Elements of Forecasting and Signal Extraction

Marc Wildi
Institute of Data Analysis and Process Design

August 2, 2012

Abstract

Keywords. Real-time signal extraction (RTSE), phase, amplitude, multivariate filter, timeliness-accuracy dilemma.

Disclaimer The paper is evolving. The date under the title can be used as reference.

1 Introduction

The following document collects and summarizes recent efforts in forecasting and (real-time) signal extraction as undertaken on SEFBlog[1]. It is also a companion to posted R-code[2]. Section 2 identifies the frequency-domain as a natural approach to forecasting and/or filtering. Section 3 discusses optimization principles and reviews univariate optimization criteria (Direct Filter Approach: DFA). The traditional mean-square perspective is extended to *customization* by addressing a fundamental uncertainty principle (this material is formalized and extended further in McElroy and Wildi (2012)). A closed-form solution, I-DFA, is derived and analyzed. A multivariate extension I-MDFA is proposed in section 4. Section 5 introduces an alternative nomenclature inspired from linear regression. Section 6 proposes inferential elements by specifying the asymptotic distribution of estimates. Section 7 extends the proposed statistical apparatus to richly parametrized (possibly ill-conditioned) designs through suitable *regularization*. This proceeding allows to tackle high-dimensional multivariate problems and/or problems involving large lag-orders. Section 8 presents and discusses useful filter constraints and proposes an integration thereof by re-formulating the previous (unconstrained) framework. Finally, section 9 summarizes results.

[1] http://blog.zhaw.ch/idp/sefblog

[2] One can check the category 'tutorial' on the left-hand side of SEFBlog in order to obtain easy access to code as well as to exercises.

2 Frequency Domain and Filter Effect

Let X_t, $t = 1, ..., T$ be a finite sample of observations and define Y_{T-r}, $r = 0, ..., T-1$ as the output of a filter with real coefficients γ_{kr}:

$$Y_{T-r} = \sum_{k=-r}^{T-r-1} \gamma_{kr} X_{T-r-k}$$

For $r = 0$ a real-time or causal filter is obtained, namely a linear combination of present and past observations. For $r > 0$, Y_{T-r} relies on 'future' observations $X_{t-r+1}, ..., X_T$ (smoothing). In order to derive the important filter effect we assume a particular (complex) input series $X_t := \exp(i\omega t),\ t \in \mathbb{Z}$. The output signal is thus

$$\begin{aligned} Y_{T-r} &= \sum_{k=-r}^{T-r-1} \gamma_{kr} \exp(i\omega(T-r-k)) && (1) \\ &= \exp(i\omega(T-r)) \sum_{k=-r}^{T-r-1} \gamma_{kr} \exp(-i\omega k) && (2) \\ &= \exp(i\omega(T-r))\Gamma_r(\omega) && (3) \end{aligned}$$

The (generally complex) function

$$\Gamma_r(\omega) := \sum_{k=-r}^{T-r-1} \gamma_{kr} \exp(-i\omega k) \tag{4}$$

is called the transfer function of the filter. We can represent the complex number $\Gamma_r(\omega)$ in polar coordinates according to

$$\Gamma_r(\omega) = A_r(\omega) \exp(-i\Phi_r(\omega)) \tag{5}$$

where $A_r(\omega) = |\Gamma_r(\omega)|$ is called the amplitude of the filter and $\Phi_r(\omega)$ is its phase.
We deduce from 1 that $X_t, t \in \mathbb{Z}$ is a periodic eigensignal of the filter with eigenvalue $\Gamma_r(\omega)$. Linearity of the filter implies that real and imaginary parts of X_t are mapped into real and imaginary parts of Y_t and therefore

$$\begin{aligned} \cos(t\omega) &\rightarrow A_r(\omega)\left[\cos(t\omega)\cos(-\Phi_r(\omega)) - \sin(t\omega)\sin(-\Phi_r(\omega))\right] \\ &= A_r(\omega)\cos(t\omega - \Phi_r(\omega)) \\ &= A_r(\omega)\cos(\omega(t - \Phi_r(\omega)/\omega)) && (6) \end{aligned}$$

The amplitude function $A_r(\omega)$ can be interpreted as the weight (damping if $A_r(\omega) < 1$, amplification if $A_r(\omega) > 1$) attributed by the filter to a sinusoidal input signal with frequency ω. The function

$$\phi_r(\omega) := \Phi_r(\omega)/\omega \tag{7}$$

can be interpreted as the time shift function of the filter in ω[3]. As we shall see in section 3, real-time signal extraction, i.e. the case $r = 0$, aims at optimal simultaneous amplitude and time shift matchings or 'fits'. Amplitude and time-shift functions describe comprehensively the effect of the filter when applied to a simple trigonometric signal of frequency omega. In order to extend the scope of the analysis and to found the validity of our approach we can rely on a well-known result stating that any sequence of numbers X_t, $t = 1, .., T$, sampled on an equidistant time-grid, can be decomposed uniquely into a weighted sum of mutually orthogonal complex exponential terms

$$X_t = \frac{1}{\sqrt{2\pi T}} \sum_{k=-T/2}^{T/2} DFT(\omega_k) \exp(-it\omega_k)$$

where $DFT(\omega_k)$ is the discrete Fourier transform of X_t and $\omega_k = \frac{k2\pi}{T}$, $k = -T/2, ..., 0, ..., T/2$ is a discrete frequency-grid in the interval $[-\pi, \pi]$. Linearity of the filter can then be invoked to extend the description of the filter effect in terms of amplitude and time-shift functions to arbitrary sequences of numbers X_t, $t = 1, ..., T$.

Note that this decomposition is a finite sample version of the fundamental spectral representation theorem and that it is fully compatible with the latter, asymptotically. However, the validity of the discrete finite-sample decomposition extends to any sequence of numbers, including realizations of non-stationary processes.

3 DFA

We here propose optimization criteria which emphasize optimal properties of asymmetric filters. For this purpose, we assume that a particular signal or, equivalently, a symmetric filter has been defined by the user. The signal could be a trend, a cycle or a seasonally adjusted component and the definition could be either ad hoc or 'model-based'. In order to simplify notations we here emphasize the practically relevant real-time or concurrent filter which approximates the signal at the end $t = T$ of the sample. We propose criteria which emphasize the revision error as well as speed (timeliness) and reliability (noise suppression) issues. Criteria in the first group are able to replicate traditional model-based filters (X-12-ARIMA,TRAMO,Stamp) perfectly, see for example section 6 in McElroy and Wildi (2012). Criteria in the second group are able to account for more complex real-time inferences (for example the detection of turning-points) and for more sophisticated user priorities (for example different levels of risk aversion).

[3]The singularity in $\omega = 0$ is resolved by noting that $\Phi(0) = 0$ for filters satisfying $\Gamma_r(0) > 0$. As a result $\phi_r(0) := \dot{\Phi}_r(0)$. For $\Gamma_r(0) = 0$ the phase could be set to any arbitrary value, including zero, of course.

3.1 Mean-Square

Let the target signal be defined by the output of a symmetric (possibly bi-infinite) filter

$$Y_T = \sum_{j=-\infty}^{\infty} \gamma_j X_{T-j} \tag{8}$$

and let $\hat{Y}_t$ denote its real-time estimate

$$\hat{Y}_T = \sum_{j=0}^{T-1} b_j X_{T-j} \tag{9}$$

Furthermore, let $\Gamma(\cdot) = \sum_{j=-\infty}^{\infty} \gamma_j \exp(-ij\cdot)$ and $\hat{\Gamma}(\cdot) = \sum_{j=0}^{T-1} b_j \exp(-ij\cdot)$ denote the corresponding transfer functions. For stationary processes X_t, the mean-square filter error can be expressed as

$$\int_{-\pi}^{\pi} |\Gamma(\omega) - \hat{\Gamma}(\omega)|^2 dH(\omega) = E[(Y_t - \hat{Y}_t)^2] \tag{10}$$

where $H(\omega)$ is the unknown spectral distribution of X_t. Consider now the following finite sample approximation of the above integral

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} w_k |\Gamma(\omega_k) - \hat{\Gamma}(\omega_k)|^2 S(\omega_k) \tag{11}$$

where $\omega_k = k2\pi/T$, $[T/2]$ is the greatest integer[4] smaller or equal to $T/2$ and the weights w_k are defined by

$$w_k = \begin{cases} 1 & , \quad |k| \neq T/2 \\ 1/2 & , \quad \text{otherwise} \end{cases} \tag{12}$$

In this expression, $S(\omega_k)$ can be interpreted as an estimate of the unknown spectral density of the process. Consistency of this estimate is not necessary because we are not interested in estimating the (unknown) spectral density but the filter mean-square error instead. In this perspective, we may take benefit of the smoothing effect provided by the summation operator in 11. So for example Wildi (1998), Wildi (2005), Wildi (2008) and Wildi (2010) propose to plug the periodogram into the above expression:

$$S(\omega_k) := I_{TX}(\omega_k) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t \exp(-it\omega_k) \right|^2$$

Formal efficiency results applying to the resulting Direct Filter Approach (DFA) are presented in Wildi (2008) and (2009). Real-world true out-of-sample performances are extensively documented

[4] In order to simplify the exposition we now assume that T is even. In our applications, the sample length is generally a multiple of 4 (quarterly data) or 12 (monthly data) in order that the important seasonal frequencies can be matched by ω_k.

and discussed in Wildi (2008).

In other parts of on-going work we propose to extend the original DFA by considering alternative spectral estimates $S(\omega_k)$ derived from *models* of the Data Generating Process (DGP)[5]. The term 'model' makes reference to explicit representations of the DGP by X-12-ARIMA, TRAMO or STAMP, for example, as well as to 'ad hoc' implicit DGP-assumptions underlying classical filters, such as HP, CF or Henderson, for example. This way, a formal link between the original DFA and traditional model-based approaches is established which allows to transpose the powerful customization principle of the latter to the former. These topics have been recently developed in Mc Elroy and Wildi (2012).

The case of non-stationary integrated processes can be handled by noting that 11 addresses the filter error $Y_t - \hat{Y}_t$, not the data X_t. The former is generally stationary even if the latter isn't[6] and therefore all spectral decomposition results are still valid in a formal mathematical perspective. In the case of integrated processes, stationarity of the filter error is obtained by imposing cointegration between the signal Y_t and the real-time estimate $\hat{Y}_t$. Formally, this amounts to impose suitable real-time filter constraints. A comprehensive treatment of the topic is given in Wildi (2008) and (2010).

3.2 Customization

Wildi (1998), (2005), (2008) and (2010) propose a decomposition of the mean-square filter error into distinct components attributable to the amplitude and the phase functions of the real-time filter. We here briefly review this decomposition and derive customized criteria which emphasize explicitly speed and/or reliability aspects subject to particular user priorities, such as, for example, various degrees of risk aversion.

The following identity holds for general transfer functions Γ and $\hat{\Gamma}$:

$$\begin{aligned}
|\Gamma(\omega) - \hat{\Gamma}(\omega)|^2 &= A(\omega)^2 + \hat{A}(\omega)^2 - 2A(\omega)\hat{A}(\omega)\cos\left(\hat{\Phi}(\omega) - \Phi(\omega)\right) \\
&= (A(\omega) - \hat{A}(\omega))^2 \\
&\quad +2A(\omega)\hat{A}(\omega)\left[1 - \cos\left(\hat{\Phi}(\omega) - \Phi(\omega)\right)\right] \qquad (13)
\end{aligned}$$

If we assume that Γ is symmetric and positive, then $\Phi(\omega) \equiv 0$. Inserting 13 into 11 and using

[5]See: http://blog.zhaw.ch/idp/sefblog/index.php?/archives/165-Real-Time-Signal-Extraction-RTSE-an-Agnostic-Perspective-Plus-a-Frivolity-of-Mine.html

[6]Optimal real-time signalextraction aims precisely at a smallest possible mean-square filter error i.e. the filter error is neither trending nor unbounded in variance.

$1-\cos(\hat{\Phi}(\omega))=2\sin(\hat{\Phi}(\omega)/2)^2$ then leads to

$$\frac{2\pi}{T}\sum_{k=-T/2}^{T/2} w_k(A(\omega_k)-\hat{A}(\omega_k))^2 S(\omega_k) \tag{14}$$

$$+\frac{2\pi}{T}\sum_{k=-T/2}^{T/2} w_k 4A(\omega_k)\hat{A}(\omega_k)\sin(\hat{\Phi}(\omega)/2)^2 S(\omega_k) \tag{15}$$

The first summand 14 is the distinctive part of the total mean-square filter error which is attributable to the amplitude function of the real-time filter (the MS-amplitude error). The second summand 15 measures the distinctive contribution of the phase or time-shift to the total mean-square error (the MS-time-shift error). The term $A(\omega_k)\hat{A}(\omega_k)$ in 15 is a scaling factor which accounts for the fact that the phase function does not convey level information.

Now consider the following generalized version of the original mean-square criterion[7]:

$$\begin{aligned}
&\frac{2\pi}{T}\sum_{k=-T/2}^{T/2} w_k(A(\omega_k)-\hat{A}(\omega_k))^2 W(\omega_k,\eta)S(\omega_k)\\
&+(1+\lambda)\frac{2\pi}{T}\sum_{k=-T/2}^{T/2} w_k 4A(\omega_k)\hat{A}(\omega_k)\sin(\hat{\Phi}(\omega_k)/2)^2 W(\omega_k,\eta)S(\omega_k)\\
=\;&\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k|\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 W(\omega_k,\eta)S(\omega_k)\\
&+4\lambda\frac{2\pi}{T}\sum_{k=-T/2}^{T/2} w_k A(\omega_k)\hat{A}(\omega_k)\sin(\hat{\Phi}(\omega_k)/2)^2 W(\omega_k,\eta)S(\omega_k)\to\min
\end{aligned} \tag{16}$$

where $W(\cdot):=W(\omega_k,\eta,\text{cutoff})$ is a weighting function defined by

$$W(\omega_k,\eta,\text{cutoff})=\left\{\begin{array}{c} 1\ ,\ \text{if } |\omega_k|<\text{cutoff}\\ (1+|\omega_k|-\text{cutoff})^\eta\ ,\ \text{otherwise}\end{array}\right. \tag{17}$$

The parameter cutoff marks the transition between pass- and stop-bands; positive values of the parameter η emphasize high-frequency matching in the stop-band (for notational simplicity we now drop the cutoff-parameter in the function-call of $W(\cdot)$). Classical mean-square optimization is obtained for $\lambda=\eta=0$: the revision error is addressed. For $\lambda>0$ the user can emphasize the contribution of the MS-time-shift error. As a result, corresponding real-time filters (typically low-pass trend or cycle extraction) will convey less delayed signals: turning-points can be detected earlier. Note that the weighting $A(\omega_k)\hat{A}(\omega_k)$ in this expression implies that λ acts on the *pass-band* frequencies exclusively and that η does not alter the time-shift error. The latter parameter emphasizes the MS-amplitude error by magnifying 'noisy' high-frequency components in the *stop-band*. As

[7]For notational simplicity it is assumed that $\Gamma(\omega)>0$ for all ω such that $\Gamma(\omega)=A(\omega)$.

a result, 'noise' is suppressed more effectively and the reliability of real-time estimates will improve accordingly.

It is generally admitted that reliability and timeliness (speed of detection) of real-time estimates are to some extent mutually exclusive requirements. It is not our intention to contradict this fundamental uncertainty principle, of course, but it seems obvious that the user can attempt to improve performances in both dimensions simultaneously by increasing λ as well as η, see McElroy and Wildi (2012), section 6, for illustration.

3.3 I-DFA

The mean-square error criterion is a quadratic function of the filter parameters and therefore the solution can be obtained analytically. The expression 16, however, is more tricky when $\lambda > 0$ because it involves non-linear functions of the filter parameters. Therefore, we here propose a new criterion which opens the way to an analytical approximation of 16 (for notational ease the additional weight w_k 12 has been dropped from all subsequent expressions). Consider the following expression:

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}\left|\Gamma(\omega_k)-\left\{\Re\left(\hat{\Gamma}(\omega_k)\right)+i\sqrt{1+4\lambda\Gamma(\omega_k)f(\omega_k)}\Im\left(\hat{\Gamma}(\omega_k)\right)\right\}\right|^2 W(\omega_k,\eta)S(\omega_k)\to\min \tag{18}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote real and imaginary parts and $i^2=-1$ is the imaginary unit. We here assume throughout that $f(\omega_k)=\text{Id}$ is an identity and call the resulting optimization criterion I-DFA[8]. Obviously, the above expression is quadratic in the filter coefficients. In analogy to 16, the weighting function $W(\omega_k,\eta)$ emphasizes the fit in the stop band. The term $\lambda\Gamma(\omega_k)$ emphasizes the imaginary part of the real-time filter in the pass band: for $\lambda > 0$ the imaginary part is artificially inflated and therefore the phase is affected. The following development allows for a direct comparison of 16 and 18:

[8] Tweaking of $f(\omega_k)$ will be treated in a separate paper.

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}\left|\Gamma(\omega_k)-\left(\Re\left(\hat{\Gamma}(\omega_k)\right)+i\sqrt{1+4\lambda\Gamma(\omega_k)}\Im\left(\hat{\Gamma}(\omega_k)\right)\right)\right|^2 W(\omega_k,\eta)S(\omega_k) \quad (19)$$

$$= \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}\left\{\left(\Gamma(\omega_k)-\Re\left(\hat{\Gamma}(\omega_k)\right)\right)^2+\Im\left(\hat{\Gamma}(\omega_k)\right)^2\right\}W(\omega_k,\eta)S(\omega_k)$$

$$+4\lambda\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}\Gamma(\omega_k)\Im\left(\hat{\Gamma}(\omega_k)\right)^2 W(\omega_k,\eta)S(\omega_k)$$

$$= \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}|\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 W(\omega_k,\eta)S(\omega_k)$$

$$+4\lambda\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}A(\omega_k)\hat{A}(\omega_k)^2\sin(\hat{\Phi}(\omega_k))^2 W(\omega_k,\eta)S(\omega_k) \quad (20)$$

A direct comparison of 16 and 20 reveals that $\hat{\Phi}(\omega_k)/2$ is replaced by $\hat{\Phi}(\omega_k)$ and a supernumerary weighting-term $\hat{A}(\omega_k)$ appears in the latter expression. Expression 20 can be solved analytically for arbitrary λ and/or weighting functions $W(\omega_k,\eta)$ because $\hat{A}(\omega_k)^2\sin(\hat{\Phi}(\omega_k))^2$ is simply the squared imaginary part of the real-time filter. For $\lambda = 0$ the original (DFA) mean-square criterion 11 is obtained. Overemphasizing the imaginary part of the real-time filter in the pass-band by augmenting $\lambda > 0$ results in filters with smaller phase (time-shifts). It should be noted, however, that the analytic I-DFA criterion 18/20 is 'less effective' in controlling the time-shift than the original DFA-criterion 16 (this is where the function $f(\omega_k)$ in 18 comes into play...). Published I-DFA code on SEFBlog relies on 18. A derivation of the analytic solution is provided in the appendix.

3.4 Replicating and Customizing HP, CF or BK-Filters

In order to replicate real-time HP, BK or CF filters one needs to plug-in the (pseudo-)spectral density underlying the implicit model of each filter (random-walk for CF, ARIMA(0,2,2) for HP) and to set $\lambda = \eta = 0$ in 18. That's all. Customization - enhancing timeliness and/or reliability of HP, CF,BK - is then obtained very easily by selecting a suitable combination of (λ,η) in 18.

4 MDFA

We here review a multivariate extension of previous results.

4.1 Mean-Square

The above (univariate) DFA has been generalized to a general multivariate framework (MDFA) in Wildi (2008.2). Specifically, theorem 7.1 proposes optimization criteria by highlighting the cointegration rank, ranging from full-rank (stationary case) to zero-rank. We here briefly summarize the main results: for ease of exposition we restrict the discussion to the stationary case. Let Y_t be defined by 8 and assume the existence of m additional explaining variables $W_{tj},\ j = 1, ..., m$ enrichening the information universe. We here rewrite 11 by adopting the traditional DFA-framework based on the periodogram ($S(\omega_k) := I_{TX}(\omega_k)$) and the DFT $\Xi_{TX}(\omega_k)$:

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}|\Gamma(\omega_k)-\hat{\Gamma}(\omega_k)|^2 I_{TX}(\omega_k) = \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}|\Gamma(\omega_k)\Xi_{TX}(\omega_k)-\hat{\Gamma}(\omega_k)\Xi_{TX}(\omega_k)|^2 \tag{21}$$

Consider the following generalization of the univariate real-time filter expression:

$$\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)+\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k) \tag{22}$$

where

$$\hat{\Gamma}_X(\omega_k) = \left(\sum_{j=0}^{L} b_{Xj}\exp(-ij\omega_k)\right)\Xi_{TX}(\omega_k) \tag{23}$$

$$\hat{\Gamma}_{W_n}(\omega_k) = \left(\sum_{j=0}^{L} b_{w_nj}\exp(-ij\omega_k)\right)\Xi_{TW_n}(\omega_k) \tag{24}$$

are the (one-sided) transfer functions applying to the 'explaining' variables and $\Xi_{TX}(\omega_k)$, $\Xi_{TW_n}(\omega_k)$ are the corresponding DFT's. Theorem 7.1 in Wildi (2008.2) shows that the following straightforward extension of 21

$$\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\left(\Gamma(\omega_k)-\hat{\Gamma}_X(\omega_k)\right)\Xi_{TX}(\omega_k)-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right|^2 \to \min_{\mathbf{B}} \tag{25}$$

inherits all efficiency properties of the (univariate) DFA and therefore the whole customization principle can be carried over to a general multivariate framework ($\mathbf{B}$ denotes the matrix of unknown filter parameters).

4.2 Customization

A generalization of the customized criterion 16 to the multivariate case can be obtained by a simple transformation applying to 25:

$$\begin{aligned}
&\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\Gamma(\omega_k)\Xi_{TX}(\omega_k)-\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right|^2\\
=\ &\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\Gamma(\omega_k)-\hat{\Gamma}_X(\omega_k)-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\frac{\Xi_{TW_n}(\omega_k)}{\Xi_{TX}(\omega_k)}\right|^2|\Xi_{TX}(\omega_k)|^2\\
=\ &\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\Gamma(\omega_k)-\tilde{\Gamma}(\omega_k)\right|^2|\Xi_{TX}(\omega_k)|^2
\end{aligned} \tag{26}$$

where

$$\tilde{\Gamma}(\omega_k) := \hat{\Gamma}_X(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k) \frac{\Xi_{TW_n}(\omega_k)}{\Xi_{TX}(\omega_k)} \tag{27}$$

Expression 26 'looks like' 11 and therefore the same customization can be applied, in principle, as in the latter expression (introducing λ and η). Specifically, we here rely on the analytically tractable customized criterion 20

$$\begin{aligned}
&\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} |\Gamma(\omega_k) - \tilde{\Gamma}(\omega_k)|^2 W(\omega_k, \eta) \, |\Xi_{TX}(\omega_k)|^2 \\
&+4\lambda \frac{2\pi}{T} \sum_{k=-T/2}^{T/2} A(\omega_k) \tilde{A}(\omega_k)^2 \sin(\tilde{\Phi}(\omega_k))^2 W(\omega_k, \eta) \, |\Xi_{TX}(\omega_k)|^2 \to \min
\end{aligned} \tag{28}$$

Note that potential singularities introduced by small values of $\Xi_{TX}(\omega_k)$ in the denominator of 27 could be ignored because they are cancelled by the outer-product with $|\Xi_{TX}(\omega_k)|^2$. However, numerical routines don't like this kind of cancelling. A numerically robust alternative, implemented in the published R-code, is presented in the appendix. Note also that 26 in combination with 20, as shown in 28, is a 'clever' solution in the sense that one does not need to define series specific $\lambda's$ or $\eta's$. The idea goes as follows:

- We are not interested in controlling for time-shifts or smoothness of series specific filter outputs of the multivariate filter.
- Instead, we are interested in having a timely (fast) and accurate (smooth) **aggregate**: indeed, the output of the multivariate filter is obtained by aggregating cross-sectionally the individual series' outputs.
- The parameters λ and η acting on $\tilde{\Gamma}(\omega_k)$, as defined by 27 (and plugged into 28), determine properties of the aggregate output, as desired, and therefore the resulting criterion 28 matches our intention.

4.3 I-MDFA

A generalization of 18 is straightforward when relying on the notational trick introduced in 26. Specifically, I-MDFA is obtained by substituting $\tilde{\Gamma}(\omega_k)$ in 26 to $\hat{\Gamma}(\omega_k)$ in 18:

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \left| \Gamma(\omega_k) - \left\{ \Re\left(\tilde{\Gamma}(\omega_k)\right) + i\sqrt{1 + 4\lambda\Gamma(\omega_k) f(\omega_k)} \Im\left(\tilde{\Gamma}(\omega_k)\right) \right\} \right|^2 W(\omega_k, \eta) S(\omega_k) \to \min \tag{29}$$

5 Nomenclature

We here port the original frequency-domain notation of the previous sections into a more elegant and simpler nomenclature inspired from ordinary linear regression. Besides esthetical considerations our purpose here is solidly anchored into reality: early this year we experienced that a 'plain-vanilla' extension of the 'old' (pre-2012) I-MDFA to Regularization can be fastidious. Therefore, we had to introduce some fundamental 'order'. As is frequently the case, 'new order' allows for 'new insights' by putting structure on previously loose concepts. Uniformity and standardization is a beneficial outcome of the proposed nomenclature as (hopefully) illustrated by step-wise generalization in section 9. Another beneficial effect is that the close connection to familiar 'linear regression' expressions allows to introduce inferential elements which were kept hidden in technical appendices until yet (Wildi (2008) and Wildi (2008.2)), see section 6. Some care has to be taken, though, because this is still frequency-domain territory, see section 5.1.

5.1 Mean-Square Framework

In a linear-regression framework, a dependent variable z_k can be linked to a set of explaining variables x_{km}, $m = 1, ..., M$ through a set of linear equations (indexed by k)

$$z_k = c + \sum_{m=1}^{M} a_m x_{km} + \epsilon_k \tag{30}$$

Determination of unknown parameters a_m is obtained by minimizing the sum of squared errors

$$\sum_k \epsilon_k^2 = \sum_k \left(z_k - \left(c + \sum_{m=1}^{M} a_m x_{km} \right) \right)^2 \to \min_{a_k} \tag{31}$$

In a classical time series context the subscript k indexes time i.e. k becomes t. But I-MDFA is set-up in the frequency-domain. So t becomes... ω_k? And what about $t - j$? And what is the dependent variable in I-MDFA? And ϵ_k?

Briefly, the criterion 31 corresponds to 25: the dependent variable z_k in 30 corresponds to the target signal $\Gamma(\omega_k)\Xi_{TX}(\omega_k)$, the unknown coefficients a_m are the unknown b_{Xj} and b_{w_nj}, $j = 0, ..., l$, in 23 and 24 and the corresponding explaining variables are $\exp(-ij\omega_k)\Xi_{TX}(\omega_k)$ and $\exp(-ij\omega_k)\Xi_{TW_n}(\omega_k)$ respectively. The residual ϵ_k must then be the DFT of the filter error in frequency ω_k. The running indices k in 31 and 25 match.

Note that $\Gamma(\omega_k)\Xi_{TX}(\omega_k)$ is observable, in the frequency-domain, whereas the output Y_T of the bi-infinite filter in 8 isn't. Also, the complex exponentials $\exp(-ij\omega_k)$ appearing in 23 and 24 correspond to the backshift operator B^j, applied j-times to X_T (giving X_{T-j}) or to W_{nT} (thus giving $W_{n,T-j}$). We obtain a straightforward analogy between the frequency-domain criterion 25 and 'plain' time-domain linear regression targeting the (unobserved) signal Y_t (in $t = T$). To complete the analogy we can derive the $(T/2+1)*(L+1)(m+1)$-dimensional *design-matrix* $\mathbf{X}$ for the linear frequency-domain regression underlying 25: its k-th row $\mathbf{X}_k$ is obtained from appending the rows in the following matrix to a (possibly very long) row-vector:

$$\mathbf{X}_k = (1+I_{k>0})\mathrm{Vec}_{\mathrm{row}}\begin{pmatrix} \Xi_{TX}(\omega_k) & \exp(-i\omega_k)\Xi_{TX}(\omega_k) & ... & \exp(-iL\omega_k)\Xi_{TX}(\omega_k) \\ \Xi_{TW_1}(\omega_k) & \exp(-i\omega_k)\Xi_{TW_1}(\omega_k) & ... & \exp(-iL\omega_k)\Xi_{TW_1}(\omega_k) \\ \Xi_{TW_2}(\omega_k) & \exp(-i\omega_k)\Xi_{TW_2}(\omega_k) & ... & \exp(-iL\omega_k)\Xi_{TW_2}(\omega_k) \\ ... & ... & ... & ... \\ \Xi_{TW_m}(\omega_k) & \exp(-i\omega_k)\Xi_{TW_m}(\omega_k & ... & \exp(-iL\omega_k)\Xi_{TW_m}(\omega_k) \end{pmatrix} \quad (32)$$

where the $\mathrm{Vec}_{\mathrm{row}}$-operator appends rows (we use this notation in order to avoid margin-overflow) and where the indicator function $(1+I_{k>0}) = \begin{cases} 1 & k=0 \\ 2 & k=1,...,T/2 \end{cases}$ accounts for the fact that frequency zero ($k=0$) occurs only once in 25 whereas all frequencies $\omega_k, k>0$ are duplicated or mirrored by a corresponding $-\omega_k = \omega_{-k}$: by taking absolute values 'observations' for positive and negative k's coincide (are mirrored). Note that the dimension of the k-th row is indeed $(L+1)(m+1)$, as indicated above, and that the dimension of the design-matrix $\mathbf{X}$ is $(T/2+1)*((L+1)(m+1))$. The corresponding coefficient-vector $\mathbf{b}$ (obtained by stacking columns of the $\mathbf{B}$-matrix) and the dependent vector $\mathbf{Y}$ are

$$\mathbf{b} = \mathrm{Vec}_{\mathrm{col}}(\mathbf{B}) = \mathrm{Vec}_{\mathrm{col}}\begin{pmatrix} b_{X0} & b_{W_10} & b_{W_20} & ... & b_{W_m0} \\ b_{X1} & b_{W_11} & b_{W_21} & ... & b_{W_m1} \\ ... & ... & ... & ... & ... \\ b_{XL} & b_{W_1L} & b_{W_2L} & ... & b_{W_mL} \end{pmatrix} \quad , \quad \mathbf{Y} = \begin{pmatrix} \Gamma(\omega_0)\Xi_{TX}(\omega_0) \\ 2\Gamma(\omega_1)\Xi_{TX}(\omega_1) \\ 2\Gamma(\omega_2)\Xi_{TX}(\omega_2) \\ . \\ 2\Gamma(\omega_{T/2})\Xi_{TX}(\omega_{T/2}) \end{pmatrix}$$

where $\mathrm{Vec}_{\mathrm{col}}$ stacks columns.Note, once again, that all frequencies larger than zero are duplicated in $\mathbf{Y}$. Neglecting the constant $\frac{2\pi}{T}$ in 25 we can now express the criterion in the more familiar form

$$(\mathbf{Y}-\mathbf{X}\mathbf{b})'(\mathbf{Y}-\mathbf{X}\mathbf{b}) \to \min_{\mathbf{b}} \quad (33)$$

One would expect that the ordinary 'least-squares' solution

$$\hat{\mathbf{b}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$$

should solve the I-MDFA mean-square criterion 25 but this assumption is wrong: the design-matrix and the target vector are complex and therefore the proposed ordinary LS-estimate would be a vector of complex numbers, in turn.

In order to obtain the correct formula (as implemented in our R-code) we first rotate all DFT's in the complex plane in such a way that the criterion 25 is not affected (the mean-square norm is insensitive to rotations in the complex plane):

$$
\begin{aligned}
& \frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\left(\Gamma(\omega_k)-\hat{\Gamma}_X(\omega_k)\right)\Xi_{TX}(\omega_k)-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right|^2 \\
= & \frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\Big||\Gamma(\omega_k)\Xi_{TX}(\omega_k)|\exp\left(i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right)-\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) \\
& -\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\Big|^2 \\
= & \frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\Big||\Gamma(\omega_k)\Xi_{TX}(\omega_k)|-\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)\exp\left(-i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right) \\
& -\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right)\Big|^2
\end{aligned}
\tag{34}
$$

As explained in the appendix, this rotation is helpful when customizing I-MDFA[9]. We now refer to expression 34 when deriving a 'correct' expression for the regression-estimate $\hat{\mathbf{b}}$ of 33. Let us re-write the (rotated) design-matrix $\mathbf{X}_{\text{rot}}$ and target vector $\mathbf{Y}_{\text{rot}}$:

$$
\mathbf{X}_{k,\text{rot}} = \mathbf{X}_k\exp\left(-i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right) \tag{35}
$$

where $\mathbf{X}_{k,\text{rot}}$ designates the k-th row (of the new design matrix) and

$$
\mathbf{Y}_{\text{rot}} = |\mathbf{Y}|
$$

is a real positive vector. Criterion 25 then becomes (up to a negligible scalar factor)

$$
(\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b}) \to \min_{\mathbf{b}} \tag{36}
$$

The general matrix derivative formula for tackling the minimization 36 with respect to $\mathbf{b}$, accounting for the presence of complex numbers, is

$$
\begin{aligned}
d/d\mathbf{b}\text{ Criterion} &= d/d\mathbf{b}\ (\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b}) \\
&= -(\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b})'\mathbf{X}_{\text{rot}}-(\mathbf{Y}_{\text{rot}}-\mathbf{X}_{\text{rot}}\mathbf{b})^{\mathbf{T}}\overline{\mathbf{X}_{\text{rot}}} \\
&= -2\mathbf{Y}_{\text{rot}}'\Re\left(\mathbf{X}_{\text{rot}}\right)-\mathbf{2b}'(\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})
\end{aligned}
$$

[9] The numerical problem discussed in section 4.2 can be avoided. A straightforward extension of 34 to the more general customized I-MDFA criterion 28 is provided in 72 in the appendix. In deriving this expression we shall assume that $\Gamma(\cdot)$ is a traditional symmetric signal extraction filter (the transfer function is real and positive). This assumption is just for convenience and does not preclude generality in any way.

where 34 is replicated up to the irrelevant scaling $\frac{2\pi}{T}$; ${\mathbf{X}_{\text{rot}}}'$ is the transposed *and* conjugate (rotated) design-matrix (Hermitian conjugate); $\mathbf{X}_{\text{rot}}^{\mathbf{T}}$ is the transposed (but not complex conjugate) matrix; $\overline{\mathbf{X}_{\text{rot}}}$ is the complex conjugate (but not transposed) matrix; $\Re\left(\mathbf{X}_{\text{rot}}\right)$ is its real part; $\mathbf{b}' = \mathbf{b}^T$ and $\mathbf{Y}' = \mathbf{Y}^T$ because both vectors are real. Note also that $\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}} = \overline{\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}}} = \mathbf{X}_{\text{rot}}^{\mathbf{T}}\overline{\mathbf{X}_{\text{rot}}}$ because the matrix-product is real (positive). Equating the previous expression to zero provides the *correct* 'linear regression' estimate (as derived in the R-code):

$$\hat{\mathbf{b}} \;=\; (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Re(\mathbf{X}_{\text{rot}})'\mathbf{Y}_{\text{rot}} \tag{37}$$

Note that $\Re(X_{\text{rot}})' = \Re(X_{\text{rot}})^T$. Without insisting too much on formal details let us note that the 'residual' of the regression (the DFT of the filter error) is a sequence of independent random variables under fairly general assumptions (typically invoked in the case of model-based approaches). Therefore the proposed least-squares estimate 37 is consistent. It is also efficient in the sense that the filter output $\hat{Y}_t$ obtained by plugging 37 into the filter-equations is asymptotically closest possible to Y_t, see Wildi (2005) and (2008).

5.2 Smoothing

The above representations emphasize a real-time concurrent filter for estimating Y_T towards the sample end $t = T$. Smoothing, i.e. estimation of Y_{T-h} could be obtained very easily by multiplying (rotating) each row $\mathbf{X}_{k,\text{rot}}$ of $\mathbf{X}_{\text{rot}}$ with $\exp(ih\omega_k)$, as done in our R-code. This simple transformation signifies that the explaining data is shifted forward by h time units relative to the fixed target (which results in optimal smoothing).

5.3 Customization

Having achieved the transcription of the mean-square I-MDFA criterion into a linear regression framework (with complex data), we now proceed to the 'customized' I-MDFA criterion 28 which is rewritten in its explicit (slightly more complex) form:

$$\sum_k \Bigg| \Gamma(\omega)\left|\Xi_{TX}(\omega_k)\right| - \Re\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)$$
$$-i*\sqrt{1+\lambda\Gamma(\omega_k)}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)\right\}\Bigg|^2 W(\omega_k,\eta) \tag{38}$$

where $W(\omega_k, \eta)$ is the weighting-function 17. Note that we here assume that $\Gamma(\omega_k) = |\Gamma(\omega_k)|$ (symmetric target filter with positive transferfunction) mainly for notational convenience, to limit 'margin-overflow'. Then, $\Gamma(\omega)\left|\Xi_{TX}(\omega_k)\right|$ are real (positive) numbers and therefore the above expression is 'rotated', see also section B in the appendix for reference. We can now derive the corresponding (rotated) design-matrix and the target vector:

$$\mathbf{X}_{k,\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta) = \left\{\Re(\mathbf{X}_{k,\mathrm{rot}}) + i\sqrt{1+\lambda\Gamma(\omega_k)}\Im(\mathbf{X}_{k,\mathrm{rot}})\right\}\sqrt{W(\omega_k,\eta)}$$

$$\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}}(\eta) = \begin{pmatrix} |\Gamma(\omega_0)\Xi_{TX}(\omega_0)|\sqrt{W(\omega_0,\eta)} \\ 2|\Gamma(\omega_1)\Xi_{TX}(\omega_1)|\sqrt{W(\omega_1,\eta)} \\ 2|\Gamma(\omega_2)\Xi_{TX}(\omega_2)|\sqrt{W(\omega_2,\eta)} \\ . \\ 2|\Gamma(\omega_{T/2})\Xi_{TX}(\omega_{T/2})|\sqrt{W(\omega_{T/2},\eta)} \end{pmatrix}$$

where $\mathbf{X}_{k,\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta)$ and $\mathbf{X}_{k,\mathrm{rot}}$ designate the k-th rows of the corresponding design-matrices. It is easily seen that $\mathbf{X}_{k,\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta) = \mathbf{X}_{k,\mathrm{rot}}$ if $\lambda = 0$ and $W(\omega_k,\eta) = 1$ (by imposing $\eta = 0$). Criterion 29 or, alternatively, 38 then become (up to a negligible scalar factor)

$$(\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}}(\eta) - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta)\mathbf{b})'(\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}}(\eta) - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta)\mathbf{b}) \to \min_{\mathbf{b}} \tag{39}$$

Accordingly, the customized coefficient estimate is obtained as

$$\hat{\mathbf{b}}^{\mathrm{Cust}}(\lambda,\eta) = \left((\mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta))'\mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta)\right)^{-\mathbf{1}}\Re(\mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta))'\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}}(\lambda,\eta) \tag{40}$$

Note that the (minimal) criterion value obtained by plugging $\hat{\mathbf{b}}^{\mathrm{Cust}}(\lambda,\eta)$ into 39 (omitting optimization, of course) cannot be interpreted as an estimate of the filter-MSE. If the latter quantity is of interest, then $\hat{\mathbf{b}}^{\mathrm{Cust}}(\lambda,\eta)$ must be plugged into 36, instead (omitting optimization, again).

Since all expressions will gain in complexity when tackling the next topic we drop the parameters (λ,η) in all customized expressions from now on: the superscript 'Cust' refers implicitly to both adjustments/parameters.

5.4 Bending

'Customization' is a generalization of the ordinary 'Mean-Square' paradigm since the latter can be replicated by setting $\lambda = \eta = 0$ (we can refer the formal reader to McElroy and Wildi (2012)[10]). On the other hand, 40 really is 'just another' MS estimate. But how 'other' is it? The only difference between 36 and 39 is that the user can express particular research priorities, in the latter, through λ,η[11]: the resulting filter will be faster and/or smoother (see section 6 in McElroy and Wildi (2012) for an illustrative example of the 'and' part). In 39, the user puts/assigns 'structure' on/to the estimation problem. This is achieved by bending geometry of the original MS-norm into a new MS-norm with some terms of the original one being magnified/inflated (phase in passband, amplitude in stop band) and some other ones being attenuated/shrunken (phase in stopband, amplitude in

[10] Replication of TRAMO-SEATS is provided as an illustrative example in section 6.

[11] SEFBlog readers asked us about 'optimal' choices of λ,η. The response is: transcript your mood.

passband). Customization bends ordinary MS-geometry into a user-perspective.

6 Inferential Elements

Given the previous transcription (encoding) of the original I-MDFA into a more familiar 'linear regression' nomenclature we here derive a convenient expression for the asymptotic distribution of I-MDFA filter coefficients. Once the distribution is available, tests are straightforward.

$$\begin{aligned}
\hat{\mathbf{b}} &= (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Re(\mathbf{X}_{\text{rot}})'\mathbf{Y}_{\text{rot}} \\
&= (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Big\{\Re(\mathbf{X}_{\text{rot}})'\Re(\mathbf{Y}_{\text{rot}}) + \Im(\mathbf{X}_{\text{rot}})'\Im(\mathbf{Y}_{\text{rot}})\Big\} \\
&= (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Big\{\Re(\mathbf{X}_{\text{rot}})'\Re(\mathbf{X}_{\text{rot}}\mathbf{b}+\mathbf{e}) + \Im(\mathbf{X}_{\text{rot}})'\Im(\mathbf{X}_{\text{rot}}\mathbf{b}+\mathbf{e})\Big\} \\
&= (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Big\{(\mathbf{X}_{\text{rot}})'\mathbf{X}_{\text{rot}}\mathbf{b} + \Re(\mathbf{X}_{\text{rot}})'\Re(\mathbf{e}) + \Im(\mathbf{X}_{\text{rot}})'\Im(\mathbf{e})\Big\} \\
&= \mathbf{b} + (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{-\mathbf{1}}\Big\{\Re(\mathbf{X}_{\text{rot}})'\Re(\mathbf{e}) + \Im(\mathbf{X}_{\text{rot}})'\Im(\mathbf{e})\Big\}
\end{aligned}$$

where $\mathbf{e} = \mathbf{Y}_{\text{rot}} - \mathbf{X}_{\text{rot}}\mathbf{b}$ is the filter error (in the frequency domain) and $\mathbf{b}$ is the 'true' value of coefficients minimizing MSE. The second equation is 'trivial' in the sense that $\mathbf{Y}_{\text{rot}}$ is a real (positive) vector and therefore $\Im(\mathbf{Y}_{\text{rot}}) = \mathbf{0}$. Under fairly general assumptions (applying for example to traditional model-based approaches) the expectation of the last matrix expression vanishes asymptotically i.e. $E[\hat{\mathbf{b}}] = \mathbf{b}$, see for example Wildi (2005), theorem 5.10, Wildi (2008) theorem 10.20 and McElroy-Wildi (2012)[12]. Furthermore one can show that

$$(\hat{\mathbf{b}} - \mathbf{b}) \sim \mathbf{N}(\mathbf{0}, \mathbf{W}) \tag{41}$$

asymptotically, where

$$\mathbf{W} = \mathbf{U}^{-\mathbf{1}}\mathbf{V}\mathbf{U}^{-\mathbf{1}}$$

and $\mathbf{V}$ and $\mathbf{U}$ are expectations of squared-gradients and Hessian matrices computed in $\mathbf{b}$. These entities are intricate/complex expressions see Wildi (2008), theorem 10.20, and McElroy-Wildi (2012). Note that $\mathbf{V}, \mathbf{U}$ are unknown and must be estimated, see Wildi (2008) theorem 10.20 for derivation of consistent estimates.

[12] These results address univariate DFA: we therefore rely on 26 in order to embed I-MDFA in the corresponding (univariate) framework.

7 Regularization

Typically, the number of filter coefficients to be estimated by I-MDFA is large. Thus overfitting is an important issue. Regularization is an attempt to alleviate overfitting by controlling degrees of freedom. A straightforward proceeding would be to impose 'ordinary' constraints on the parameter space by applying a (more or less 'primitive') dimension-squasher: for example by imposing a small filter length. The trick behind regularization is to restrict degrees of freedom without harming performances i.e. without shrinking the parameter-space to 'fatal misspecification'. For this purpose we shall emphasize properties of filter-coefficients which are felt to be desirable; for simplicity of exposition we assume these properties to be 'ideal' (say, in a Platonic sense). Regularization then shrinks the parameter space to our ideal: shrinkage implies that a smaller fraction of the available degrees of freedom is 'lost' and thus out-of-sample and in-sample performances are more likely to be in accordance i.e. overfitting is alleviated.

Let's attempt to summarize schematically the logic: if our 'ideal' requirements are effectively 'universal', then they should be pertinent too and therefore we would expect in-sample performances to be good; we then conclude that out-of-sample performances should follow accordingly.

7.1 Regularization-Troika

We first specify universal characteristics of filter-coefficients which are felt to be desirable a priori:

- Smoothness: coefficients should not change 'too erratically' as a function of lag; coefficients should be smooth and decays should be progressive[13].
- Decay: we expect that filter coefficients should converge 'sufficiently fast' towards zero.
- Similarity: coefficients should be 'as similar as possible' across time series (apply similar filters to similar series).

These three requirements set-up a so-called 'Regularization-Troika'. The terms of this 'Troika' are potentially conflicting: a fast decay can affect smoothness, for example. Also, each requirement limits overfitting by conflicting with in-sample performances (if properly implemented). Before proceeding to a formal definition of the Troika we first need to reparametrize the set of filter coefficients b_l^u, $l = 0, ..., L$, $u = 0, ..., m$[14].

$$b_l^u = \begin{cases} b_l + \delta b_l^u, & u > 0 \\ b_l - \sum_{u=1}^{m} \delta b_l^u, & u = 0 \end{cases} \tag{42}$$

[13]Note that we are working with seasonally adjusted data: eliminating seasonal components would eventually conflict with our simple smoothness requirement (more sophisticated concepts would be available in this case).

[14]The coefficient b_l^u is assigned to lag l of series u whereby $u = 0$ indicates X_t and $u > 0$ stands for W_{tu}.

In the case of univariate designs we have $m = 0$ so that the above reparametrization is groundless. Note that 42 imposes $\delta b_l^0 = -\sum_{u=1}^m \delta b_l^u$ such that b_l can be interpreted in terms of a 'central' parameter and δb_l^u, $u > 0$ can be viewed as series specific 'effects'. For formal developments we reformulate this alternative parametrization of the coefficient vector in matrix notation:

$$
\begin{aligned}
\mathbf{b} &= \mathbf{A}\tilde{\mathbf{b}} \qquad (43)\\
\mathbf{A} &= \begin{pmatrix} \mathbf{Id} & -\mathbf{Id} & -\mathbf{Id} & ... & ... & -\mathbf{Id} \\ \mathbf{Id} & \mathbf{Id} & \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{Id} & \mathbf{0} & \mathbf{Id} & \mathbf{0} & ... & \mathbf{0} \\ ::: & & & & & \\ \mathbf{Id} & \mathbf{0} & \mathbf{0} & ... & ... & \mathbf{Id} \end{pmatrix}\\
\mathbf{b}' &= \left(b_0^0, b_1^0, ..., b_L^0 \mid b_0^1, b_1^1, ..., b_L^1 \mid ... \mid ... \mid ... \mid b_0^m, b_1^m, ..., b_L^m\right)'\\
\tilde{\mathbf{b}}' &= \left(b_0, b_1, ..., b_L \mid \delta b_0^1, \delta b_1^1, ..., \delta b_L^1 \mid \delta b_0^2, \delta b_1^2, ..., \delta b_L^2 \mid ... \mid ... \mid ... \mid \delta b_0^m, \delta b_1^m, ..., \delta b_L^m\right)'
\end{aligned}
$$

where $\mathbf{Id}$ is an $(L+1)*(L+1)$ identity. Consider now the following generalization of the customized I-MDFA criterion 39:

$$
\begin{aligned}
&(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})\\
+ \quad &\lambda_{\text{smooth}}\mathbf{b}'\mathbf{Q_{smooth}}\mathbf{b} + \lambda_{\text{cross}}\mathbf{b}'\mathbf{Q_{cross}}\mathbf{b} + \lambda_{\text{decay}}\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b} \to \min \qquad (44)
\end{aligned}
$$

where the matrices $\mathbf{Q_{smooth}}$, $\mathbf{Q_{cross}}$ and $\mathbf{Q_{decay}}$ define bilinear forms[15]. The Regularization Troika is determined by specifying these matrices.

$$
\begin{aligned}
\mathbf{Q_{smooth}} &= \begin{pmatrix} \mathbf{Q} & \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & ... \\ ::: & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & \mathbf{Q} \end{pmatrix}\\
\mathbf{Q} &= \begin{pmatrix}
1 & -2 & 1 & 0 & 0 & 0 & 0 & .... & & & & & 0\\
-2 & 5 & -4 & 1 & 0 & 0 & 0 & .... & & & & & 0\\
1 & -4 & 6 & -4 & 1 & 0 & 0 & .... & & & & & 0\\
0 & 1 & -4 & 6 & -4 & 1 & 0 & .... & & & & & 0\\
:: & & & & & & & & & & & & \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 1 & -4 & 6 & -4 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 1 & -4 & 6 & -4 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 1 & -4 & 5 & -2\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & 1 & -2 & 1
\end{pmatrix}
\end{aligned}
$$

[15] Our notation here is in accordance with the R-code: regularization matrices (bilinear forms) have identical names.

where $\mathbf{Q}$ is a $(L+1)*(L+1)$ matrix. It is not difficult (but cumbersome) to show that

$$\mathbf{b'Q_{smooth}b} = \sum_{u=0}^{m}\sum_{l=2}^{L}\left((1-B)^2 b_l^u\right)^2 \tag{45}$$

where $(1-B)^2 b_l = b_l - 2b_{l-1} + b_{l-2}$ denote second-order differences. Therefore $\mathbf{b'Q_{smooth}b}$ is a measure for the quadratic curvature - smoothness - of filter coefficients: if coefficients decay linearly, as a function of the lag, then this term vanishes. Increasing λ_{smooth} in 44 will assign preference to I-MDFA solutions whose filters are 'smooth': in the limit, when $\lambda_{\text{smooth}} \to \infty$, filter coefficients must be linear (as a function of lag). We now consider the cross-sectional term of the Troika. For this purpose we define

$$\begin{aligned} \mathbf{Q_{cross}} &= \mathbf{A^{-1\prime}\tilde{Q}_{cross}A^{-1}} \\ \mathbf{\tilde{Q}_{cross}} &= \begin{pmatrix} \mathbf{0}_{11} & | & \mathbf{0}_{12} \\ \mathbf{0}_{21} & | & \mathbf{Id}_{22} \end{pmatrix} \end{aligned} \tag{46}$$

where $\mathbf{0}_{11}$, $\mathbf{0}_{12}$ and $\mathbf{0}_{21}$ are $(L+1)*(L+1)$, $(L+1)*((L+1)m)$ and $(m(L+1))*(L+1)$ zero-matrices and $\mathbf{Id}_{22}$ is a $(L+1)m*(L+1)m$ identity. Note that $\mathbf{A^{-1}}$ transforms $\mathbf{b}$ to $\tilde{\mathbf{b}}$ and that the bilinear form (in transformed space) defined by $\mathbf{\tilde{b}'\tilde{Q}_{cross}\tilde{b}}$ sums up the squared deviations ('effects') δb_l^u i.e.

$$\begin{aligned} \mathbf{b'Q_{cross}b} &= \mathbf{\tilde{b}'\tilde{Q}_{cross}\tilde{b}} \\ &= \sum_{u=1}^{m}\sum_{l=0}^{L}(\delta b_l^u)^2 \end{aligned}$$

If this expression vanishes then filter coefficients are identical across time series since 'effects' vanish. Increasing λ_{cross} in 44 will assign preference to I-MDFA solutions whose filters are 'similar' across time series: in the limit, when $\lambda_{\text{cross}} \to \infty$, filter coefficients must be identical. To conclude we specify the decay-requirement in our Troika.

$$\begin{aligned} \mathbf{Q_{decay}} &= \begin{pmatrix} \text{diag}(\mathbf{q}) & 0 & 0 & 0 & \dots & 0 \\ 0 & \text{diag}(\mathbf{q}) & 0 & 0 & \dots & 0 \\ 0 & 0 & \text{diag}(\mathbf{q}) & 0 & \dots & 0 \\ :: & & & & & \\ 0 & 0 & 0 & 0 & \dots & \text{diag}(\mathbf{q}) \end{pmatrix} \\ \mathbf{q} &= (q^{\max(0,h)}, q^{|1-\max(0,h)|}, q^{|2-\max(0,h)|}, \dots, q^{|L-\max(0,h)|}) \end{aligned}$$

where $q > 1$ and $\text{diag}(\mathbf{q})$ is a diagonal matrix with $\mathbf{q}$ as its diagonal. If $h = 0$ then the *concurrent* filter is emphasized (estimation of Y_T[16]). In this case the elements of $\mathbf{q}$ are monotonically increasing

[16]For $h < 0$ the I-MDFA filter forecasts Y_{T-h}: this is not a (classical) h-step ahead forecast of the *data* but of the *signal*, instead.

and

$$\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b} = \sum_{u=0}^{m}\sum_{l=0}^{L} q^l (b_l^u)^2 \tag{47}$$

The weight attributed by $\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b}$ to filter-coefficients inflates at an exponential rate as a function of the lag l: Increasing λ_{decay} in 44 will assign preference to I-MDFA solutions whose coefficients decay towards zero and q determines the rate (of decay). In practice we have found it convenient to concatenate the dual effect into a single parameter by setting $q := 1 + \lambda_{\text{decay}}$, as currently implemented in our R-code.
When estimating Y_{T-h}, $h > 0$ (smoothing) we would like to assign most (filter-) weight to observations coinciding with Y_{T-h}. This requirement would conflict with 47 since the regularization would enforce the corresponding coefficients to be close to zero. Therefore, $\mathbf{q}$ relies on a balanced design: minimum regularization is imposed to lag h (namely $q^{|h-\max(0,h)|} = q$) and then 'decay' is emphasized symmetrically on both sides and away from the target lag h (i.e. the vector $\mathbf{q}$ is no more monotonic).

7.2 A Potential Flaw of the Cross-Sectional Troika-Term and a (straightforward) Solution

The grand-mean parametrization 42 is very convenient for deriving a simple (diagonal) bilinear form $\tilde{\mathbf{Q}}_{\mathbf{cross}}$ in 46. However, a potential flaw - kind of inconsistency - of this parametrization is that it assigns regularization constraints in an asymmetric way. In order to understand the problem recall 42:

$$b_l^u \quad = \quad \begin{cases} b_l + \delta b_l^u, & u > 0 \\ b_l - \sum_{u=1}^m \delta b_l^u, & u = 0 \end{cases}$$

In this expression all coefficients for $u > 0$ receive a single deviance term δb_l^u whereas $b_l^0 = b_l - \sum_{u=1}^m \delta b_l^u$. Therefore, b_l^0 has more 'degrees of freedom' than $b_l^u, u > 0$. Stated differently, the variance of $\sum_{u=1}^m \delta b_l^u$ is larger than the variance of any single δb_l^u (assuming independence of the deviances). As a result of this asymmetry, results (estimates) are sensitive to the ordering of the columns in the data-matrix. There are other disadvantages that I do not discuss here but obviously the aforementioned flaw is sufficiently strong to think about alternatives.

A straightforward solution consists in ignoring the proposed 'grand-mean' parametrization and to impose regularization constraints directly on $\mathbf{b}$ instead of $\tilde{\mathbf{b}}$. Of course, we'll loose the nice diagonal form of the bilinear form but, as it appears, the resulting expression is quite simple too. So let's assume we want to impose (regularization) constraints on $\mathbf{b}$ according to

$$\sum_{u=0}^{m}\left(\left(b_0^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_0^{u'}\right)^2 + \left(b_1^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_1^{u'}\right)^2 + ... + \left(b_L^u - \frac{1}{m+1}\sum_{u'=0}^{m} b_L^{u'}\right)^2\right) \tag{48}$$

It is quite easy to derive a suitable symmetric bilinear form for this expression according to

$$\mathbf{Q_{cross}} = \begin{pmatrix} \mathbf{q_{cross,1}}' \\ \mathbf{q_{cross,2}}' \\ ... \\ \mathbf{q_{cross,(m+1)*(L+1)}}' \end{pmatrix} \tag{49}$$

where

$$\begin{aligned}
\mathbf{q_{cross,1}}' &= (1-\frac{1}{m+1},0,...,0 \mid -\frac{1}{m+1},0,...,0 \mid -\frac{1}{m+1},0,...,0 \mid ...) \\
\mathbf{q_{cross,2}}' &= (0,1-\frac{1}{m+1},0,...,0 \mid 0,-\frac{1}{m+1},0,...,0 \mid 0,-\frac{1}{m+1},0,...,0 \mid ...) \\
\mathbf{q_{cross,3}}' &= (0,0,1-\frac{1}{m+1},0,...,0 \mid 0,0,-\frac{1}{m+1},0,...,0 \mid 0,0,-\frac{1}{m+1},0,...,0 \mid ...) \\
&... \\
\mathbf{q_{cross,(m+1)*(L+1)}}' &= (0,0,...,-\frac{1}{m+1} \mid 0,0,...,-\frac{1}{m+1} \mid 0,0,...,-\frac{1}{m+1} \mid ... \mid 0,0,...,1-\frac{1}{m+1})
\end{aligned}$$

and the blocks (separated by $|$) are of length $L+1$. We thus have 1's on the diagonal of $\mathbf{Q_{cross}}$ and suitably periodically arranged $-\frac{1}{m+1}$ which account for the central means in 48.

The main advantage of this 'direct' cross-sectional regularization is its symmetry: all series are treated equally and therefore swapping columns in the data matrix does not affect estimates anymore. This parametrization is introduced in our R-code since 30.07.2012.

7.3 Solution

Criterion 44 is a *regularized* and *customized* version of the original mean-square criterion 36: setting $\lambda = \eta = \lambda_{\text{smooth}} = \lambda_{\text{cross}} = \lambda_{\text{decay}} = 0$ just replicates plain MS-performances. It is also a *regularized* version of the customized criterion 39: setting $\lambda_{\text{smooth}} = \lambda_{\text{cross}} = \lambda_{\text{decay}} = 0$ just replicates customized designs. Since the criterion is more general, we are interested in deriving the solution explicitly. Consider therefore

$$\begin{aligned}
& (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b}) \\
+ \quad & \lambda_{\text{smooth}}\mathbf{b}'\mathbf{Q_{smooth}}\mathbf{b} + \lambda_{\text{cross}}\mathbf{b}'\mathbf{Q_{cross}}\mathbf{b} + \lambda_{\text{decay}}\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b} \to \min
\end{aligned} \tag{50}$$

Since the Regularization Troika is defined by bilinear terms the criterion is still quadratic in unknown coefficients and therefore a closed-form solution exists. Derivation of the criterion with respect to parameters leads to

$$\begin{aligned}
d/d\mathbf{b}\text{ Criterion} &= d/d\mathbf{b}\ (\mathbf{Y}_{\text{rot}}^{\text{Cust}}-\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}}-\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})\\
&+\ d/d\mathbf{b}\ \left(\lambda_{\text{smooth}}\mathbf{b'Q_{smooth}b}+\lambda_{\text{cross}}\mathbf{b'Q_{cross}b}+\lambda_{\text{decay}}\mathbf{b'Q_{decay}b}\right)\\
&= -2(\mathbf{Y}_{\text{rot}}^{\text{Cust}})'\Re\left(\mathbf{X}_{\text{rot}}^{\text{Cust}}\right)+\mathbf{2b'}((\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}})\\
&\quad +2\lambda_{\text{smooth}}\mathbf{b'Q_{smooth}}+2\lambda_{\text{cross}}\mathbf{b'Q_{cross}}+2\lambda_{\text{decay}}\mathbf{b'Q_{decay}}
\end{aligned}$$

where we used the fact that bilinear forms were symmetric. Equating this expression to zero, the generalized solution is obtained as

$$\begin{aligned}
&\hat{\mathbf{b}}^{\text{Cust-Reg}}(\lambda,\eta,\lambda_{\text{smooth}},\lambda_{\text{cross}},\lambda_{\text{decay}})= \\
&\left((\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}+\lambda_{\text{smooth}}\mathbf{Q_{smooth}}+\lambda_{\text{cross}}\mathbf{Q_{cross}}+\lambda_{\text{decay}}\mathbf{Q_{decay}}\right)^{\mathbf{-1}}\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{Y}_{\text{rot}}^{\text{Cust}}
\end{aligned}\tag{51}$$

As expected, each term of the Regularization Troika contributes in ‘regularizing’ the matrix subject to inversion. Ill-posed problems with L large (large filter order) and m large (high-dimensional design) can be solved effectively by imposing ‘desirable’ shrinkage towards idealized filter characteristics. Users can familiarize with the new features by relying on a tutorial on SEFBlog[17] which analyzes partial and combined effects when applied to the EURI.

8 Filter Constraints

8.1 Specification

Regularization favours ‘universal’ characteristics of filters - of coefficients - which are felt to be desirable ‘a priori’ irrespective of the particular application. Terms of the Troika fight against each other as well as against in-sample performances in criterion 44. The resulting solution 51 is always some kind of compromise between conflicting requirements which reflects the Helvetic origin of its Designer. In contrast, filter constraints are hard-coded constraints which have to be satisfied in ‘absolute’ terms but their justification cannot claim ‘universality’: sometimes they are useful and sometimes not. We here distinguish two types of constraints namley level and time-shifts requirements: both are potentially relevant depending on *the purpose* of the application. The level constraint imposes pre-specified values for the amplitude functions in frequency zero. Typically, for *bandpass* filters one would impose $\hat{A}(0) = 0$, for example. For univariate *lowpass* filters this would become $\hat{A}(0) = 1$, typically. For multivariate low-pass designs, however, the topic is more complex/sophisticated because different time series do not always support common trends (the concept of cointegration can be linked to this topic). The time-shift constraint imposes a vanishing

[17] http://blog.zhaw.ch/idp/sefblog/index.php?/archives/246-Live-Tutorial-Round-2-What-is-Regularization-A-Follow-UP.html.

time-shift of filters in frequency zero: whereas this requirement is pretty trivial in the case of symmetric filters, it may become tricky in the case of asymmetric (for example concurrent) filters. A vanishing time-shift is highly desirable because turning-points in the filtered series are concomitant with turning-points in the original data.
From a model-based perspective, the level constraint is related to a single unit-root of the data generating process (DGP) in frequency zero whereas the time-shift constraint can be invoked when the DGP shows evidence of a double unit-root[18]. Our perspective, here, is very different though since we treat level and time-shift constraints as independent requirements: we can impose none, either one, or both, irrespective of 'unit-roots'. We really feel that the merit of these constraints is closely linked to research priorities, assigned by the user, rather than to abstract properties of the data[19].

Prosaically, the first order (level-) restriction is

$$b_0^u + b_1^u + ... + b_L^u = w^u \tag{52}$$

Specifically: the restriction imposes a *level constraint* according to $\hat{\Gamma}_{W_u}(0) = w^u$ (in notational terms we assume, again, that $\hat{\Gamma}_{W_0}$ is $\hat{\Gamma}_X$ if $u = 0$). Note that by constraining the transfer function we allow for control of the sign, too, which is slightly more stringent than the previously mentioned amplitude restriction. Thats fine! The second order restriction imposes a vanishing time-shift in frequency zero. For this purpose the derivative of the transfer function in frequency zero must vanish: $\frac{\partial}{\partial\omega}\big|_{\omega=0} \sum_{j=0}^{L} b_j^u \exp(-ij\omega) = 0$. This condition results in the following coefficient constraint:

$$b_1^u + 2b_2^u + 3b_3^u + ... + Lb_L^u = 0 \tag{53}$$

Note that this last expression assumes a real-time (concurrent) filter $\hat{\Gamma}(\omega) = b_0 + b_1 \exp(-i\omega) + ... + b_L \exp(-iL\omega)$ whose derivative with respect to ω in $\omega = 0$ reduces to the left hand-side of 53 (up to irrelevant scaling by the constant $-i$). The general case, including smoothing, is addressed in the next section.

Imposing both constraints simultaneously leads to:

$$\begin{aligned} b_{L-1}^u &= -Lb_0^u - (L-1)b_1^u - ... - 2b_{L-2}^u + Lw^u & (54) \\ b_L^u &= (L-1)b_0^u + (L-2)b_1^u + (L-4)b_3^u + ... + b_{L-2}^u - (L-1)w^u & (55) \end{aligned}$$

[18] We have never seen/observed/heard of an economic time series whose DGP is likely to be close to I(2). Never! Yet, non-stationary (economic) time series are frequently identified as I(2) realizations: TRAMO/SEATS, X-12-ARIMA or STAMP frequently select such (more or less severely) misspecified models.

[19] Wildi (2008) proposes a generalized test for verifying pertinence of the constraints, see chapter 6 in http://blog.zhaw.ch/idp/sefblog/index.php?/archives/168-RTSE-My-Good-Old-Book.html.

where $u = 0, ..., m$. These restrictions can be imposed independently by specifying i1$< -$T (level restriction) and/or i2$< -$T (time-shift) in our R-code.

Although the particular parametrization of the filter constraints proposed in 54 and 55 is formally correct, it strictly applies to the concurrent filter and does not account for smoothing (estimation of Y_{T-h}, $h > 0$). Moreover, as we shall see next, it potentially conflicts with the Regularization-Troika. A more general and consistent implementation of filter constraints is addressed in the next section.

8.2 Implementation

Smoothing concerns estimation of a signal Y_{T-h} shifted in time relative to $t = T$ i.e. $h > 0$. In this context, the specification of the time-shift constraint is affected: zero-shift is assigned to time point $T - h$. Formally, this requirement is obtained by equating the derivative of the smoothing filter to zero in frequency zero: $\frac{\partial}{\partial\omega}\Big|_{\omega=0} \sum_{j=-h}^{L-h} b_j^u \exp(-ij\omega) = 0$:

$$(-h)b^u_{-h} + (1-h)b^u_{1-h} + (2-h)b^u_{2-h} + ... + (-1)b^u_{-1} + b^u_1 + 2b^u_2 + ... + (L-h)b^u_{L-h} = 0 \qquad (56)$$

The level constraint becomes:

$$b^u_{-h} + b^u_{-(h-1)} + ... + b^u_{L-h} = w^u \qquad (57)$$

Evidently, given these restrictions (equations), any of the coefficients $b^u_{-h}, ..., b^u_{L-h}$ could be selected for implementing the constraint(s): equations 54 and 55 considered b^u_{L-1} and b^u_L as possible candidates. We now select coefficient(s) such that our choice does not conflict with the Regularization Troika. Indeed, the decay-term of the Troika assigns a preference to filters whose coefficients decay towards zero sufficiently rapidly on both sides of b_0. It would be unwise, in this context, to give-up grip - control - on b^u_{L-h} or b^u_{-h} (assuming $h >> 0$) by imposing constraints (on these specific parameters) which would potentially conflict with the natural decay-argument. Stated otherwise: we cannot expect that the filter restrictions imposed to b^u_{L-h} or b^u_h, for example, will ensure that these coefficients will be 'small', as intended by regularization (assuming $h >> 0$). This concept distinguishes b_0 and/or b_1 as *natural candidate(s)* for imposing constraints since then filter-restrictions would not conflict with Regularization requirements.

Since our filter constraints are linear we can express them in the form

$$\mathbf{b} \;\; = \;\; \mathbf{R}\mathbf{b_f} + \mathbf{c} \qquad (58)$$

where $\mathbf{b_f}$ is the vector of freely determined coefficients. In the following we specify the terms on the right of this equation in the three relevant cases i1=T, i2=F (simple level-constraint), i1=F, i2=T (simple time-shift constraint) and i1=i2=T (both constraints imposed). The fourth case i1=i2=F is trivial. In the first case we obtain

$$b_0^u = w^u - \sum_{k=-h,k\neq 0}^{L-h} b_k^u$$

and the entries in 58 become

$$\begin{aligned}
\mathbf{R} &= \begin{pmatrix} \mathbf{C} & 0 & ... & 0 \\ 0 & \mathbf{C} & ... & 0 \\ ::: & & & \\ 0 & 0 & ... & \mathbf{C} \end{pmatrix} && (59)\\
\mathbf{C} &= \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ -1 & -1 & -1 & ... & -1 & -1 \\ ::: & & & & & \\ 0 & 0 & 0 & ... & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} && (60)\\
\mathbf{c}' &= (0, ..., 0, w^0, 0, ..., 0 \mid\mid 0, ..., 0, w^1, 0, ..., 0 \mid\mid ... \mid\mid 0, ..., 0, w^m, 0, ..., 0)\\
\mathbf{b_f}' &= (b_{-h}^0, ..., b_{-1}^0, b_1^0, ..., b_{L-h}^0 \mid\mid b_{-h}^1, ..., b_{-1}^1, b_1^1, ..., b_{L-h}^1 \mid\mid ... \mid\mid b_{-h}^m, ..., b_{-1}^m, b_1^m, ..., b_{L-h}^m)
\end{aligned}$$

The row of -1's in $\mathbf{C}$ is in position $h+1$ whereas the constant w^u $(u = 0, ..., m)$ in $\mathbf{c}$ is in position $u * L + h + 1$. The vector $\mathbf{b_f'}$ collects all freely determined parameters (thus b_0^u is missing) whereas $\mathbf{b}$ collects all coefficients: the former vector is used for optimization and the latter is required for filtering (once $\mathbf{b_f'}$ has been determined).

The second case that we consider is i1=F and i2=T: a simple time-shift constraint without level requirement[20]. Then, from 56 we obtain

$$b_1^u = -(-h)b_{-h}^u - (1-h)b_{1-h}^u - ... - (-1)b_{-1}^u - 0b_0^u - 2b_2^u - ... - (L-h)b_{L-h}^u$$

[20]This case cannot be replicated by model-based approaches because of the hierarchical structure of unit-roots i.e. the I(2)-constraint conditiones the I(1)-requirement.

Such that

$$
\begin{array}{rcl}
\mathbf{C} & = & \left(\begin{array}{ccccccccc}
1 & 0 & 0 & ... & ... & ... & ... & 0 & 0\\
0 & 1 & 0 & ... & ... & ... & ... & 0 & 0\\
::: & & & & & & & & \\
h & h-1 & h-2 & ... & 1 & 0 & -2 & ... & -(L-h)\\
::: & & & & & & & & \\
0 & 0 & 0 & ... & ... & ... & ... & 1 & 0\\
0 & 0 & 0 & ... & ... & ... & ... & 0 & 1
\end{array}\right) \\
\mathbf{c} & = & \mathbf{0} \\
\mathbf{b_f}' & = & (b^0_{-h},...,b^0_0,b^0_2,...,b^0_{L-h}\,||\,b^1_{-h},...,b^1_0,b^1_2,...,b^1_{L-h}\,||\,...\,||\,b^m_{-h},...,b^m_0,b^m_2,...,b^m_{L-h})
\end{array}
\tag{61}
$$

in 58. Note that b_0^u does not enter explicitly into this constrainst because the observations X_{T-h} and $W_{T-h,u}$ are coincident with Y_{T-h}; therefore, altering their weights b_0^u does not affect the time-shift of the estimate $\hat{Y}_{T-h}$ (relative to the target Y_{T-h}). The non-trivial weighting vector $(h, h-1, ..., -(L-h))$ in the matrix $\mathbf{C}$ is now located in position $h+2$ (instead of $h+1$ in the previous case). The vector of constants $\mathbf{c}$ vanishes here. Finally, $\mathbf{b_f}$ collects all free coefficients i.e. all coefficients except b_1^u (instead of b_0^u in the previous case). The case $h=0$ (concurrent filter) is handled by

$$
\mathbf{C}_{h=0} = \left(\begin{array}{ccccc}
1 & 0 & 0 & ... & 0\\
0 & -2 & -3 & ... & -L\\
0 & 1 & 0 & ... & 0\\
::: & & & & \\
0 & 0 & 0 & ... & 1
\end{array}\right)
$$

The last case i1=i2=T assumes that both constraints are imposed. Solving for b_0^u and b_1^u in 56 and 57 leads to

$$
\begin{array}{rcl}
b_1^u & = & hb^u_{-h} + (h-1)b^u_{-(h-1)} + ... + b^u_{-1} - 2b^u_2 - 3b_3 - ... - (L-h)b^u_{L-h}\\
b_0^u & = & w^u - (h+1)b^u_{-h} - hb^u_{-(h-1)} - ... - 2b^u_{-1} + b^u_2 + 2b^u_3 + ... + ((L-1)-h)b^u_{L-h}
\end{array}
$$

In our general notation 58 we obtain

$$
\begin{aligned}
\mathbf{C} &= \begin{pmatrix}
1 & 0 & 0 & ... & & & & 0 & 0 \\
0 & 1 & 0 & ... & & & & 0 & 0 \\
::: & & & & & & & & \\
-(h+1) & -h & -(h-1) & ... & -2 & 1 & 2 & ... & (L-1-h) \\
h & h-1 & h-2 & ... & 1 & -2 & -3 & ... & -(L-h) \\
::: & & & & & & & & \\
0 & 0 & 0 & ... & & & & 1 & 0 \\
0 & 0 & 0 & ... & & & & 0 & 1
\end{pmatrix} \\
\mathbf{c}' &= (0,...,0,w^0,0,...,0 \;||\; 0,...,0,w^1,0,...,0 \;||\; ... \;||0,...,0,w^m,0,...,0) \\
\mathbf{b_f}' &= (b^0_{-h},...,b^0_{-1},b^0_2,...,b^0_{L-h} \;||\; b^1_{-h},...,b^1_{-1},b^1_2,...,b^1_{L-h} \;||\; ... \;||\; b^m_{-h},...,b^m_{-1},b^m_2,...,b^m_{L-h})
\end{aligned}
\tag{62}
$$

The non-trivial weighting rows in the matrix $\mathbf{C}$ are located in positions $h+1$ and $h+2$ whereas the constants w^u $(u = 0,...,m)$ in $\mathbf{c}$ are to be found in position $u*(L-1)+h+1$. Note that both b_0^u and b_1^u are now missing in the vector of freely determined parameters $\mathbf{b_f}$. The cases $h = 0, 1$ are accounted for by

$$
\begin{aligned}
\mathbf{C}_{h=0} &= \begin{pmatrix}
1 & 2 & 3 & ... & (L-1) \\
-2 & -3 & -4 & ... & -L \\
1 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & ... & 1
\end{pmatrix} \\
\mathbf{C}_{h=1} &= \begin{pmatrix}
1 & 0 & 0 & 0 & ... & 0 \\
-2 & 1 & 2 & 3 & ... & (L-2) \\
1 & -2 & -3 & -4 & ... & -(L-1) \\
0 & 1 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & 1
\end{pmatrix}
\end{aligned}
$$

Obviously, all entities on the right-hand side of 58 depend on i1, i2 as well as on h (even the dimensions involved depend on i1 and i2). However, for notational simplicity we refrained from attaching a cumbersome triple index to them. We assume from now on that the reader is aware of this coquetry when interpreting expressions involving any of the incriminated terms.

8.3 Optimization

In order to derive the constrained customized and regularized parameter vector we can plug 58 into 44 and take derivatives

$$
\begin{aligned}
d/d\mathbf{b_f}\ \text{Criterion} &= d/d\mathbf{b_f}\ (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{R}\mathbf{b_f}+\mathbf{c}))'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{R}\mathbf{b_f}+\mathbf{c}))\\
&+ d/d\mathbf{b_f}\ \lambda_{\text{smooth}}(\mathbf{R}\mathbf{b_f}+\mathbf{c})'\,\mathbf{Q_{smooth}}\,(\mathbf{R}\mathbf{b_f}+\mathbf{c})\\
&+ d/d\mathbf{b_f}\ \lambda_{\text{cross}}(\mathbf{R}\mathbf{b_f}+\mathbf{c})'\,\mathbf{Q_{cross}}\,(\mathbf{R}\mathbf{b_f}+\mathbf{c})\\
&+ d/d\mathbf{b_f}\ \lambda_{\text{decay}}(\mathbf{R}\mathbf{b_f}+\mathbf{c})'\,\mathbf{Q_{decay}}\,(\mathbf{R}\mathbf{b_f}+\mathbf{c})\\
&= -2(\mathbf{Y}_{\text{rot}}^{\text{Cust}})'\Re\left(\mathbf{X}_{\text{rot}}^{\text{Cust}}\right)\mathbf{R} - \mathbf{2}(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{c})'(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R}) + \mathbf{2b_f'}((\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})\\
&\quad +2\lambda_{\text{smooth}}\mathbf{b_f'}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + 2\lambda_{\text{smooth}}(\mathbf{c}'\mathbf{Q}_{\text{smooth}}\mathbf{R})\\
&\quad +2\lambda_{\text{cross}}\mathbf{b_f'}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + 2\lambda_{\text{cross}}(\mathbf{c}'\mathbf{Q}_{\text{cross}}\mathbf{R})\\
&\quad +2\lambda_{\text{decay}}\mathbf{b_f'}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} + 2\lambda_{\text{decay}}(\mathbf{c}'\mathbf{Q}_{\text{decay}}\mathbf{R})
\end{aligned}
$$

where, again, we used the fact that bilinear forms are symmetric. Equating this expression to zero, the *customized, regularized and constrained* solution is obtained as

$$
\begin{aligned}
&\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda,\eta,\lambda_{\text{smooth}},\lambda_{\text{cross}},\lambda_{\text{decay}},i1,i2)\\
=&\ \left\{(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R}\right\}^{-1}\\
&\ \left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} - (\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{c} - \mathbf{R}'(\lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{cross}})\mathbf{c}\right)\\
=&\ \left\{\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R}\right\}^{-1}\\
&\ \left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const}\right) \qquad (63)
\end{aligned}
$$

where

$$\mathbf{Const} = -\mathbf{R}'\left\{\mathbf{X}_{\text{rot}}^{\text{Cust}\prime}\mathbf{X}_{\text{rot}}^{\text{Cust}} + \lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{decay}}\right\}\mathbf{c}$$

This last term collects all level constraints. A comparison with 51 illustrates that both expressions - with or without constraints - are formally quite similar, up to the additional transformation of previous matrices by $\mathbf{R}$ and the emergence of a new level-shift **Const**[21]. The result of the optimization is $\mathbf{b_f}$, the vector of freely determined coefficients. The sought-after 'full-coefficient' vector $\mathbf{b}$, which is indispensable for the filtering-task, is then obtained from 58.

8.4 Forecasting

The case $h < 0$ signifies that the future signal Y_{T-h} is targeted: unlike traditional model-based approaches our design tackles this estimation problem 'directly', without relying on data-forecasts;

[21] For those readers who followed progress of this work on SEFBlog: the i1-bug in our R-code was due to the fact that we ignored **Const**. The reason is that we worked with bandpass filters for which all $w^u = 0$ and thus $\mathbf{c} = \mathbf{0}$ thus the new level-shift vanished i.e. the error could not be detected.

it is thus less sensitive to misspecification issues since we are not interested in identifying the DGP. The only difference in this case occurs when i1=i2=T. Specifically the vector of constants $\mathbf{c}$ then becomes

$$\mathbf{c} = (-(h-1)w^0, hw^0, ..., 0 \;||\; -(h-1)w^1, hw^1, ..., 0 \;||\; ... \;||\; -(h-1)w^m, hw^m, ..., 0)$$

which generalizes the case $h = 0$ (nowcast). Note that we observed a similar weighting of the constants w^u in 54 and 55, already.

8.5 Alternative Parametrization

It is convenient to work with $\tilde{\mathbf{b}}$ as defined in 43 when implementing cross-sectional regularization. Imposing constraints on $\tilde{\mathbf{b}}$ rather than $\mathbf{b}$ could be done very easily by swapping between both spaces, using $\mathbf{A^{-1}}$ i.e.

$$\begin{aligned}
\mathbf{b} &= \mathbf{R}\mathbf{b_f} + \mathbf{c} \\
\tilde{\mathbf{b}} &= \mathbf{A^{-1}}\left(\mathbf{R}\mathbf{b_f} + \mathbf{c}\right)
\end{aligned}$$

We make use of this facility in our R-code.

9 Summary

We here list all optimization criteria in a hierarchical way.

9.1 Mean-Square

$$(\mathbf{Y}_{\text{rot}} - \mathbf{X}_{\text{rot}}\mathbf{b})'(\mathbf{Y}_{\text{rot}} - \mathbf{X}_{\text{rot}}\mathbf{b}) \to \min_{\mathbf{b}}$$

$$\hat{\mathbf{b}} = (\mathbf{X}_{\text{rot}}'\mathbf{X}_{\text{rot}})^{\mathbf{-1}}\Re(\mathbf{X}_{\text{rot}})'\mathbf{Y}_{\text{rot}}$$

9.2 Customization

Let

$$\begin{aligned}
\mathbf{X}_{k,\text{rot}}^{\text{Cust}}(\lambda,\eta) &= \left\{\Re(\mathbf{X}_{k,\text{rot}}) + i\sqrt{1+\lambda\Gamma(\omega_k)}\Im(\mathbf{X}_{k,\text{rot}})\right\}\sqrt{W(\omega_k,\eta)} \\
\mathbf{Y}_{\text{rot}}^{\text{Cust}}(\eta) &= \begin{pmatrix} |\Gamma(\omega_0)\Xi_{TX}(\omega_0)|\sqrt{W(\omega_0,\eta)} \\ 2|\Gamma(\omega_1)\Xi_{TX}(\omega_1)|\sqrt{W(\omega_1,\eta)} \\ 2|\Gamma(\omega_2)\Xi_{TX}(\omega_2)|\sqrt{W(\omega_2,\eta)} \\ . \\ 2|\Gamma(\omega_{T/2})\Xi_{TX}(\omega_{T/2})|\sqrt{W(\omega_{T/2},\eta)} \end{pmatrix}
\end{aligned}$$

Then

$$(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b}) \to \min_{\mathbf{b}}$$

Accordingly, the customized coefficient estimate is obtained as

$$\hat{\mathbf{b}}^{\text{Cust}}(\lambda, \eta) \quad = \quad \left((\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\right)^{\mathbf{-1}} \Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{Y}_{\text{rot}}^{\text{Cust}}$$

9.3 Customization and Regularization

Please recall that all expressions marked with the superscript 'Cust' depend on (λ, η), which have been dropped to avoid notational overflow.

$$\begin{aligned} & (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b}) \\ + \quad & \lambda_{\text{smooth}}\mathbf{b}'\mathbf{Q}_{\mathbf{smooth}}\mathbf{b} + \lambda_{\text{cross}}\mathbf{b}'\mathbf{Q}_{\mathbf{cross}}\mathbf{b} + \lambda_{\text{decay}}\mathbf{b}'\mathbf{Q}_{\mathbf{decay}}\mathbf{b} \to \min \end{aligned}$$

$$\begin{aligned} & \hat{\mathbf{b}}^{\text{Cust-Reg}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}) = \\ & \left((\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}} + \lambda_{\text{smooth}}\mathbf{Q}_{\mathbf{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\mathbf{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\mathbf{decay}}\right)^{\mathbf{-1}} \Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} \end{aligned}$$

9.4 Integrating Level and Time-Shift Constraints

We have to distinguish four cases:

- i1=i2=F (no constraints imposed): $\mathbf{R} = \mathbf{Id}, \mathbf{c} = \mathbf{0}$.
- i1=T, i2=F (level constraint without time-shift): $\mathbf{C}, \mathbf{c}$ are based on 60 and $\mathbf{R}$ is specified by 59.
- i1=F, i2=T (time-shift without level-constraint): then $\mathbf{C}, \mathbf{c}$ are based on 61 and $\mathbf{R}$ is specified by 59.
- i1= i2=T (both constraints imposed): then $\mathbf{C}, \mathbf{c}$ are based on 62 and $\mathbf{R}$ is specified by 59.

The estimate in each of these cases is:

$$\begin{aligned} & \hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2) \\ = \quad & \left\{\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q}_{\mathbf{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q}_{\mathbf{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q}_{\mathbf{decay}}\mathbf{R}\right\}^{-1} \\ & \left(\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const}\right) \end{aligned}$$

where

$$\mathbf{Const} = -\mathbf{R}'\left\{\mathbf{X}_{\text{rot}}^{\text{Cust}\prime}\mathbf{X}_{\text{rot}}^{\text{Cust}} + \lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{decay}}\right\}\mathbf{c}$$

The sought-after parameter vector is then obtained as

$$\hat{\mathbf{b}} \quad = \quad \mathbf{R}\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2) + \mathbf{c}$$

Let us briefly remind of the fact that all expressions on the right-hand side of this last expression (including $\mathbf{R}$ and $\mathbf{c}$) depend on i1, i2 as well as on the target-lead/-lag h, see section 8.2.

9.5 Stepwise Generalization

Any criterion in this list is a generalization of previous ones. Setting $\mathbf{R} = \mathbf{Id}$ and $\mathbf{c} = \mathbf{0}$ means that no restrictions are imposed and therefore the last criterion reduces to the previous 'customization and regularization' case. Setting $\lambda_{\text{smooth}} = \lambda_{\text{cross}} = \lambda_{\text{decay}} = 0$ removes the new regularization feature, thus replicating I-MDFA performances as available before Jan-2012. Setting $\lambda = \eta = 0$ then replicates MS-performances. Of course, MS-performances could be regularized i.e. any dosage of 'Customization' can be combined with any dosage of 'Regularization'. In applications so far we have found the decay-term of the Reg-Troika to be most relevant. Adding some cross-sectional control can be meaningful too. We have still to find an application where the smoothness requirement is useful beyond 'esthetical' requirements, yet. At least, knowing coefficients to be smooth may assist in improving the quality of sleep which is the ultimate purpose of the Regularization Troika.

10 To Do's and Links to R-Code

To do's:

- Develop the level-constraint for non-stationary time series in a multivariate setting (the theory is proposed in Wildi (2008.2)).
- Develop the link between I-(M)DFA and (M)DFA. A 'confidential' paper exists. I'll disclose some of the results in a later revision of the current paper.
- A lot of little dirty correction work...

Links to posted R-code (incomplete and in loose order):

- Indexing of vectors in R-code starts with 1: thus b_0 (in this paper) corresponds to b[1] (in our code). This will shift all indices in our code (when compared to the paper) and L in our code means $L - 1$ in this paper.
- The time index h signifying estimation of Y_{T-h} is called Lag in our code. Lag=0 means concurrent filter (estimation of Y_T at the sample end $t = T$), $Lag > 0$ means smoothing and $Lag < 0$ forecasting (forecast of the signal Y_{T+Lag}).

- All regularization features (in particular the matrices $\mathbf{Q}$ and $\mathbf{R}$ as well as the vector $\mathbf{c}$ and the matrix $\mathbf{Const}$) are defined and computed in a new function called 'mat-func' in the file I-MDFA-new.r
  - The combined effect of the matrices $\mathbf{A}$ and $\mathbf{R}$ is called des-mat in our code. For i1=i2=F (no restriction imposed) $\mathbf{R}$ is an identity and des-mat is simply $\mathbf{A}$.
  - The term $\mathbf{Const}$ collecting level-constraints is split into xtxy and reg-xtxy in our code (reg-xtxy is the contribution by the Regularization Troika whereas xtxy is the contribution of the ordinary (unregularized) part).
- Complex conjugation and transposition $\mathbf{X}'\mathbf{X}$ should give a real number. In order to enforce this feature we work explicitly with real and imaginary parts in our code i.e. one doesn't find $\mathbf{X}'\mathbf{X}$ but $\Re(\mathbf{X}')\Re(\mathbf{X}) + \Im(\mathbf{X}')\Im(\mathbf{X})$ instead.
- Numerical considerations: if L is large then the decay-part ($\mathbf{Q}_{\text{decay}}$) will inflate at an exponential rate and the problem might become intractable in numerical terms, particularly when m (the number of explaining series) is large too. One should take care to impose 'meaningful' regularizations. We frequently use magnitudes of $\lambda_{\text{decay}} = 0.06$. Cross-sectional and smoothness terms might be added, if necessary/useful.
- The variable 'rever' in our code is the value of the optimization criterion 50 possibly with constraints imposed through 58. This is not an estimate of the mean-square error of the filter (unless $\lambda = \eta = \lambda_{\text{smooth}} = \lambda_{\text{cross}} = \lambda_{\text{decay}} = 0$) because it accounts for customization as well as for regularization. We mention this fact because some users tend to misinterpret this variable. We'll probably up-date our code by providing both numbers: the value of the optimization criterion as well as the estimate of the filter-MSE.

A I-DFA: Working out Trigonometric Terms

The following calculations apply to the univariate filter criterion 18 (for simplicity of exposition the constant multiplication term 4 has been concatenated into λ):

$$\begin{aligned}
& \sum_k |\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k)) - i * \sqrt{1+\lambda\Gamma(\omega_k)} Im(\hat{\Gamma}(\omega_k))|^2 I_{TX}(\omega_k) \\
= & \sum_k \left(\left[\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k))\right]^2 + (1+\lambda\Gamma(\omega_k)) Im\left(\hat{\Gamma}(\omega_k)\right)^2 \right) I_{TX}(\omega_k)
\end{aligned}$$

We now differentiate this expression with respect to filter parameters:

$$\sum_k \Big((\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k)))(-d/db_j(Re(\hat{\Gamma}(\omega_k)))) + (1+\lambda\Gamma(\omega_k)) Im(\hat{\Gamma}(\omega_k)) d/db_j(Im(\hat{\Gamma}(\omega_k))) \Big) I_{TX}(\omega_k) = 0$$

Now $-d/db_j(Re(\hat{\Gamma}(\omega_k))) = -\cos(j\omega_k)$ and $d/db_j(Im(\hat{\Gamma}(\omega_k))) = \sin(j\omega_k)$[22]. Therefore we obtain

$$\sum_k \Big((\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k)))(-\cos(j\omega_k)) + (1+\lambda\Gamma(\omega_k)) Im(\hat{\Gamma}(\omega_k))\sin(j\omega_k)) \Big) I_{TX}(\omega_k) = 0$$

Or

$$\sum_k (\Gamma(\omega_k)\cos(j\omega_k)) I_{TX}(\omega_k) = \sum_k \Big(Re(\hat{\Gamma}(\omega_k))\cos(j\omega_k) + (1+\lambda\Gamma(\omega_k)) Im(\hat{\Gamma}(\omega_k))\sin(j\omega_k)) \Big) I_{TX}(\omega_k) \tag{64}$$

Now $Re(\hat{\Gamma}(\omega_k)) = \sum_l b_l \cos(l\omega_k)$ and $Im(\hat{\Gamma}(\omega_k)) = \sum_l b_l \sin(l\omega_k)$. Therefore we obtain the following set of equations on the right-hand side of 64:

$$\begin{aligned}
& b_0 \sum_k (\cos(j\omega_k)\cos(0\omega_k) + (1+\lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(0\omega_k)) I_{TX}(\omega_k) + \\
& b_1 \sum_k (\cos(j\omega_k)\cos(1\omega_k) + (1+\lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(1\omega_k)) I_{TX}(\omega_k) + \\
& ... + \\
& b_L \sum_k (\cos(j\omega_k)\cos(L\omega_k) + (1+\lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(L\omega_k)) I_{TX}(\omega_k)
\end{aligned}$$

Let $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Note that all expressions are real here, in contrast to 37, 40, 51 and 63 which require usage of the $\Re$-operator. Then the right-hand side of 64 implies

$$\mathbf{X}'\mathbf{X} = \left(\sum_k (\cos(j\omega_k)\cos(m\omega_k) + (1+\lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(m\omega_k)) I_{TX}(\omega_k) \right)_{jm} \tag{65}$$

where both indices $0 \leq j, m \leq L$. Finally, the left-hand side of 64 implies that

$$\mathbf{X}'\mathbf{Y} = \left(\sum_k (\Gamma(\omega_k)\cos(j\omega_k)) I_{TX}(\omega_k) \right)_j \tag{66}$$

where $j = 0, ..., L$. These formulas are used in the published R-code.

B I-MDFA: Working out Trigonometric Terms

The criterion is:

$$\sum_k \left| \Gamma(\omega_k) - \tilde{\Gamma}(\omega_k) \right|^2 |\Xi_{TX}(\omega_k)|^2 \to \min_{\mathbf{B}} \tag{67}$$

[22]Please note that we inverted the sign of the complex exponential functions i.e. one should read $\cos(-j\omega_k)$ instead of $\cos(j\omega_k)$ (which does not change anything...) and $\sin(-j\omega_k)$ instead of $\sin(j\omega_k)$. Obviously, this arbitrary change of sign is completely irrelevant to the derivation of parameters. The only 'relevant' modification concerns the phase of the real-time filter whose sign must be inverted in order to allow for a meaningful interpretation of filter diagnostics.

where

$$\tilde{\Gamma}(\omega_k) := \hat{\Gamma}_X(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k) \frac{\Xi_{TW_n}(\omega_k)}{\Xi_{TX}(\omega_k)} \tag{68}$$

Instead of the numerically potentially unstable ratio of DFT's appearing in 68, we here re-write the criterion such that the potential singularity is avoided (the R-code relies on this re-formulated expression). Let us first consider the following multivariate generalization of 18 (for notational simplicity the constant multiplier 4 is concatenated into λ and the weighting function $W(\omega_k, \eta)$ has been ignored):

$$\sum_k \left| \Gamma(\omega)\Xi_{TX}(\omega_k) - \Re\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right.$$
$$\left. -i * \sqrt{1+\lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right\}\right|^2 \to \min$$

where i is the imaginary unit. Although this is not the right way to proceed we here first follow this line of attack (the necessary modification is provided below). One obtains:

$$\sum_k \left| \Gamma(\omega)\Xi_{TX}(\omega_k) - \Re\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right. \tag{69}$$

$$\left. -i * \sqrt{1+\lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right\}\right|^2 \tag{70}$$

$$= \sum_k \left(\Re()^2 + \Im()^2\right)$$

$$= \sum_k \left(\Gamma(\omega)\Re\left(\Xi_{TX}(\omega_k)\right) - \Re\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)\right) - \sum_n \Re\left(\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right)^2$$

$$+ \sum_k \left(\Gamma(\omega)\Im\left(\Xi_{TX}(\omega_k)\right) - \sqrt{1+\lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)\right) + \sum_{n=1}^{m} \Im\left(\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)\right\}\right)^2 \tag{71}$$

We can recognize/identify two problems related to this expression:

- A nuisance term $\Gamma(\omega)\Im\left(\Xi_{TX}(\omega_k)\right)$ appears in the imaginary part 71.
- Requiring a smaller imaginary part (reduced phase) of the aggregate filter

  $$\Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right)$$

  by augmenting λ in 70 would not necessarily lead to the expected improvement because the target signal $\Gamma(\omega)\Xi_{TX}(\omega_k)$ in 69 is a complex number with a non-vanishing imaginary part too.

Both problems could be avoided in 67, by isolating $\Xi_{TX}(\omega_k)$ outside of the filter expression. In doing this, we note that we don't need to 'isolate' the whole DFT: its argument would be sufficient. So let's have a look at the following modified expression

$$\begin{aligned}
&\sum_k \Bigg|\Gamma(\omega)\left|\Xi_{TX}(\omega_k)\right| - \Re\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)\\
&-i*\sqrt{1+\lambda\Gamma(\omega_k)}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)\right\}\Bigg|^2\\
&\left|\exp\left(i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right|^2
\end{aligned}$$

where we isolate $\exp\left(i\arg\left(\Xi_{TX}(\omega_k)\right)\right)$ 'only'. Since $\left|\exp\left(i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right|^2 = 1$ we can simplify the above expression to obtain:

$$\begin{aligned}
&\sum_k \Bigg|\Gamma(\omega)\left|\Xi_{TX}(\omega_k)\right| - \Re\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)\\
&-i*\sqrt{1+\lambda\Gamma(\omega_k)}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right)\right\}\Bigg|^2
\end{aligned}\tag{72}$$

As can be seen, all previous problems are solved: the nuisance term has vanished in the imaginary part; imposing a smaller imaginary part of the aggregate multivariate filter (by augmenting λ) would meet our target signal $\Gamma(\omega)\left|\Xi_{TX}(\omega_k)\right|$ which is now *real*; finally, this expression and the resulting criterion are stable numerically. In order to simplify notations let us denote the rotated DFT's in 72 by:

$$\begin{aligned}
\tilde{\Xi}_{TX}(\omega_k) &= \left|\Xi_{TX}(\omega_k)\right|\\
\tilde{\Xi}_{TW_n}(\omega_k) &= \Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right)
\end{aligned}$$

We can now proceed to the formal solution by differentiating expression 72 with respect to b_j^m (the $j-$th MA-coefficient of the filter applied to W_{mt}) and equating to zero:

$$\begin{aligned}
&\sum_k\left(\Gamma(\omega)\tilde{\Xi}_{TX}(\omega_k) - \Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) - \sum_{n=1}^{m}\Re\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right)(-1)\Re\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)\\
&-\sum_k\left(\sqrt{1+\lambda\Gamma(\omega_k)}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) + \sum_{n=1}^{m}\Im\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right\}\right)(-1)\Im\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)\\
=\ &0
\end{aligned}$$

where we assumed $\tilde{\Xi}_{TW_m}(\omega_k) = \tilde{\Xi}_{TX}(\omega_k)$ if $m = 0$. One then obtains

$$\begin{aligned}
&\sum_k \Gamma(\omega)\tilde{\Xi}_{TX}(\omega_k)\Re\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)\\
=\ &\sum_k\left\{\Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) + \sum_{n=1}^{m}\Re\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right\}\Re\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)\\
&+\sum_k\sqrt{(1+\lambda\Gamma(\omega_k))}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) + \sum_{n=1}^{m}\Im\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right\}\Im\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)
\end{aligned}\tag{73}$$

We show that this expression reduces to the classical univariate mean-square DFA-criterion when $m=0$ and $\lambda=0$. The left-hand side becomes (recall that $\tilde{\tilde{\Xi}}_{TX}(\omega_k)$ is real after rotation)

$$\Gamma(\omega)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\Re\left(\exp(ij\omega_k)\tilde{\tilde{\Xi}}_{TW_m}(\omega_k)\right) = \Gamma(\omega)\cos(j\omega_k)I_{TX}(\omega_k)$$

which corresponds to 66. The right-hand side simplifies to

$$\begin{aligned}
& \Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)\Re\left(\exp(ij\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)+\Im\left(\hat{\Gamma}_X(\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)\Im\left(\exp(ij\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)\\
= {} & \Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)\Re\left(\overline{\exp(ij\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)}\right)-\Im\left(\hat{\Gamma}_X(\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\right)\Im\left(\overline{\exp(ij\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)}\right)\\
= {} & \Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\tilde{\Xi}}_{TX}(\omega_k)\overline{\tilde{\tilde{\Xi}}_{TX}(\omega_k)\exp(ij\omega_k)}\right)\\
= {} & \Re\left(\hat{\Gamma}_X(\omega_k)\overline{\exp(ij\omega_k)}\right)I_{TX}(\omega_k)
\end{aligned}$$

which corresponds to the data-matrix 65.

The right-hand side of equation 73 (the criterion differentiated with respect to b_j^m) attributes the following weight to the filter coefficient b_l^u:

$$\begin{aligned}
&\Re\left(\exp(il\omega_k)\Xi_{TW_u}(\omega_k)\right)\Re\left(\exp(ij\omega_k)\Xi_{TW_m}(\omega_k)\right)\\
&+\sqrt{(1+\lambda\Gamma(\omega_k))}\Im\left(\exp(il\omega_k)\Xi_{TW_u}(\omega_k)\right)\Im\left(\exp(ij\omega_k)\Xi_{TW_m}(\omega_k)\right)
\end{aligned}$$

where, once again, we assume that $\Xi_{TW_0}(\omega_k)=\Xi_{TX}(\omega_k)$ for $u=0$. This generalized $\mathbf{X}'\mathbf{X}$-matrix reduces to 65 in the univariate case. This expression is used in the published R-code.

References

[1] McElroy T and Wildi, M. (2012) On a Trilemma Between Accuracy, Timeliness and Smoothness in Real-Time Forecasting and Signal Extraction. *Soon on SEFBlog.*

[2] Wildi, M. (1998) Detection of Compatible Turning Points and Signal Extraction for Non-Stationary Time Series. *Operation Research Proceedings, Springer.*

[3] Wildi, M. (2004) Signal Extraction: How (In)efficient Are Model-Based Approaches? An Empirical Study Based on TRAMO/SEATS and Census X-12-ARIMA. *KOF-Working Paper Nr. 96*, ETH-Zurich.

[4] Wildi, M. (2005) Signal Extraction: Efficient Estimation, 'Unit-Root'-Tests and Early Detection of Turning-Points. *Lecture Notes in Economic and Mathematical Systems, 547*, Springer-Verlag Berlin Heidelberg.

[5] Wildi, M. (2008) Real-Time Signal-Extraction: Beyond Maximum Likelihood Principles. http://www.idp.zhaw.ch/fileadmin/user_upload/engineering/_Institute_und_Zentren/IDP/forschungsschwerp

[6] Wildi, M. (2008.2) Efficient Multivariate Real-Time Filtering and Cointegration. *IDP-working paper, IDP-WP-08Sep-01.* http://www.idp.zhaw.ch/fileadmin/user_upload/engineering/_Institute_und_Zentren/IDP/forschungsschwerp WP-08Sep-01.pdf

[7] Wildi, M. (2009) Real-Time US-Recession Indicator (USRI): A Classical Cycle Perspective with Bounceback. *IDP-working paper, IDP-WP-09Jun-06.* http://www.idp.zhaw.ch/en/engineering/idp-institute-of-data-analysis-and-process-design/research/finance-risk-management-and-econometrics/economic-indices/us-economic-recession-indicator.html

[8] Wildi, M. (2010) Real-Time Signalextraction: a Shift of Perspective. *Estudios de Economia Aplicada. Vol 28-3, 497-518.* Working paper: http://blog.zhaw.ch/idp/sefblog/uploads/customized_final.pdf