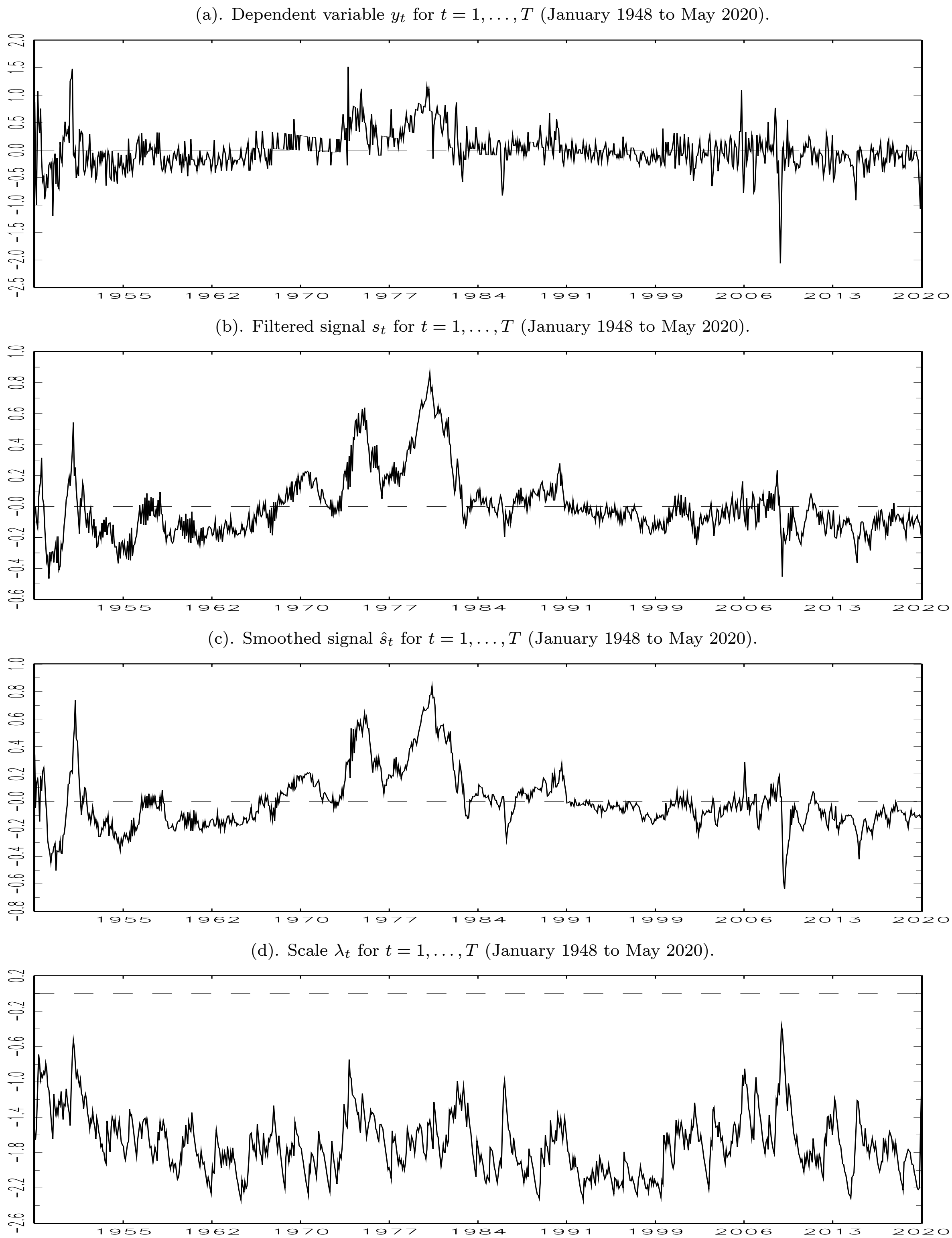


Figure 2: Seasonally adjusted US inflation for score-driven location (score-driven scale), QAR(2)-Beta-t-EGARCH(1,1).
Notes: y_t is monthly seasonally adjusted US inflation rate minus its sample mean.

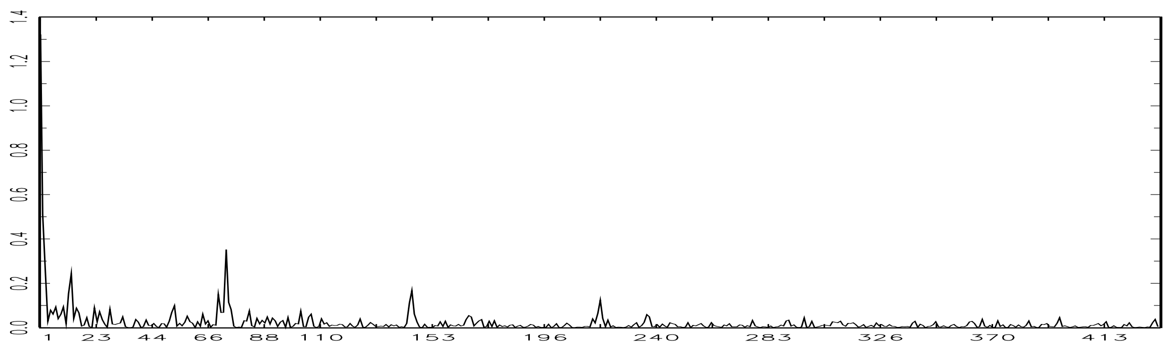
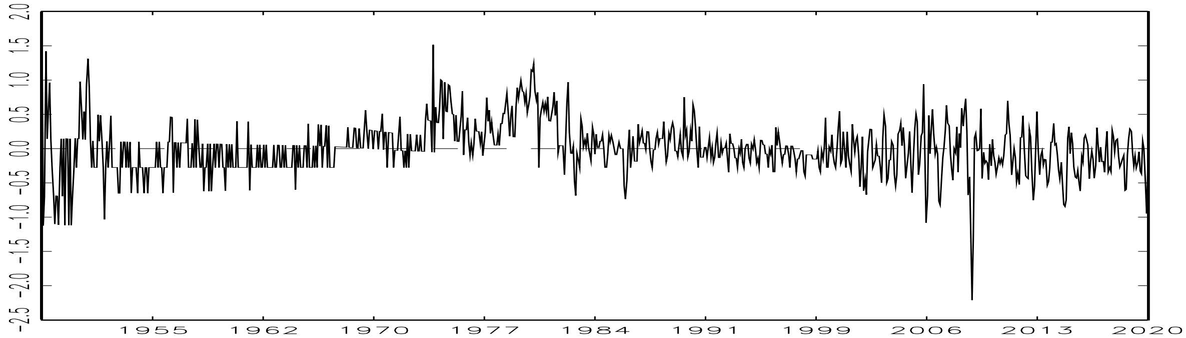
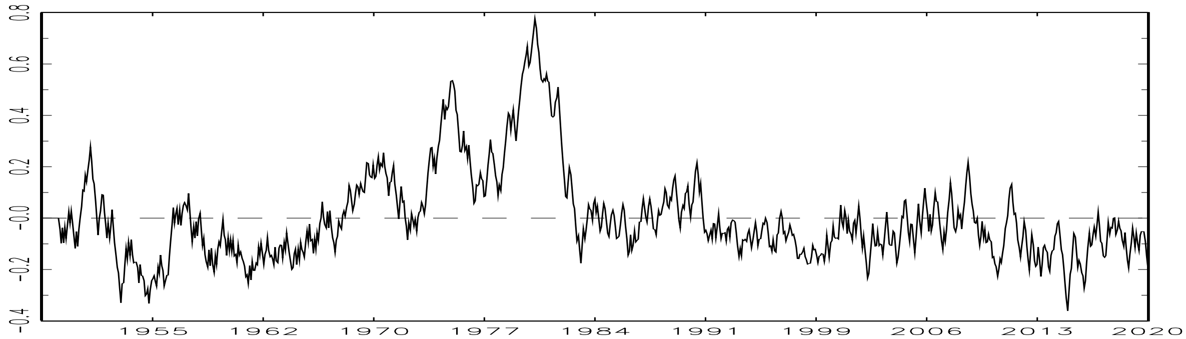


Figure 3: Periodogram for monthly NSA US inflation rate.
Notes: The x-axis corresponds to Fourier frequencies $2\pi j/T$ for $j = 1, \ldots, T/2$, and the ticks on the x-axis correspond to j. Three local maximum values correspond to $j = 73$, 145, and 218. These values divided by $T = 869$ are 0.0840, 0.1669, and 0.2509, respectively. As monthly data are used, we expect that the peaks in the periodogram shall occur close to the frequencies $2\pi k/12$ for $k = 1, 2, \ldots, 6$. The values of $k/12$ for $k = 1$, 2, and 3 are 0.0833, 0.1667, and 0.2500, respectively, which are in close correspondence to the local maximum-based estimates of j/T, and support the use of annual seasonality ($m = 12$).

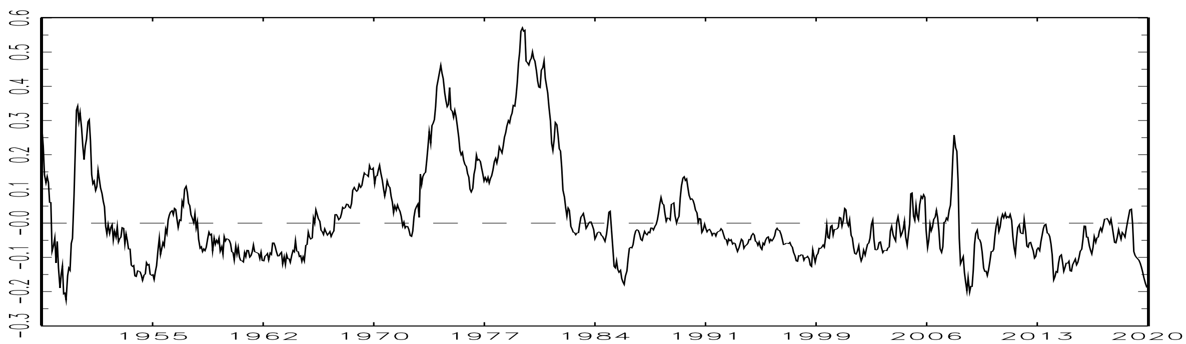
(a). Dependent variable y_t for $t = 1, \ldots, T$ (January 1948 to May 2020).



(b). Filtered signal s_t for $t = 1, \ldots, T$ (January 1948 to May 2020).



(c). Smoothed signal $\hat{s}_t$ for $t = 1, \ldots, T$ (January 1948 to May 2020).



(d). Seasonality ρ_t for $t = 1, \ldots, T$ (January 1948 to May 2020).

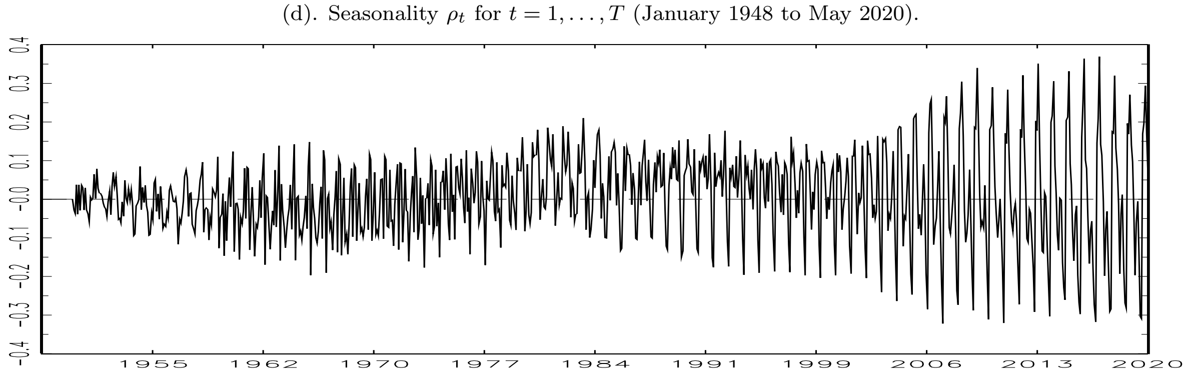
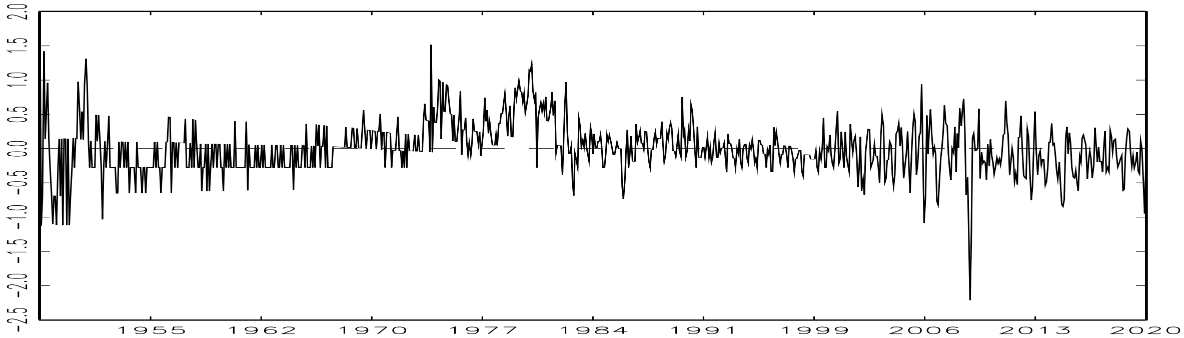
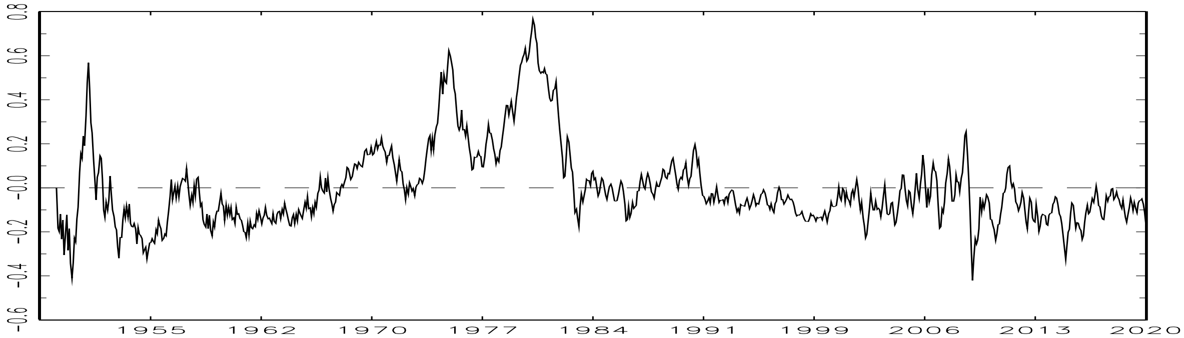


Figure 4: NSA US inflation for score-driven trend and seasonality (constant scale), QAR(1)-QAR(12).
Notes: y_t is monthly NSA US inflation rate minus its sample mean.

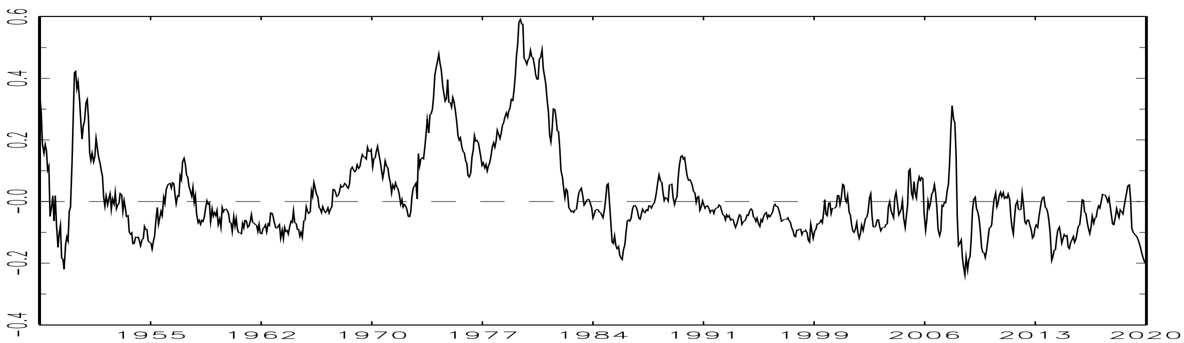
(a). Dependent variable y_t for $t = 1, \ldots, T$ (January 1948 to May 2020).



(b). Filtered signal s_t for $t = 1, \ldots, T$ (January 1948 to May 2020).



(c). Smoothed signal $\hat{s}_t$ for $t = 1, \ldots, T$ (January 1948 to May 2020).



(d). Seasonality ρ_t for $t = 1, \ldots, T$ (January 1948 to May 2020).

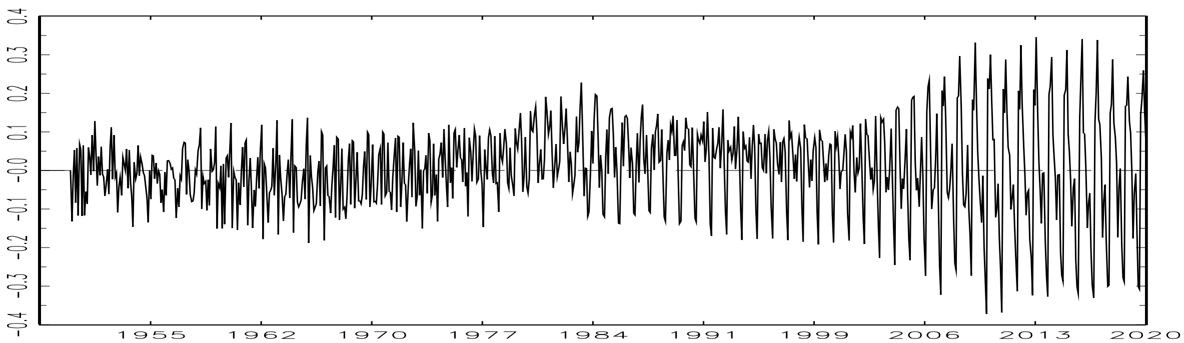


Figure 5(a)-(d): NSA US inflation for score-driven trend and seasonality (score-driven scale), QAR(1)-QAR(12)-Beta-t-EGARCH(1,1).
Notes: y_t is monthly NSA US inflation rate minus its sample mean.

(e). Scale λ_t for $t = 1, \dots, T$ (January 1948 to May 2020).

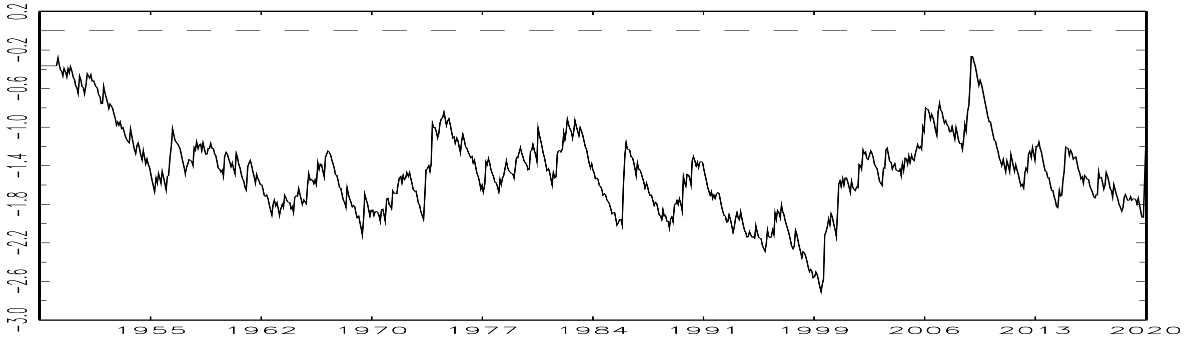


Figure 5(e): NSA US inflation for score-driven trend and seasonality (score-driven scale), QAR(1)-QAR(12)-Beta-t-EGARCH(1,1).
Notes: y_t is monthly NSA US inflation rate minus its sample mean.

Supplementary Material:

Optimal signal extraction for score-driven models

Astrid Ayala, Szabolcs Blazsek*, and Adrian Licht

School of Business, Universidad Francisco Marroquín, Guatemala City

Abstract: In this paper, a novel approach of signal smoothing for score-driven models is suggested, by using results from the literature on minimum mean squared error (MSE) signals. The new smoothing procedure can be applied to more general score-driven models than the state space smoothing recursions procedures from the literature. Score-driven location, trend, and seasonality models with constant and score-driven scale parameters for macroeconomic variables are used. The two-step smoothing procedure is computationally fast, and it uses closed-form formulas for smoothed signals. In the first step, the score-driven models are estimated by using the maximum likelihood (ML) method. In the second step, the ML estimates of the score functions are substituted into the minimum MSE signal extraction filter. Applications for monthly data of the seasonally adjusted and the not seasonally adjusted (NSA) United States (US) inflation rate variables for the period of 1948 to 2020 are presented.

Keywords: Signal extraction, Minimum mean squared error (MSE) signals, Dynamic conditional score (DCS), Generalized autoregressive score (GAS)

*School of Business, Universidad Francisco Marroquín, Calle Manuel F. Ayau, Guatemala City, 01010, Guatemala.

E-mails: aayala@ufm.edu (Ayala), sblazsek@ufm.edu (Blazsek), adrian.licht@ufm.edu (Licht)

Supplementary Material A

For the score-driven models of this paper, the scaled score function with respect to location l_t, and the score function with respect to log-scale z_t have the following properties: (i) We assume that $f(y_t|\mathcal{F}_{t-1},\Theta)$, where $\mathcal{F}_{t-1}=(y_1,\ldots,y_{t-1})$ and Θ represents the time-invariant parameters, is a correctly specified conditional density (Wooldridge 1994). Therefore,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial\Theta'}\right]=E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial s_t}\right]\times\frac{\partial s_t}{\partial\Theta'}=0 \tag{A.1}$$

where index $t-1$ indicates expectations that are conditional on $\mathcal{F}_{t-1}$. Since $\partial s_t/\partial\Theta'\neq 0$,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial s_t}\right]=E_{t-1}\left[\frac{\nu+1}{\nu\exp(2\lambda_t)}l_t\right]=E_{t-1}(l_t)\frac{\nu+1}{\nu\exp(2\lambda_t)}=0 \tag{A.2}$$

As a consequence, $E_{t-1}(l_t)=0$, i.e. l_t is a martingale difference sequence (MDS). Moreover,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial\Theta'}\right]=E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \lambda_t}\right]\times\frac{\partial \lambda_t}{\partial\Theta'}=0 \tag{A.3}$$

Since $\partial\lambda_t/\partial\Theta'\neq 0$,

$$E_{t-1}\left[\frac{\partial \ln f(y_t|\mathcal{F}_{t-1},\Theta)}{\partial \lambda_t}\right]=E_{t-1}(z_t)=0 \tag{A.4}$$

Thus, z_t is a MDS. (ii) $E(l_t)=0$ and $E(z_t)=0$, due to the law of iterated expectations. (iii) l_t and z_t are contemporaneously correlated, as both are functions of v_t. (iv) Scaled score function l_t is not i.i.d., as it depends on λ_t. (v) We assume that $|\lambda_t|<\lambda_{\max}<\infty$ for all t (Blazsek et al. 2020), which sets an exogenous bound for dynamic scale. The consequence of this assumption is that l_t is a bounded function of ϵ_t (Blazsek et al. 2020). Therefore, $\text{Var}(l_t)<\infty$, and l_t is white noise. If the roots of $(1-\alpha_1 z-\ldots-\alpha_p z^p)=0$ lie outside the unit circle, then s_t is covariance stationary for (4.3). (vi) z_t is a bounded function of ϵ_t (Harvey 2013). Therefore, $\text{Var}(z_t)<\infty$, and z_t is white noise. If $|b|<1$, then λ_t is covariance stationary. (vii) Due to

$|\lambda_t| < \infty$, $\partial l_t/\partial\lambda_t$ and $\partial z_t/\partial s_t$ are bounded functions of ϵ_t (Blazsek et al. 2020). (viii) $\partial l_t/\partial s_t$ and $\partial z_t/\partial\lambda_t$ are bounded functions of ϵ_t (Blazsek et al. 2020). (ix) Scaled score function l_t is an $\mathcal{F}$-measurable function of ϵ_t (White 2001), because l_t is a continuous function of ϵ_t. Scaled score function l_t is strictly stationary and ergodic, because l_t is an $\mathcal{F}$-measurable function of $(\epsilon_1,\ldots,\epsilon_t)$, and because ϵ_t is strictly stationary and ergodic (White 2001). (x) Score function z_t is i.i.d., because z_t is a continuous function of ϵ_t, and because ϵ_t is i.i.d. (White 2001). (xi) Score function z_t is an $\mathcal{F}$-measurable function of ϵ_t (White 2001), because z_t is a continuous function of ϵ_t (Harvey 2013). Score function z_t is strictly stationary and ergodic, because z_t is an $\mathcal{F}$-measurable function of ϵ_t, and because ϵ_t is strictly stationary and ergodic (White 2001).

Supplementary Material B

Matrix $C_{U,V}$ with dimensions $(T-p)\times T$ is represented as:

$$C_{U,V} = \begin{bmatrix} B & D_2 & \cdots & \cdots & \cdots & \cdots & D_T \\ C_0 & B & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_1 & C_0 & B & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_J & \cdots & C_1 & C_0 & B & D_2 & \vdots \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ C_0 & 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & C_0 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C_0 & 0 & \cdots & 0 \end{bmatrix} \tag{B.1}$$

where $J = (T-p-2)$. In the following, we formulate the elements B, C_0, C_i for $i = 1,\ldots,J$, and D_j for $j = 2,\ldots,T$.

First, $B = \text{Cov}(\psi_1 l_{t-1}, v_t) = \text{Cov}[\psi_1 l_{t-1}, \exp(\lambda_t)\epsilon_t] = 0$, because v_t is a MDS:

$$E(v_t|\mathcal{F}_{t-1}) = E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-1}] = \exp(\lambda_t)E[\epsilon_t|\mathcal{F}_{t-1}] = 0 \tag{B.2}$$

Therefore, $E(l_{t-1}v_t|\mathcal{F}_{t-1}) = l_{t-1}E(v_t|\mathcal{F}_{t-1}) = 0$, hence, $E(v_t l_{t-1}) = 0$. Second, $C_0 = \text{Cov}(\psi_1 l_t, v_t) = \psi_1\nu E[\exp(2\lambda_t)]/(\nu+1)$, where (4.14) and (4.15) are used for the computation of $E[\exp(2\lambda_t)]$. Third, $C_i = \text{Cov}[\psi_1 l_{t+i}, \exp(\lambda_t)\epsilon_t] = E[\psi_1 l_{t+i}\exp(\lambda_t)\epsilon_t]$ for $1 \le i \le J$, for which we use the following conditional expectation:

$$E[\psi_1 l_{t+i}\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t+i-1}] = \psi_1 \underbrace{E[l_{t+i}|\mathcal{F}_{t+i-1}]}_{0}\exp(\lambda_t)\epsilon_t = 0 \tag{B.3}$$

which is true because l_t is a MDS. Hence, $C_i = 0$ due to the law of iterated expectations. Fourth, $D_j = \mathrm{Cov}[\psi_1 l_{t-j}, \exp(\lambda_t)\epsilon_t] = 0$ for $j = 2, \ldots, T$, because of the following arguments. We use the following conditional expectation:

$$E[\psi_1 l_{t-j}\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-j}] = \psi_1 l_{t-j} E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-j}] \tag{B.4}$$

We use the law of iterated expectations (White 2001) as follows:

$$E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-j}] = E\{E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-j}\} = E\{\exp(\lambda_t)\underbrace{E[\epsilon_t|\mathcal{F}_{t-1}]}_{0}|\mathcal{F}_{t-j}\} = 0 \tag{B.5}$$

which holds because $E(|v_t|) < \infty$ (White 2001). We use (B.5), and the law of iterated expectations for (B.4), and we obtain that $D_j = 0$.

Supplementary Material C

Matrix $C_{U,V}$ with dimensions $(T-p) \times (T-m)$ is represented as:

$$C_{U,V} = \begin{bmatrix} B & D_2 & \cdots & \cdots & D_{T-p} & \cdots & \cdots & D_m & \cdots & \cdots & D_{T-m} \\ C_0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & D_m \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ C_J & \cdots & \cdots & \cdots & C_1 & C_0 & B & D_2 & \cdots & \cdots & D_{T-p} \end{bmatrix} \tag{C.1}$$

where $J = (T-p-2)$. In the following, we formulate the elements B, C_0, C_i for $i = 1, \ldots, J$, and D_j for $j = 2, \ldots, T-m$, respectively:

First, $B = \text{Cov}(\psi_1 l_{t-1}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = 0$ for $m > 1$, because:

$$\text{Cov}(\psi_1 l_{t-1}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \psi_1 \underbrace{E(l_{t-1} I_t)}_{\text{(i)}} - \psi_1 \beta_m \underbrace{E(l_{t-1} I_{t-m})}_{\text{(ii)}} + \psi_1 \Psi_m \underbrace{E(l_{t-1} l_{t-m})}_{\text{(iii)}} = 0 \tag{C.2}$$

In the following, we present (i), (ii), and (iii):

(i) I_t is a MDS, because

$$E(I_t|\mathcal{F}_{t-1}) = E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-1}] = \exp(\lambda_t)E[\epsilon_t|\mathcal{F}_{t-1}] = \exp(\lambda_t)E(\epsilon_t) = 0 \tag{C.3}$$

Hence, $E(l_{t-1}I_t|\mathcal{F}_{t-1}) = l_{t-1}E(I_t|\mathcal{F}_{t-1}) = 0$, and $E(l_{t-1}I_t) = 0$.

(ii) We write the following conditional mean:

$$E(l_{t-1}I_{t-m}|\mathcal{F}_{t-m}) = E(l_{t-1}|\mathcal{F}_{t-m})I_{t-m} = 0 \tag{C.4}$$

We use the law of iterated expectations:

$$E(l_{t-1}|\mathcal{F}_{t-m}) = E[E(l_{t-1}|\mathcal{F}_{t-2})|\mathcal{F}_{t-m}] = E[0|\mathcal{F}_{t-m}] = 0 \tag{C.5}$$

which holds due to the MDS property of l_t, and $E(|l_{t-1}|) < \infty$ (White 2001). By using the law of iterated expectations, $E(l_{t-1}I_{t-m}) = 0$.

(iii) $E(l_{t-1}l_{t-m})$ is zero because l_t is a MDS with finite variance.

Second, $C_0 = \text{Cov}(\psi_1 l_t, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \psi_1 \nu E[\exp(2\lambda_t)]/(\nu + 1)$, where $E[\exp(2\lambda_t)]$ is given by (4.14) and (4.15), due to the following arguments.

$$C_0 = \psi_1 \underbrace{\text{Cov}(l_t, I_t)}_{\text{(i)}} - \psi_1 \beta_m \underbrace{\text{Cov}(l_t, I_{t-m})}_{\text{(ii)}} + \psi_1 \Psi_m \underbrace{\text{Cov}(l_t, l_{t-m})}_{\text{(iii)}} \tag{C.6}$$

In the following, we present (i), (ii), and (iii):

(i) $\text{Cov}(l_t, I_t) = \nu E[\exp(2\lambda_t)]/(\nu + 1)$, where $E[\exp(2\lambda_t)]$ is given by (4.14) and (4.15).

(ii) For $\text{Cov}(l_t, I_{t-m})$, we use the following conditional expectation:

$$E(l_t I_{t-m}|\mathcal{F}_{t-m}) = E(l_t|\mathcal{F}_{t-m})I_{t-m} \tag{C.7}$$

We use the law of iterated expectations (White 2001):

$$E(l_t|\mathcal{F}_{t-m}) = E[\underbrace{E(l_t|\mathcal{F}_{t-1})}_{0}|\mathcal{F}_{t-m}] = 0 \tag{C.8}$$

which holds because $E(|l_t|) < \infty$, and l_t is a MDS. By using the law of iterated expectations, $E(l_t I_{t-m}) = 0$. Hence, $\text{Cov}(l_t, I_{t-m}) = 0$ for all m.

(iii) $\text{Cov}(l_t, l_{t-m})$ is zero because l_t is a MDS with finite variance.

Third, $C_i = \text{Cov}(\psi_1 l_{t+i}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = 0$, for $i = 1, \ldots, J$, due to the following arguments:

$$C_i = \psi_1 \underbrace{\text{Cov}(l_{t+i}, I_t)}_{\text{(i)}} - \psi_1 \beta_m \underbrace{\text{Cov}(l_{t+i}, I_{t-m})}_{\text{(ii)}} + \psi_1 \Psi_m \underbrace{\text{Cov}(l_{t+i}, l_{t-m})}_{\text{(iii)}} \tag{C.9}$$

In the following, we present (i), (ii), and (iii):

(i) We use the following conditional expectation:

$$E(l_{t+i} I_t|\mathcal{F}_t) = E(l_{t+i}|\mathcal{F}_t)I_t \tag{C.10}$$

We use the law of iterated expectations (White 2001):

$$E(l_{t+i}|\mathcal{F}_t) = E[\underbrace{E(l_{t+i}|\mathcal{F}_{t+i-1})}_{0}|\mathcal{F}_t] = 0 \tag{C.11}$$

which holds because $E(|l_t|) < \infty$, and l_t is a MDS. Hence, by using the law of iterated expectations, $\mathrm{Cov}(l_{t+i}, I_t) = 0$ for $i = 1, \ldots, J$.

(ii) We use the following conditional expectation:

$$E(l_{t+i}I_{t-m}|\mathcal{F}_{t-m}) = E(l_{t+i}|\mathcal{F}_{t-m})I_{t-m} \tag{C.12}$$

We use the law of iterated expectations (White 2001):

$$E(l_{t+i}|\mathcal{F}_{t-m}) = E[\underbrace{E(l_{t+i}|\mathcal{F}_{t+i-1})}_{0}|\mathcal{F}_{t-m}] = 0 \tag{C.13}$$

which holds because $E(|l_t|) < \infty$, and l_t is a MDS. Hence, by using the law of iterated expectations, $\mathrm{Cov}(l_{t+i}, I_{t-m}) = 0$ for $i = 1, \ldots, J$.

(iii) $\mathrm{Cov}(l_{t+i}, l_{t-m}) = 0$ for $i = 1, \ldots, J$, because l_t is a MDS with finite variance.

Fourth, $D_j = \mathrm{Cov}(\psi_1 l_{t-j}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m})$ for $j = 2, \ldots, T-m$ is computed as follows:

$$D_j = \psi_1 \underbrace{\mathrm{Cov}(l_{t-j}I_t)}_{\text{(i)}} - \psi_1\beta_m \underbrace{\mathrm{Cov}(l_{t-j}I_{t-m})}_{\text{(ii)}} + \psi_1\Psi_m \underbrace{\mathrm{Cov}(l_{t-j}l_{t-m})}_{\text{(iii)}} \tag{C.14}$$

In the following, we present (i), (ii), and (iii):

(i) We use the following conditional expectation:

$$E(l_{t-j}I_t|\mathcal{F}_{t-j}) = E(I_t|\mathcal{F}_{t-j})l_{t-j} = E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-j}]l_{t-j} \tag{C.15}$$

We use the law of iterated expectations (White 2001):

$$E\{E[\exp(\lambda_t)\epsilon_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-j}\} = E\{\exp(\lambda_t)\underbrace{E[\epsilon_t|\mathcal{F}_{t-1}]}_{0}|\mathcal{F}_{t-j}\} = 0 \tag{C.16}$$

which is due to $E|I_t| < \infty$ and $E(\epsilon_t) = 0$. As a consequence, $\mathrm{Cov}(l_{t-j}I_t) = 0$ for $j =$

$2, \ldots, T-m$.

(ii) If $j > m$, then we use the following conditional expectation:

$$E(l_{t-j}I_{t-m}|\mathcal{F}_{t-j}) = E(I_{t-m}|\mathcal{F}_{t-j})l_{t-j} = E[\exp(\lambda_{t-m})\epsilon_{t-m}|\mathcal{F}_{t-j}]l_{t-j} \tag{C.17}$$

$$= E\{E[\exp(\lambda_{t-m})\epsilon_{t-m}|\mathcal{F}_{t-m-1}]|\mathcal{F}_{t-j}\}l_{t-j} = E\{\exp(\lambda_{t-m})\underbrace{E[\epsilon_{t-m}|\mathcal{F}_{t-m-1}]}_{0}|\mathcal{F}_{t-j}\}l_{t-j} = 0$$

which is due to $E(|I_t|) < \infty$ (White 2001). In this case, $E(l_{t-j}I_{t-m}) = 0$.

If $j = m$, then (Harvey 2013):

$$E(l_{t-m}I_{t-m}) = \frac{\nu E[\exp(2\lambda_t)]}{\nu + 1} \tag{C.18}$$

If $j < m$, then we use the following conditional expectation:

$$E(l_{t-j}I_{t-m}|\mathcal{F}_{t-m}) = E(l_{t-j}|\mathcal{F}_{t-m})I_{t-m} = E[\underbrace{E(l_{t-j}|\mathcal{F}_{t-j-1})}_{0}|\mathcal{F}_{t-m}]I_{t-m} = 0 \tag{C.19}$$

which is due to $E(|l_t|) < \infty$ (White 2001), and l_t is a MDS. In this case, $E(l_{t-j}I_{t-m}) = 0$.

(iii) If $j > m$, then we use the following conditional expectation:

$$E(l_{t-j}l_{t-m}|\mathcal{F}_{t-j}) = E(l_{t-m}|\mathcal{F}_{t-j})l_{t-j} = E[\underbrace{E(l_{t-m}|\mathcal{F}_{t-m-1})}_{0}|\mathcal{F}_{t-j}]l_{t-j} = 0 \tag{C.20}$$

which is due to $E(|l_t|) < \infty$ (White 2001), and l_t is a MDS. For this case, $E(l_{t-j}l_{t-m}) = 0$.

If $j = m$, then (Harvey 2013):

$$E(l_{t-m}^2) = \frac{\nu^2 E[\exp(2\lambda_t)]}{(\nu + 3)(\nu + 1)} \tag{C.21}$$

If $j < m$, then we use the following conditional expectation:

$$E(l_{t-j}l_{t-m}|\mathcal{F}_{t-m}) = E(l_{t-j}|\mathcal{F}_{t-m})l_{t-m} = E[\underbrace{E(l_{t-j}|\mathcal{F}_{t-j-1})}_{0}|\mathcal{F}_{t-m}]l_{t-m} = 0 \qquad \text{(C.22)}$$

which is due to $E(|l_t|) < \infty$ (White 2001), and l_t is a MDS. For this case, $E(l_{t-j}l_{t-m}) = 0$.

In summary, matrix $C_{U,V}$ with dimensions $(T-p) \times (T-m)$ is:

$$C_{U,V} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & D_m & 0 & \cdots & 0 \\ C_0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & D_m \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & C_0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \qquad \text{(C.23)}$$

where $C_0 = \text{Cov}(\psi_1 l_t, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = \psi_1 \nu E[\exp(2\lambda_t)]/(\nu+1)$, and

$$D_m = \text{Cov}(\psi_1 l_{t-i}, I_t - \beta_m I_{t-m} + \Psi_m l_{t-m}) = -\psi_1 \beta_m \frac{\nu E[\exp(2\lambda)]}{\nu+1} + \psi_1 \Psi_m \frac{\nu^2 E[\exp(2\lambda)]}{(\nu+3)(\nu+1)} \qquad \text{(C.24)}$$

where $E[\exp(2\lambda_t)]$ is given by (4.14) and (4.15).

References

Blazsek, Szabolcs, Alvaro Escribano, and Adrian Licht (2020) Prediction accuracy of bivariate score-driven risk premium and volatility filters: an illustration for the Dow Jones. Working Paper 2020-10, University Carlos III of Madrid, Department of Economics.

Harvey, Andrew C. (2013) *Dynamic Models for Volatility and Heavy Tails.* Cambridge: Cambridge University Press.

White, Halbert (2001) *Asymptotic Theory for Econometricians.* San Diego: Academic Press.

Wooldridge, Jeffrey M. (1994) Estimation and inference for dependent processes. In Robert F. Engle and Daniel L. McFadden (eds.), *Handbook of Econometrics*, Volume 4, pp. 2639–2738. Amsterdam: North-Holland.