Elements of Forecasting and Signal Extraction

Marc Wildi

Institute of Data Analysis and Process Design

October 23, 2014

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 10 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

$$X_t = \frac{\sqrt{2\pi}}{\sqrt{T}} \sum_{k=-[T/2]}^{[T/2]} w_k \Xi_{TX}(\omega_k) \exp(it\omega_k)$$

where

- $[T/2]$ is the greatest integer smaller or equal to $T/2$.

- $\Xi_{TX}(\omega_k) := \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t \exp(-it\omega_k)$ is the Discrete Fourier Transform (DFT) of X_t

- $\omega_k = \frac{k2\pi}{T}$, $k = -T/2, ..., 0, ..., T/2$ is a discrete frequency-grid in the interval $[-\pi, \pi]$

- the weights $w_k = \begin{cases} 1 & , [-T/2] \leq k \leq [T/2] \text{ if } T \text{ is odd} \\ \begin{cases} 1 & |k| < T/2 \\ 1/2 & |k| = T/2 \end{cases} & \text{if } T \text{ is even} \end{cases}$. Not to be confounded with ω_k.

Note that the weights w_k are needed in the case of even T only and then their effect is negligible by all practical means: therefore we can neglect them (as is frequently done in the literature on the topic). Given the above decomposition, linearity of the filter can 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, for example.

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

---

[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.

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).

3.1 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

---

[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.

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 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,

---

[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.

various degrees of risk aversion.

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

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

If we assume that Γ is symmetric and positive, then $\Phi(\omega) \equiv 0$. Inserting 13 into 11 and using $1 - \cos(\hat{\Phi}(\omega)) = 2\sin(\hat{\Phi}(\omega)/2)^2$ 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

$$
W(\omega_k, \eta, \text{cutoff}) = \begin{cases} 1 \;, \text{ if } |\omega_k| < \text{cutoff} \\ (1 + |\omega_k| - \text{cutoff})^\eta \;, \text{ otherwise} \end{cases}
\tag{17}
$$

---

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

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 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:

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

(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 compar-

---

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

ison of 16 and 18:

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

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

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

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

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

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 $\Xi_{TX}(\omega_k)$:

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

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

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

where

$$\hat{\Gamma}_X(\omega_k) = \sum_{j=0}^{L} b_{Xj} \exp(-ij\omega_k) \qquad (23)$$

$$\hat{\Gamma}_{W_n}(\omega_k) = \sum_{j=0}^{L} b_{w_nj} \exp(-ij\omega_k) \qquad (24)$$

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

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

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:

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

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

$$=\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\left|\Gamma(\omega_k)-\tilde{\Gamma}(\omega_k)\right|^2|\Xi_{TX}(\omega_k)|^2 \qquad (26)$$

where

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

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

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

$$+4\lambda\frac{2\pi}{T}\sum_{k=-T/2}^{T/2}A(\omega_k)\tilde{A}(\omega_k)^2\sin(\tilde{\Phi}(\omega_k))^2 W(\omega_k,\eta)\,|\Xi_{TX}(\omega_k)|^2\to\min \qquad (28)$$

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.

10

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 10. 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   Matrix Notation (Frequency Domain) and 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 $b_{w_n j}$,

11

$j = 0, ..., l$, 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:

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

where the Vec_{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

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

where Vec_{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

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

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):

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

$$= \frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\Bigg||\Gamma(\omega_k)\Xi_{TX}(\omega_k)|\exp\left(i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right)-\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)$$

$$-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\Bigg|^2$$

$$= \frac{2\pi}{T}\sum_{k=-T/2}^{T/2}\Bigg||\Gamma(\omega_k)\Xi_{TX}(\omega_k)|-\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)\exp\left(-i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right)$$

$$-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i*\arg\left(\Gamma(\omega_k)\Xi_{TX}(\omega_k)\right)\right)\Bigg|^2 \tag{34}$$

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)

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

---

[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 136 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.

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

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

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. Equating the previous expression to zero provides the *correct* 'linear regression' estimate (as derived in the R-code):

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

(note that $\Re(\mathbf{X}_{\text{rot}})' = \Re(\mathbf{X}_{\text{rot}})^{\mathbf{T}}$). 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:

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

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

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, $\Gamma(\omega) |\Xi_{TX}(\omega_k)|$ 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:

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

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

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

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

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

---

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

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 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.

$$
\begin{aligned}
\hat{\mathbf{b}} &= \left(\Re\left(\mathbf{X}'_{\mathrm{rot}}\mathbf{X}_{\mathrm{rot}}\right)\right)^{-1}\Re(\mathbf{X}_{\mathrm{rot}})'\mathbf{Y}_{\mathrm{rot}} \\
&= \left(\Re\left(\mathbf{X}'_{\mathrm{rot}}\mathbf{X}_{\mathrm{rot}}\right)\right)^{-1}\left\{\Re(\mathbf{X}_{\mathrm{rot}})'\Re(\mathbf{Y}_{\mathrm{rot}}) + \Im(\mathbf{X}_{\mathrm{rot}})'\Im(\mathbf{Y}_{\mathrm{rot}})\right\} \\
&= \left(\Re\left(\mathbf{X}'_{\mathrm{rot}}\mathbf{X}_{\mathrm{rot}}\right)\right)^{-1}\left\{\Re(\mathbf{X}_{\mathrm{rot}})'\Re(\mathbf{X}_{\mathrm{rot}}\mathbf{b}+\mathbf{e}) + \Im(\mathbf{X}_{\mathrm{rot}})'\Im(\mathbf{X}_{\mathrm{rot}}\mathbf{b}+\mathbf{e})\right\} \\
&= \left(\Re\left(\mathbf{X}'_{\mathrm{rot}}\mathbf{X}_{\mathrm{rot}}\right)\right)^{-1}\left\{\Re\left[(\mathbf{X}_{\mathrm{rot}})'\mathbf{X}_{\mathrm{rot}}\right]\mathbf{b} + \Re(\mathbf{X}_{\mathrm{rot}})'\Re(\mathbf{e}) + \Im(\mathbf{X}_{\mathrm{rot}})'\Im(\mathbf{e})\right\} \\
&= \mathbf{b} + \left(\Re\left(\mathbf{X}'_{\mathrm{rot}}\mathbf{X}_{\mathrm{rot}}\right)\right)^{-1}\left\{\Re(\mathbf{X}_{\mathrm{rot}})'\Re(\mathbf{e}) + \Im(\mathbf{X}_{\mathrm{rot}})'\Im(\mathbf{e})\right\}
\end{aligned}
$$

where $\mathbf{e} = \mathbf{Y}_{\mathrm{rot}} - \mathbf{X}_{\mathrm{rot}}\mathbf{b}$ 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 $\mathbf{Y}_{\mathrm{rot}}$ is a real (positive) vector and therefore $\Im(\mathbf{Y}_{\mathrm{rot}}) = \mathbf{0}$. 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

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

---

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

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

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.

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

---

[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).

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].

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

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:

$$
\mathbf{b} = \mathbf{A}\tilde{\mathbf{b}} \tag{43}
$$

$$
\mathbf{A} = \begin{pmatrix} \mathbf{Id} & -\mathbf{Id} & -\mathbf{Id} & ... & ... & -\mathbf{Id} \\ \mathbf{Id} & \mathbf{Id} & \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{Id} & \mathbf{0} & \mathbf{Id} & \mathbf{0} & ... & \mathbf{0} \\ \vdots\vdots\vdots & & & & & \\ \mathbf{Id} & \mathbf{0} & \mathbf{0} & ... & ... & \mathbf{Id} \end{pmatrix}
$$

$$
\mathbf{b}' = \left(b_0^0, b_1^0, ..., b_L^0 \mid b_0^1, b_1^1, ..., b_L^1 \mid ... \mid ... \mid ... \mid b_0^m, b_1^m, ..., b_L^m \right)'
$$

$$
\tilde{\mathbf{b}}' = \left(b_0, b_1, ..., b_L \mid \delta b_0^1, \delta b_1^1, ..., \delta b_L^1 \mid \delta b_0^2, \delta b_1^2, ..., \delta b_L^2 \mid ... \mid ... \mid ... \mid \delta b_0^m, \delta b_1^m, ..., \delta b_L^m \right)'
$$

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

$$
(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})
$$
$$
+ \lambda_{\text{smooth}}\mathbf{b}'\mathbf{Q}_{\text{smooth}}\mathbf{b} + \lambda_{\text{cross}}\mathbf{b}'\mathbf{Q}_{\text{cross}}\mathbf{b} + \lambda_{\text{decay}}\mathbf{b}'\mathbf{Q}_{\text{decay}}\mathbf{b} \to \min \tag{44}
$$

---

[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}.

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.

$$\mathbf{Q_{smooth}} = \begin{pmatrix} \mathbf{Q} & \mathbf{0} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & ... \\ ::: & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & \mathbf{Q} \end{pmatrix}$$

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

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

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

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 $\tilde{\mathbf{b}}'\tilde{\mathbf{Q}}_{\mathbf{cross}}\tilde{\mathbf{b}}$ sums up the squared deviations ('effects') δb_l^u i.e.

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

---

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

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 $\lambda_{\mathrm{cross}} \to \infty$, filter coefficients must be identical. To conclude we specify the decay-requirement in our Troika.

$$
\mathbf{Q_{decay}} = \begin{pmatrix} \mathrm{diag}(\mathbf{q}) & 0 & 0 & 0 & \ldots & 0 \\ 0 & \mathrm{diag}(\mathbf{q}) & 0 & 0 & \ldots & 0 \\ 0 & 0 & \mathrm{diag}(\mathbf{q}) & 0 & \ldots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \ldots & \mathrm{diag}(\mathbf{q}) \end{pmatrix}
$$

$$
\mathbf{q} = (q^{\max(0,h)}, q^{|1-\max(0,h)|}, q^{|2-\max(0,h)|}, ..., q^{|L-\max(0,h)|}) \tag{47}
$$

where $q > 1$ and $\mathrm{diag}(\mathbf{q})$ 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 and

$$
\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b} = \sum_{u=0}^{m} \sum_{l=0}^{L} q^l (b_l^u)^2 \tag{48}
$$

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

$$
q := 1 + \lambda_{\mathrm{decay}} \tag{49}
$$

as currently implemented in our R-code[17].

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 48 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).

---

[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.

[17]At least this last statement applied up to July 2012: since then we splitted both effects, see section 7.2.2 below: on 30-Oct.2012 new code and a tutorial dedicated to this topic were posted on SEFBlog.

7.2 Potential Flaws of the Previously Defined Troika-Terms and (straightforward) Solutions

7.2.1 The Cross-Sectional Term

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:

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

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

$$
\sum_{u=0}^{m} \left(\left(b_0^u - \frac{1}{m+1} \sum_{u'=0}^{m} b_0^{u'} \right)^2 + \left(b_1^u - \frac{1}{m+1} \sum_{u'=0}^{m} b_1^{u'} \right)^2 + ... + \left(b_L^u - \frac{1}{m+1} \sum_{u'=0}^{m} b_L^{u'} \right)^2 \right) \tag{50}
$$

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

$$
\mathbf{Q}_{\mathbf{cross}} = \begin{pmatrix} \mathbf{q}_{\mathbf{cross,1}}' \\ \mathbf{q}_{\mathbf{cross,2}}' \\ ... \\ \mathbf{q}_{\mathbf{cross,(m+1)*(L+1)}}' \end{pmatrix} \tag{51}
$$

where

$$\mathbf{q_{cross,1}}' = (1 - \frac{1}{m+1}, 0, ..., 0 \mid -\frac{1}{m+1}, 0, ..., 0 \mid -\frac{1}{m+1}, 0, ..., 0 \mid ...)$$

$$\mathbf{q_{cross,2}}' = (0, 1 - \frac{1}{m+1}, 0, ..., 0 \mid 0, -\frac{1}{m+1}, 0, ..., 0 \mid 0, -\frac{1}{m+1}, 0, ..., 0 \mid ...)$$

$$\mathbf{q_{cross,3}}' = (0, 0, 1 - \frac{1}{m+1}, 0, ..., 0 \mid 0, 0, -\frac{1}{m+1}, 0, ..., 0 \mid 0, 0, -\frac{1}{m+1}, 0, ..., 0 \mid ...)$$

$$...$$

$$\mathbf{q_{cross,(m+1)*(L+1)}}' = (0, 0, ..., -\frac{1}{m+1} \mid 0, 0, ..., -\frac{1}{m+1} \mid 0, 0, ..., -\frac{1}{m+1} \mid ... \mid 0, 0, ..., 1 - \frac{1}{m+1})$$

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 50.

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: we give access to both parametrizations through a boolean called 'grand-mean': if grand-mean=F then the new (symmetric) parametrization 51 is selected, otherwise the previous (asymmetric) grand-mean encoding 43 is used.

7.2.2 The Decay-Term

The decay-term defined by λ_{decay} and $\mathbf{Q_{decay}}$ in section 7.1 mixes-up and confounds two different effects: λ_{decay} controls for the *strength* of the decay-regularization in 44 as well as for the *shape* of the decay along the diagonal of $\mathbf{Q_{decay}}$ (by setting $q := 1 + \lambda_{\text{decay}}$ in 52). With hindsight we found this concatenation to be 'confusing' and in the end untenable. We therefore split the decay-term of the Troika into two separate and distinguished effects controlled by a two-dimensional vector $\boldsymbol{\lambda}_{\text{decay,shape}} = (\lambda_{\text{decay}}, \lambda_{\text{shape}})$ where the first component λ_{decay} controls for the *strength* of the decay-term in 44 and the second component λ_{shape} parametrizes the *shape* of the regularization constraint in 52

$$q := 1 + \lambda_{\text{shape}} \tag{52}$$

As an example, setting $\lambda_{\text{shape}} = 0$ would enforce a homogeneous shrinkage of all filter coefficients, irrespective of their lag, towards zero: $\mathbf{Q_{decay}}$ as defined by the $\mathbf{q}$-vector in 47 would be an identity i.e. the diagonal elements would be constant. Moreover, λ_{decay} would then define the priority that we are likely to assign to this 'particularly shaped' regularization. On the other hand, a 'large' value of λ_{shape} would lead to a very strong emphasize of high-lag coefficients (weighting the remote

past of the time series) and therefore the resulting regularization would lead to a strong shrinkage of high-lag coefficients - only -, as emphasized by the diagonal of $\mathbf{Q_{decay}}$ defined by 47[18].

7.3 General H_0-Shrinkage

A brief look at 44 confirms that the Regularization Troika imposes a shrinkage of filter coefficients towards zero: if all coefficients vanish, then the regularization terms vanish equally[19]. Obviously, this zero-shrinkage may conflict with some general hypotheses about the 'true' parameters (a concrete example is proposed in section 12.1.1). So let's assume that we entertain such a vision – a null-hypothesis – expressed in the form:

$$\mathbf{b_c} = \mathbf{b_0} + \mathbf{b_{res}}$$

where the 'completed' effective filter parameter $\mathbf{b_c}$ can be decomposed into a hypothetical $\mathbf{b_0}$ (not to be estimated) and an unknown 'residual' $\mathbf{b_{res}}$ which is to be estimated. If our null-hypothesis is pertinent, then we expect the residual $\mathbf{b_{res}}$ to be 'small' and thus zero-shrinkage would be a beneficial feature – a positive asset – i.e. the Regularization Troika would not induce 'misspecification' anymore.

7.3.1 A 'Bad' Good Idea

One could try to solve the aforementioned problem by focusing the estimation criterion 44 on $\mathbf{b_{res}}$ (instead of $\mathbf{b_c}$). Specifically, define a 'residual' target $\mathbf{Y_{res}}$ as follows

$$\mathbf{Y_{res}} := \mathbf{Y} - \mathbf{Xb_0} \tag{53}$$

We can now substitute $\mathbf{Y_{res}}$ to $\mathbf{Y}$ in section 5.1 and obtain a rotated (real) target

$$\mathbf{Y_{rot,res}} = |\mathbf{Y_{res}}|$$

and a new rotated 'residual' design matrix:

$$\mathbf{X}_{k,\mathrm{rot,res}} = \mathbf{X}_k \exp\left(-i * \arg\left(\mathbf{Y_{k,res}}\right)\right) \tag{54}$$

where $\mathbf{X}_{k,\mathrm{rot, res}}$ designates the k-th row (of the new rotated design matrix). At this stage the whole I-MDFA formalism applies and the general criterion 44, applied to $\mathbf{b_{res}}$, becomes

$$(\mathbf{Y}_{\mathrm{rot,res}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot,res}}^{\mathrm{Cust}}\mathbf{b_{res}})'(\mathbf{Y}_{\mathrm{rot,res}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot,res}}^{\mathrm{Cust}}\mathbf{b_{res}})$$
$$+ \quad \lambda_{\mathrm{smooth}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{smooth}}\mathbf{b_{res}} + \lambda_{\mathrm{cross}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{cross}}\mathbf{b_{res}} + \lambda_{\mathrm{decay}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{decay}}\mathbf{b_{res}} \to \min$$

---

[18]See a corresponding tutorial introduced Oct.30, 2012 on SEFBlog.

[19]More precisely: the decay-term imposes a shrinkage of original coefficients (in levels) whereas smoothness and cross-sectional terms impose shrinkage of *differenced* coefficients: longitudinal and lateral differences, respectively.

In particular, the zero-shrinkage entailed by the Regularization Troika now applies to $\mathbf{b}_{\mathrm{res}}$, as desired.

OK, but this idea is not very clever! Why? Please cogitate a bit before looking at the formally correct solution...

7.3.2 Criteria for General H_0-Shrinkage

When customizing the filter we are not interested in addressing the phase between the(rotated) filter-residual $\mathbf{Y}_{\mathrm{rot,res}}$ and the residual filter output obtained by $\mathbf{b}_{\mathrm{res}}$, as suggested in the previous section. Instead, we want to emphasize the phase-differential between the original (rotated) target $\mathbf{Y}_{\mathrm{rot}}$ and the output of the completed effective filter defined by $\mathbf{b_c}$: the output of the real-time filter should be synchronized with the original target! The solution in the previous section tackles another problem[20] (which might be interesting too but it's not our primary concern). Consequently the right criterion is

$$
\begin{aligned}
& (\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\mathbf{b_0} + \mathbf{b}_{\mathrm{res}}))'(\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}(\mathbf{b_0} + \mathbf{b}_{\mathrm{res}})) \\
+ \quad & \lambda_{\mathrm{smooth}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{smooth}}\mathbf{b}_{\mathrm{res}} + \lambda_{\mathrm{cross}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{cross}}\mathbf{b}_{\mathrm{res}} + \lambda_{\mathrm{decay}}\mathbf{b}'_{\mathrm{res}}\mathbf{Q_{decay}}\mathbf{b}_{\mathrm{res}} \to \min_{\mathbf{b_{res}}}
\end{aligned} \tag{55}
$$

Note that regularization (the second line in the above criterion) is applied to $\mathbf{b}_{\mathrm{res}}$ and not to $\mathbf{b_c} = \mathbf{b_0} + \mathbf{b}_{\mathrm{res}}$: the Regularization Troika is now compatible with a shrinkage of $\mathbf{b_c} = \mathbf{b_0} + \mathbf{b}_{\mathrm{res}}$ towards $\mathbf{b_0}$. Moreover, customization addresses the phase-differential between the original target $\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}}$ and the real-time filter, as desired.

Equivalently, we could express the optimization problem in the following form

$$
\begin{aligned}
& (\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}\mathbf{b_c})'(\mathbf{Y}_{\mathrm{rot}}^{\mathrm{Cust}} - \mathbf{X}_{\mathrm{rot}}^{\mathrm{Cust}}\mathbf{b_c}) \\
+ \quad & \lambda_{\mathrm{smooth}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{smooth}}(\mathbf{b_c} - \mathbf{b_0}) + \lambda_{\mathrm{cross}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{cross}}(\mathbf{b_c} - \mathbf{b_0}) \\
& + \lambda_{\mathrm{decay}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{decay}}(\mathbf{b_c} - \mathbf{b_0}) \to \min_{\mathbf{b_c}}
\end{aligned} \tag{56}
$$

Note that minimization is achieved with respect to $\mathbf{b_c}$ (in contrast to 55 which addresses $\mathbf{b_{res}}$). Once again, shrinkage by the Troika is compatible with the null-hypothesis H_0. Criterion 55 addresses $\mathbf{b_{res}}$ and therefore $\mathbf{b_0}$ must be added to the obtained solution, see the appendix. In contrast, criterion 56 addresses $\mathbf{b_c}$ 'directly' and therefore we don't need to add $\mathbf{b_0}$ to the obtained solution.

Critera 44 (zero-shrinkage) and 55 or 56 (general H_0-shrinkage) are *regularized* and *customized* versions of the original mean-square criterion 36: setting $\lambda = \eta = \lambda_{\mathrm{smooth}} = \lambda_{\mathrm{cross}} = \lambda_{\mathrm{decay}} = 0$ just

---

[20]Basically because the phase is not a linear operator.

replicates plain MS-performances. They are also *regularized* versions of the customized criterion 39: setting $\lambda_{\text{smooth}} = \lambda_{\text{cross}} = \lambda_{\text{decay}} = 0$ just replicates customized designs. We now derive explicit solutions.

7.3.3 Solution: Zero-Shrinkage

As explained in the previous section, the Regularization Troika can be made compatible with a shrinkage of the estimated filter coefficient-vector $\mathbf{b_c} = \mathbf{b_0} + \mathbf{b}_{\text{res}}$ towards a pre-specified null-hypothesis H_0 defined by $\mathbf{b_0}$. If $H_0 : \mathbf{b_0} = \mathbf{0}$ then we may focus on criterion 44 (the simpler variant of the general criteria 55 or 56). Consider

$$(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b})$$
$$+ \quad \lambda_{\text{smooth}}\mathbf{b}'\mathbf{Q_{smooth}}\mathbf{b} + \lambda_{\text{cross}}\mathbf{b}'\mathbf{Q_{cross}}\mathbf{b} + \lambda_{\text{decay}}\mathbf{b}'\mathbf{Q_{decay}}\mathbf{b} \to \min \qquad (57)$$

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

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

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

$$\hat{\mathbf{b}}^{\text{Cust-Reg}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, \mathbf{b_0} = \mathbf{0}) = \qquad\qquad (58)$$
$$\left(\Re\left\{(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\right\} + \lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}} \right)^{-\mathbf{1}} \Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{Y}_{\text{rot}}^{\text{Cust}}$$

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[21] which analyzes partial and combined effects when applied to the EURI.

---

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

7.3.4 Solution: General H_0-Shrinkage

In general we prefer criterion 56 (over 55) because it can be generalized to allow for important filter constraints, see section 8.3.2. A formal solution of 55 is proposed in the appendix: both criteria lead to identical solutions so that we may ignore the latter. Let's now focus on

$$(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_c})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_c})$$
$$+ \quad \lambda_{\text{smooth}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{smooth}}(\mathbf{b_c} - \mathbf{b_0}) + \lambda_{\text{cross}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{cross}}(\mathbf{b_c} - \mathbf{b_0})$$
$$+\lambda_{\text{decay}}(\mathbf{b_c} - \mathbf{b_0})'\mathbf{Q_{decay}}(\mathbf{b_c} - \mathbf{b_0}) \rightarrow \min_{\mathbf{b_c}}$$

We obtain

$$
\begin{aligned}
d/d\mathbf{b_c} \text{ Criterion} \quad &= \quad d/d\mathbf{b_c} \ (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_c})'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_c}) \\
&+ \quad d/d\mathbf{b_c} \ \lambda_{\text{smooth}}(\mathbf{b_c} - \mathbf{b_0})' \, \mathbf{Q_{smooth}} \, (\mathbf{b_c} - \mathbf{b_0}) \\
&+ \quad d/d\mathbf{b_c} \ \lambda_{\text{cross}}(\mathbf{b_c} - \mathbf{b_0})' \, \mathbf{Q_{cross}} \, (\mathbf{b_c} - \mathbf{b_0}) \\
&+ \quad d/d\mathbf{b_c} \ \lambda_{\text{decay}}(\mathbf{b_c} - \mathbf{b_0})' \, \mathbf{Q_{decay}} \, (\mathbf{b_c} - \mathbf{b_0}) \\
&= \quad -2(\mathbf{Y}_{\text{rot}}^{\text{Cust}})'\Re\left(\mathbf{X}_{\text{rot}}^{\text{Cust}}\right) + 2\mathbf{b_c'}\Re\left\{(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\right\} \\
&\quad +2\lambda_{\text{smooth}}\mathbf{b_c'}\mathbf{Q_{smooth}} - 2\lambda_{\text{smooth}}\mathbf{b_0'}\mathbf{Q_{smooth}} \\
&\quad +2\lambda_{\text{cross}}\mathbf{b_c'}\mathbf{Q_{cross}} - 2\lambda_{\text{cross}}\mathbf{b_0'}\mathbf{Q_{cross}} \\
&\quad +2\lambda_{\text{decay}}\mathbf{b_c'}\mathbf{Q_{decay}} - 2\lambda_{\text{decay}}\mathbf{b_0'}\mathbf{Q_{decay}}
\end{aligned}
$$

Equating this expression to zero, the *customized and regularized* H_0-shrinkage solution is obtained as

$$
\begin{aligned}
&\hat{\mathbf{b}}_c^{\text{Cust-Reg}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, \mathbf{b_0}) \\
&= \quad \left\{\Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\right] + \lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\right\}^{-1} \\
&\quad \left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}))'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \left(\lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\right)\mathbf{b_0}\right) \\
&= \quad \left\{\Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\right] + \lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\right\}^{-1} \\
&\quad \left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}))'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const}\right) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (59)
\end{aligned}
$$

where

$$\mathbf{Const} = \left\{\lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\right\}\mathbf{b_0}$$

The constant depends on the null-hypothesis H_0. In particular, if $H_0 : \mathbf{b_0} = \mathbf{0}$, then the constant vanishes and 59 simplifies to 58.

8 Filter Constraints: the Case of no Cointegration

We here assume that the user would like to apply meaningful constraints to the individual filters which reflect 'desirable' real-time properties such as level-match or time-shift constraints, see below for details. The case of 'no cointegration' emphasizes a particular framework in which restrictions imposed to a filter are independent from restrictions imposed to the other ones. In contrast, the so-called 'cointegration case' links constraints across filters, see section 9.

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 solutions 58 and 59 are always some kind of compromise between conflicting requirements which reflects the Helvetic origin of its Designer i.e. of myself. 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 they are 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, see the cointegration case in section 9. The time-shift constraint imposes a vanishing time-shift of filters in frequency zero: whereas this requirement is a natural outcome in the case of symmetric (positive) 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[22]. Our approach and our perspective here are different - more general - because we treat level and time-shift constraints as 'independent' user-requirements: we can impose none, either one, or both, irrespective of 'unit-roots' of the DGP. We feel that the merits of these constraints are closely linked to research priorities, as assigned by the user, rather than to abstract properties of the DGP[23].

---

[22]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.

[23]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.

Prosaically, the first order (level-) restriction is

$$b_0^u + b_1^u + ... + b_L^u = w^u \tag{60}$$

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: more generally, we here propose to impose an arbitrary time-shift constraint to series u, say s_u, where $s^u = 0$ just replicates the zero-shift. For this purpose the derivative of the transfer function in frequency zero must equate to s^u: $\frac{\partial}{\partial \omega}\big|_{\omega=0} \sum_{j=0}^{L} b_j^u \exp(-ij\omega) = s^u$. This condition results in the following coefficient constraint:

$$b_1^u + 2b_2^u + 3b_3^u + ... + Lb_L^u = s^u \tag{61}$$

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 61 (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:

$$b_{L-1}^u = -Lb_0^u - (L-1)b_1^u - ... - 2b_{L-2}^u + Lw^u - s^u \tag{62}$$

$$b_L^u = (L-1)b_0^u + (L-2)b_1^u + (L-4)b_3^u + ... + b_{L-2}^u - (L-1)w^u + s^u \tag{63}$$

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 62 and 63 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: a shift of s^u time units is

assigned to time point $T - h$. Formally, this requirement is obtained by equating the derivative of the smoothing filter to s^u in frequency zero: $\frac{\partial}{\partial \omega}\big|_{\omega=0} \sum_{j=-h}^{L-h} b_j^u \exp(-ij\omega) = s^u$:

$$(-h)b_{-h}^u + (1-h)b_{1-h}^u + (2-h)b_{2-h}^u + ... + (-1)b_{-1}^u + b_1^u + 2b_2^u + ... + (L-h)b_{L-h}^u = s^u \quad (64)$$

The level constraint becomes:

$$b_{-h}^u + b_{-(h-1)}^u + ... + b_{L-h}^u = w^u \quad (65)$$

Evidently, given these restrictions (equations), any of the coefficients $b_{-h}^u, ..., b_{L-h}^u$ could be selected for implementing the constraint(s): equations 62 and 63 considered b_{L-1}^u and b_L^u 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_{L-h}^u or b_{-h}^u (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_{L-h}^u or b_h^u, 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

$$\mathbf{b} = \mathbf{R}\mathbf{b_f} + \mathbf{c} \quad (66)$$

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 (no time-shift constraint but a level constraint) we obtain

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

and the entries in 66 become

$$\mathbf{R} = \begin{pmatrix} \mathbf{C} & 0 & ... & 0 \\ 0 & \mathbf{C} & ... & 0 \\ ::: & & & \\ 0 & 0 & ... & \mathbf{C} \end{pmatrix} \tag{67}$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ -1 & -1 & -1 & ... & -1 & -1 \\ ::: & & & & & \\ 0 & 0 & 0 & ... & 1 & 0 \\ 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix} \tag{68}$$

$$\mathbf{c}' = (0, ..., 0, w^0, 0, ..., 0 \ || \ 0, ..., 0, w^1, 0, ..., 0 \ || \ ... \ || 0, ..., 0, w^m, 0, ..., 0)$$

$$\mathbf{b_f}' = (b^0_{-h}, ..., b^0_{-1}, b^0_1, ..., b^0_{L-h} \ || \ b^1_{-h}, ..., b^1_{-1}, b^1_1, ..., b^1_{L-h} \ || \ ... \ || \ b^m_{-h}, ..., b^m_{-1}, b^m_1, ..., b^m_{L-h})$$

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^u_0 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[24]. Then, from 64 we obtain

$$b^u_1 = s^u - (-h)b^u_{-h} - (1-h)b^u_{1-h} - ... - (-1)b^u_{-1} - 0b^u_0 - 2b^u_2 - ... - (L-h)b^u_{L-h}$$

Such that

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & ... & ... & ... & ... & 0 & 0 \\ 0 & 1 & 0 & ... & ... & ... & ... & 0 & 0 \\ ::: & & & & & & & & \\ h & h-1 & h-2 & ... & 1 & 0 & -2 & ... & -(L-h) \\ ::: & & & & & & & & \\ 0 & 0 & 0 & ... & ... & ... & ... & 1 & 0 \\ 0 & 0 & 0 & ... & ... & ... & ... & 0 & 1 \end{pmatrix} \tag{69}$$

$$\mathbf{c}' = (0, ..., 0, s^0, 0, ..., 0 \ || \ 0, ..., 0, s^1, 0, ..., 0 \ || \ ... \ || 0, ..., 0, s^m, 0, ..., 0)$$

$$\mathbf{b_f}' = (b^0_{-h}, ..., b^0_0, b^0_2, ..., b^0_{L-h} \ || \ b^1_{-h}, ..., b^1_0, b^1_2, ..., b^1_{L-h} \ || \ ... \ || \ b^m_{-h}, ..., b^m_0, b^m_2, ..., b^m_{L-h})$$

---

[24]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.

in 66. 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). Analogously, the constant s^u $(u = 0, ..., m)$ in $\mathbf{c}$ is located in position $u * L + h + 2$ (instead of $u * L + h + 1$ for the previous level constraint). 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

$$
\mathbf{C}_{h=0} \;=\; \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & -2 & -3 & ... & -L \\ 0 & 1 & 0 & ... & 0 \\ \vdots \\ 0 & 0 & 0 & ... & 1 \end{pmatrix}
$$

The last case i1=i2=T assumes that both constraints are imposed. Solving for b_0^u and b_1^u in 64 and 65 leads to

$$
\begin{aligned}
b_1^u &= s^u + hb_{-h}^u + (h-1)b_{-(h-1)}^u + ... + b_{-1}^u - 2b_2^u - 3b_3 - ... - (L-h)b_{L-h}^u \\
b_0^u &= w^u - s^u - (h+1)b_{-h}^u - hb_{-(h-1)}^u - ... - 2b_{-1}^u + b_2^u + 2b_3^u + ... + ((L-1)-h)b_{L-h}^u
\end{aligned}
$$

In our general notation 66 we obtain

$$
\mathbf{C} \;=\; \begin{pmatrix} 1 & 0 & 0 & ... & & 0 & 0 \\ 0 & 1 & 0 & ... & & 0 & 0 \\ \vdots \\ -(h+1) & -h & -(h-1) & ... & -2 \quad 1 \quad 2 & ... & (L-1-h) \\ h & h-1 & h-2 & ... & 1 \quad -2 \quad -3 & ... & -(L-h) \\ \vdots \\ 0 & 0 & 0 & ... & & 1 & 0 \\ 0 & 0 & 0 & ... & & 0 & 1 \end{pmatrix} \tag{70}
$$

$$
\begin{aligned}
\mathbf{c}' &= (0, ..., 0, w^0 - s^0, s^0, 0, ..., 0 \;||\; 0, ..., 0, w^1 - s^1, s^1, 0, ..., 0 \;||\; ... \;||0, ..., 0, w^m - s^m, s^m, 0, ..., 0) \\
\mathbf{b_f}' &= (b_{-h}^0, ..., b_{-1}^0, b_2^0, ..., b_{L-h}^0 \;||\; b_{-h}^1, ..., b_{-1}^1, b_2^1, ..., b_{L-h}^1 \;||\; ... \;||\; b_{-h}^m, ..., b_{-1}^m, b_2^m, ..., b_{L-h}^m)
\end{aligned}
$$

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 $u * (L-1) + h + 1$ (and s^u accordingly). 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

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

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

Obviously, all entities on the right-hand side of 66 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

The case of general H_0-shrinkage has been proposed in sections 7.3 and 7.3.4. We here distinguish accordingly the (simpler) case of zero-shrinkage and the more general case of arbitrary H_0-shrinkage.

8.3.1 The Case of Zero-Shrinkage

In order to derive the constrained customized and regularized parameter vector we can plug 66 into 44 and take derivatives

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

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

$$
\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2, \mathbf{b_0} = \mathbf{0})
$$

$$
= \left\{ \Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} \right] + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} \right\}^{-1}
$$

$$
\left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \Re\left\{ (\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{c} \right\} - \mathbf{R}'(\lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{cross}})\mathbf{c} \right)
$$

$$
= \left\{ \Re\left[\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} \right] + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} \right\}^{-1}
$$

$$
\left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const} \right) \tag{71}
$$

where

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

This last term collects all level constraints. A comparison with 58 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 $\mathbf{Const}$: setting $\mathbf{R} = \mathbf{Id}$ and $\mathbf{c} = \mathbf{0}$ (no constraints involved) in the above solution indeed replicates 58. 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 66.

8.3.2   General H_0-Shrinkage

In the case of general H_0-shrinkage we consider the following derivative:

$$
\begin{aligned}
d/d\mathbf{b_f} \text{ Criterion} \quad = \quad & d/d\mathbf{b_f}\ (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{Rb_f} + \mathbf{c}))'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{Rb_f} + \mathbf{c})) \\
+\quad & d/d\mathbf{b_f}\ \lambda_{\text{smooth}}(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0})'\,\mathbf{Q_{smooth}}\,(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0}) \\
+\quad & d/d\mathbf{b_f}\ \lambda_{\text{cross}}(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0})'\,\mathbf{Q_{cross}}\,(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0}) \\
+\quad & d/d\mathbf{b_f}\ \lambda_{\text{decay}}(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0})'\,\mathbf{Q_{decay}}\,(\mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0}) \\
=\quad & -2(\mathbf{Y}_{\text{rot}}^{\text{Cust}})'\Re\left(\mathbf{X}_{\text{rot}}^{\text{Cust}}\right)\mathbf{R} - 2\Re\left\{ (\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{c})'(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R}) \right\} + 2\mathbf{b_f}'\Re\left\{ (\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R} \right\} \\
& +2\lambda_{\text{smooth}}\mathbf{b_f}'\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + 2\lambda_{\text{smooth}}(\mathbf{c} - \mathbf{b_0})'\mathbf{Q_{smooth}}\mathbf{R} \\
& +2\lambda_{\text{cross}}\mathbf{b_f}'\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + 2\lambda_{\text{cross}}(\mathbf{c} - \mathbf{b_0})'\mathbf{Q_{cross}}\mathbf{R} \\
& +2\lambda_{\text{decay}}\mathbf{b_f}'\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} + 2\lambda_{\text{decay}}(\mathbf{c} - \mathbf{b_0})'\mathbf{Q_{decay}}\mathbf{R}
\end{aligned}
$$

where shrinkage by the Troika is exerted on $\mathbf{b}_{\text{res}} = \mathbf{Rb_f} + \mathbf{c} - \mathbf{b_0}$ instead of $\mathbf{b} = \mathbf{Rb_f} + \mathbf{c}$. Equating this expression to zero, the *customized, regularized and constrained* H_0-shrinkage solution is

obtained as

$$
\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2, \mathbf{b_0})
$$

$$
= \left\{ \Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R}\right] + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} \right\}^{-1}
$$

$$
\left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \Re\left\{ (\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{c} \right\} \right.
$$

$$
\left. -\mathbf{R}'\left(\lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{decay}} \right)(\mathbf{c} - \mathbf{b_0}) \right)
$$

$$
= \left\{ \Re\left[\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R}\right] + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R} \right\}^{-1}
$$

$$
\left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R})'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const} \right) \tag{72}
$$

where

$$
\mathbf{Const} = -\mathbf{R}'\left\{ \Re\left[\mathbf{X}_{\text{rot}}^{\text{Cust}\prime}\mathbf{X}_{\text{rot}}^{\text{Cust}}\right]\mathbf{c} + \left[\lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{decay}}\right](\mathbf{c} - \mathbf{b_0}) \right\}
$$

If $\mathbf{b_0} = \mathbf{0}$ then 72 simplifies to 71. If $\mathbf{R} = \mathbf{Id}$ and $\mathbf{c} = \mathbf{0}$ (no restrictions imposed), then 72 simplifies to 59.

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; 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

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

which generalizes the case $h = 0$ (nowcast). Note that we observed a similar weighting of the constants w^u in 62 and 63, already.

9 Filter Constraints: the Case of Cointegration

Cointegration addresses a particular property of the DGP of a set of multivariate non-stationary (integrated) time series. As we shall see, this structuring - 'shaping' - property of the DGP will lead to a set of constraints which link filters *across* series. In contrast, amplitude and time-shift constraints in the previous section were mutually unrelated across series. To be clear let us repeat that the abstract DGP is of no immediate concern or interest here: we just take the opportunity to derive and to propose a set of potentially meaningful filter constraints which might be interesting in

a particular research context whether the 'true' process is cointegrated or not: as for the 'unit roots' in the previous section the theoretical concept of 'cointegration' is assimilated into a pretext (namely to derive interesting constraints). As a matter of fact, we here continue to use the *stationary* MDFA-criteria (customized and/or possibly regularized) based on the DFT of the data. An extension of the optimization criteria to non-stationary *integrated* processes is proposed later in section 11.

9.1 Cointegration: the Rank One Case

Let's take pretext from the following structuring hypothesis: X_t and W_{it}, $i = 1, ..., m$ are integrated I(1)-series with a single unit root in frequency zero (arbitrary integration orders are discussed in McElroy and Wildi (2013)) and there exists a single non-null cointegrating vector $(a_0, a_1, ..., a_{m+1})'$ such that $a_0 X_t + \sum_{i=1}^{m} a_{i+1} W_{it}$ is stationary (the rank one case). For notational ease we now assume that $a_0 \neq 0$ i.e. the trend of the target series is related to the trends in the explaining variables through the cointegration relation[25]. This assumption does not preclude the scope of the following analysis since we can always interchange the ordering of the series such that the re-ordered a_0 is different from zero. Note that for any number w the linear combination $wa_0 X_t + w \sum_{i=1}^{m} a_{i+1} W_{it}$ must be stationary too and therefore we might select $w = 1/a_0$ in order to define a new *normalized* cointegrating vector: $\alpha_0 = 1, \alpha_i = -a_i/a_0$. Then

$$X_t - \alpha_1 W_{1t} - ... - \alpha_m W_{mt} \tag{73}$$

must be a stationary series. As in the previous sections we consider estimation of Y_{T-h}: if $h = 0$ then the *real-time* filter is obtained; for $h < 0$ a forecast results; for $h > 0$ an optimal smoother is derived. For simplicity of exposition we assume that all series are *coincident* i.e. the Lag h (for estimating Y_{T-h}) is the same for all explaining series[26]. Whereas our previous developments emphasized a strict frequency-domain approach, we now acknowledge the usefulness of the time-domain by deriving a universal decomposition of the filter error which emphasizes all relevant cointegration aspects: this decomposition will be suitably transposed into the frequency-domain a

---

[25]If X_t is an interesting macro series, say GDP, and Y_t is the signal, say the trend, then the analyst generally attempts to select series W_{it} whose trends are related to the GDP-trend in 'some way' because the trend is typically a 'salient' interesting feature.

[26]Different (integer or fractional) lead/lags are accounted for 'automatically' by the cross-products of the DFT's in the MDFA-criteria: in this general setting the time-index t always refers to the sampling period i.e. the time points at which the data are collected (this is not necessarily identical with the period to which a possibly leading or lagging series W_{it} refers). For ease of exposition (notational simplicity) we therefore assume that all series are sampled simultaneously. A sampling-lag L_j (integer or possibly fractional) of $W_j(t)$ could be accounted for very easily by multiplying $\hat{\Gamma}_{W_j}(\omega_k)$ by $\exp(-iLag_j\omega_k)$ in the MDFA-criterion: but this additional 'feature' would unnecessarily complicate all expressions.

bit later, in section 11. Let the filter error be

$$r_t \;=\; Y_t - \sum_{k=-h}^{L-h} b_k^0 X_{t-k} - \sum_{k=-h}^{L-h} b_k^1 W_{1,t-k} - \dots - \sum_{k=-h}^{L-h} b_k^m W_{m,t-k}$$

Note that r_t critically depends on h but we decline the opportunity to hamper the notation. Denote by $\hat{X}_t, \hat{W}_{it}$ the corresponding filter outputs in the time domain (which also heavily depend on h) so that

$$r_t \;=\; Y_t - \hat{X}_t - \hat{W}_{1t} - \dots - \hat{W}_{mt}$$

The error can be decomposed according to

$$
\begin{aligned}
r_t \;=\; & \sum_{k=-\infty}^{\infty} (\gamma_k - b_k^0) X_{t-k} - \sum_{k=-h}^{L-h} b_k^1 W_{1,t-k} - \dots - \sum_{k=-h}^{L-h} b_k^m W_{m,t-k} \\
\;=\; & \sum_{k=-\infty}^{\infty} (\gamma_k - b_k^0)\Big(X_{t+h} + (X_{t-k} - X_{t+h})\Big) - \sum_{k=-h}^{L-h} b_k^1 \Big(W_{1,t+h} + (W_{1,t-k} - W_{1,t+h})\Big) \\
& - \dots - \sum_{k=-h}^{L-h} b_k^m \Big(W_{m,t+h} + (W_{m,t-k} - W_{m,t+h})\Big) \\
\;=\; & (\Gamma(0) - \hat{\Gamma}_X(0)) X_{t+h} - \hat{\Gamma}_{W_1}(0) W_{1,t+h} - \dots - \hat{\Gamma}_{W_m}(0) W_{m,t+h} \qquad (74)\\
& + \sum_{k=-\infty}^{\infty} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) - \sum_{k=-h}^{L-h} b_k^1 (W_{1,t-k} - W_{1,t+h}) - \sum_{k=-h}^{L-h} b_k^m (W_{m,t-k} - W_{m,t+h}) \;(75)
\end{aligned}
$$

where $b_j^0 = 0$ if $j < -h$ or $j > L - h$. One can show that the differences in 75 are stationary under mild assumptions about the filters involved, see section 11. Therefore, stationarity (finite variance) of the filter error r_t critically and solely depends on the level-term 74:

$$(\Gamma(0) - \hat{\Gamma}_X(0)) X_{t+h} - \hat{\Gamma}_{W_1}(0) W_{1,t+h} - \dots - \hat{\Gamma}_{W_m}(0) W_{m,t+h}$$

Comparing this expression with 73 (and ignoring the shift by h because the linear combination is stationary) implies that the following filter restrictions must be satisfied in order that r_t remains stationary (finite variance)

$$\hat{\Gamma}_{W_i}(0) = \alpha_i(\Gamma(0) - \hat{\Gamma}_X(0)) \qquad (76)$$

for $i = 1, ..., m$. In particular, the restriction $\Gamma(0) = \hat{\Gamma}_X(0)$ (necessary in the univariate I(1)-case) together with $\hat{\Gamma}_{W_i}(0) = 0, i = 1, ..., m$, would ensure a finite variance of the filter error. Unfortunately, the additional 'level' information conveyed by $W_{it}, i = 1, ...m$ would be 'lost' because the restriction $\hat{\Gamma}_{W_i}(0) = 0$ implies a 'squashing' of the trend of W_{it} by $\hat{\Gamma}_{W_i}(\cdot)$. So, ultimately, cointegration amounts to *relax* filter restrictions by allowing $\hat{\Gamma}_{W_i}(0)$ to depend on $\hat{\Gamma}_X(0)$ in 76: a

cross-sectional link between filters is established in the unit-root frequency. In order to derive the precise link we rewrite the above filter restriction:

$$b_0^u = \alpha_u \left(\Gamma(0) - \sum_{k=-h}^{L-h} b_k^0 \right) - \sum_{k=-h, k \neq 0}^{L-h} b_k^u \qquad (77)$$

Let us remind that we could, in principle, impose the retriction to any b_k^u, $k = -h, ..., L - h$. Our natural preference for b_0^u in 77 is related to the decay-term of the regularization Troika, see 46: the weighting-scheme imposed by the decay-term is arranged symmetrically about b_0^u i.e. b_0^u is least affected by regularization and therefore regularization does not interfer with constraints. So, if

$$\mathbf{b} = \mathbf{R}\mathbf{b_f} + \mathbf{c}$$

where the $(L + 1 + mL))$-dimensional vector $\mathbf{b_f'}$ collects all freely determined parameters and the $((m + 1)(L + 1))$-dimensional vector $\mathbf{b}$ collects all coefficients, then 77 implies that

$$\mathbf{R} = \begin{pmatrix} \mathbf{Id} & 0 & & 0... & 0 & 0 \\ \alpha_1 \mathbf{C}_0 & \mathbf{C}_1 & 0 & ... & 0 & 0 \\ \alpha_2 \mathbf{C}_0 & 0 & \mathbf{C}_1 & ... & 0 & 0 \\ ::: & & & & & \\ \alpha_m \mathbf{C}_0 & 0 & ... & ... & 0 & \mathbf{C}_1 \end{pmatrix}, \mathbf{C}_0 = \begin{pmatrix} 0 & 0 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ -1 & -1 & -1 & ... & -1 & -1 \\ 0 & 0 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ 0 & 0 & 0 & ... & 0 & 0 \end{pmatrix} \qquad (78)$$

$$\mathbf{C}_1 = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 & 0 \\ ::: & & & & & & \\ -1 & -1 & -1 & ... & -1 & -1 & -1 \\ 0 & 0 & 0 & ... & 1 & ... & 0 \\ ::: & & & & & & \\ 0 & 0 & 0 & ... & ... & 0 & 1 \end{pmatrix} \quad \text{for} \quad u > 0 \qquad (79)$$

$$\mathbf{c'} = (0, ..., 0, 0, 0, ..., 0 \; || \; 0, ..., 0, \alpha_1 \Gamma(0), 0, ..., 0 \; || \; ... \; ||0, ..., 0, \alpha_m \Gamma(0), 0, ..., 0)$$

$$\mathbf{b_f'} = (b_{-h}^0, ..., b_{-1}^0, b_0^0, b_1^0, ..., b_{L-h}^0 \; || \; b_{-h}^1, ..., b_{-1}^1, b_1^1, ..., b_{L-h}^1 \; || \; ... \; || \; b_{-h}^m, ..., b_{-1}^m, b_1^m, ..., b_{L-h}^m)$$

The rows of -1's in $\mathbf{C}_u$, $u = 0, 1$ are in position $h + 1$ and the constants $\alpha_u \Gamma(0)$ $(u = 1, ..., m)$ in $\mathbf{c}$ are in position $L + 1 + (u - 1) * L + h + 1$. The matrix $\mathbf{R}$ is $((m + 1)(L + 1)) * (L + 1 + mL)$: the upper-left identity-block corresponds to $b_{-h}^0, ..., b_{L-h}^0$; it is $(L + 1) * (L + 1)$ because no constraint is imposed to the first filter $\hat{\Gamma}_X$ (it stays on the right-side of equation 76); all subsequent blocks, corresponding to $\mathbf{C}_1$ are of dimension $(L + 1) * L$ because $b_0^u, u > 0$ is determined by 77. Note that the lag-zero coefficient b_0^0 does appear in $\mathbf{b_f}$ (the first filter is unrestricted) whereas b_0^u are deleted

for $u > 0$.

In the above derivations we assumed throughout $\Gamma(0) > 0$ so that the signal Y_t must be non-stationary. If $\Gamma(0) = 0$ (and $\Gamma(\cdot)$ is sufficiently regular in the unit-root frequency zero) then the unit root of the input series X_t is cancelled by the symmetric target filter and the output series Y_t must be stationary (a typical example is a bandpass target filter applied to the non-stationary X_t). Imposing $\hat{\Gamma}_X(0) = \hat{\Gamma}_{W_i}(0) = 0$ would cancel all trends in the data and therefore $\hat{Y}_t$ would be stationary too i.e. the filter error $r_t = Y_t - \hat{Y}_t$ would have finite variance, as desired. However, efficiency would eventually be lost by ignoring the cointegration relation linking the series. A more flexible - general - approach is obtained by inserting $\Gamma(0) = 0$ in 76:

$$\hat{\Gamma}_{W_i}(0) = -\alpha_i \hat{\Gamma}_X(0) \tag{80}$$

Obviously, $\hat{\Gamma}_X(0) = \hat{\Gamma}_{W_i}(0) = 0$, $i = 1, ..., m$, is a solution but due to the cancelling of the unit root obtained by imposing the cointegration constraint 80 we can allow for real-time filters whose transfer functions do not have to vanish in frequency zero: an additional degree of freedom is obtained in order to promote efficiency by reducing the mean-square filter error 'further'.

Remarks

- In principle we could include *stationary* explaining data W_{it} without imposing constraints to the corresponding filters since the variance of the filter error r_t would remain finite 'unconditionally'. A typical example would be to include (bounded/stationary) business-survey data in the determination of 'band-passed' (or 'low-passed') GDP. However, we do not recommend a mix of different integration order(s) due to the different dynamics supported by the data. For sake of clarity let us specify this statement: we advise against the utilization of stationary business-survey data when tracking the business-cycle as obtained by applying a bandpass target to GDP (or IPI) in *levels*: the 'old' OECD-CLI is a typical representative. However, we can recommend business-surveys when tracking *growth* in the *differenced* business-cycle of GDP obtained either by applying the bandpass (or a lowpass) to differenced GDP or by differencing the bandpassed (or lowpassed) GDP: the EURI[27] is a typical representative.

- The above arguments concerning cancellation of unit-roots by filter constraints rely on mild regularity assumptions discussed in section 11.

---

[27]See $http://www.idp.zhaw.ch/euri$

9.2 Cointegration : General Case (Rank≥ 1)

Consider the case of multiple $(f > 1)$ linearly independent cointegration constraints

$$a_{10}X_t + a_{11}W_{1t} + ... + a_{1m}W_{mt}$$

$$a_{20}X_t + a_{21}W_{1t} + ... + a_{2m}W_{mt}$$

$$...$$

$$a_{f0}X_t + a_{f1}W_{1t} + ... + a_{fm}W_{mt}$$

Any linear combination with weights $\mathbf{w}' := (w_1, ..., w_f)$

$$\sum_{j=1}^{f} w_j a_{j0} X_t + \sum_{j=1}^{f} w_j a_{j1} W_{1t} + ... + \sum_{j=1}^{f} w_j a_{jm} W_{mt}$$

must be stationary too. As in the previous section we now assume that the trend of X_t is not independent of the other series such that a particular $a_{i0} \neq 0$. As for the rank-one case, this assumption is not strictly necessary for the derivation of our results (if it is not satisfied then the time series could be re-ordered until it is); instead it allows for a nice interpretation of the more complex non-stationary extension of the MDFA-criteria in section 11; although a sensible 'macro-indicator' (real-time) design would typically comply with this assumption, it ultimately serves a didactical purpose in our context. Assume that $i = f$ i.e. $a_{f0} \neq 0$ which could be obtained by re-ordering the cointegration vectors. Since the stationarity argument holds irrespective of the length of $\mathbf{w}$ we now impose the following normalization

$$\sum_{j=1}^{f} w_j a_{j0} = 1 \text{ or, equivalently } w_f = \frac{1 - \sum_{j=1}^{f-1} a_{j0}}{a_{f0}} \tag{81}$$

which replicates results in the previous section for $f = 1$ (cointegration-rank one) i.e. $w = 1/a_0$. Let now $\alpha_{ji} := -a_{ji}$ for $i > 0$[28] so that

$$\sum_{j=1}^{f} w_j a_{j0} X_t - \sum_{j=1}^{f} w_j \alpha_{j1} W_{1t} - ... - \sum_{j=1}^{f} w_j \alpha_{jm} W_{mt}$$

$$= X_t - \sum_{j=1}^{f} w_j \alpha_{j1} W_{1t} - ... - \sum_{j=1}^{f} w_j \alpha_{jm} W_{mt} \tag{82}$$

(where we used 81) is a stationary series. Comparing this expression with 74

$$(\Gamma(0) - \hat{\Gamma}_X(0))X_t - \hat{\Gamma}_{W_1}(0)W_{1t} - ... - \hat{\Gamma}_{W_m}(0)W_{mt}$$

implies that stationarity (finite variance) of the filter error r_t is obtained if

$$\hat{\Gamma}_{W_i}(0) = \left(\sum_{j=1}^{f} w_j \alpha_{ji} \right) (\Gamma(0) - \hat{\Gamma}_X(0)) \tag{83}$$

---

[28]This convention slightly departs from the rank-one case where we set $\alpha_i = -a_i/a_0$.

We now derive the exact multivariate parametrization of the filter constraints corresponding to these restrictions. For this purpose we first need to replace the weights w_j in 83 by corresponding filter-expressions. Using 81, the vector $\mathbf{w}$ can be expressed as

$$
\begin{pmatrix}
w_1 \\
w_2 \\
. \\
w_{f-1} \\
\frac{1-\sum_{j=1}^{f-1} a_{j0}}{a_{f0}}
\end{pmatrix}
= \mathbf{w} = \mathbf{W}\mathbf{w_{f-1}} + \mathbf{e_f}
\tag{84}
$$

where

$$
\mathbf{W} =
\begin{pmatrix}
1 & 0 & 0 & \dots & 0 \\
0 & 1 & 0 & \dots & 0 \\
& \vdots & & & \\
0 & 0 & 0 & \dots & 1 \\
-\frac{a_{10}}{a_{f0}} & -\frac{a_{20}}{a_{f0}} & -\frac{a_{30}}{a_{f0}} & \dots & -\frac{a_{f-1,0}}{a_{f0}}
\end{pmatrix}
, \mathbf{w_{f-1}} =
\begin{pmatrix}
w_1 \\
w_2 \\
. \\
w_{f-1}
\end{pmatrix}
, \mathbf{e_f} =
\begin{pmatrix}
0 \\
0 \\
. \\
0 \\
\frac{1}{a_{f0}}
\end{pmatrix}
$$

$\mathbf{W}$ is a $f * (f-1)$ matrix and $\mathbf{e_f}$ is a f-dimensional vector. We now rewrite the first $f-1$ equations in 83 in vector-form:

$$
\begin{aligned}
\hat{\mathbf{\Gamma}}_{1:f-1}(0) &= (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{1:f-1}\mathbf{w} \\
&= (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{1:f-1} (\mathbf{W}\mathbf{w}_{f-1} + \mathbf{e}_f)
\end{aligned}
\tag{85}
$$

where

$$
\hat{\mathbf{\Gamma}}_{1:f-1}(\omega) :=
\begin{pmatrix}
\hat{\Gamma}_{W_1}(\omega) \\
\hat{\Gamma}_{W_2}(\omega) \\
. \\
\hat{\Gamma}_{W_{f-1}}(\omega)
\end{pmatrix}
\text{ and } \mathbf{A}_{1:f-1} :=
\begin{pmatrix}
\alpha_{11} & \alpha_{21} & \dots & \alpha_{f1} \\
\alpha_{12} & \alpha_{22} & \dots & \alpha_{f2} \\
& \vdots & & \\
\alpha_{1,f-1} & \alpha_{2,f-1} & \dots & \alpha_{f,f-1}
\end{pmatrix}
$$

Thus we can express $\mathbf{w}_{f-1}$ as a function of $\hat{\mathbf{\Gamma}}_{1:f-1}(0)$ and $\hat{\Gamma}_X(0)$:

$$
\mathbf{w}_{f-1} = (\mathbf{A}_{1:f-1}\mathbf{W})^{-1} \left(\frac{\hat{\mathbf{\Gamma}}_{1:f-1}(0)}{\Gamma(0) - \hat{\Gamma}_X(0)} - \mathbf{A}_{1:f-1}\mathbf{e}_f \right)
\tag{86}
$$

Remarks:

- The matrix $\mathbf{A}_{1:f-1}\mathbf{W}$ must have full rank (the cointegration rank is f) and therefore it is invertible.

- We implicitly assumed $|\Gamma(0) - \hat{\Gamma}_X(0)| > 0$ in the above derivation. Otherwise 83 would 'degenerate' to $\hat{\Gamma}_{W_i}(0) = 0$ at least for all those W_{it} with a non-degenerated (non-null)

40

cointegration vector α_{ji}. In the null-case, i.e. if $\alpha_{ji} = 0$ for all j, then 83 could not be invoked, but in such a case the trend of W_{it} would be independent of all other trends, including X_t, and therefore $\hat{\Gamma}_{W_i}(0) = 0$ would be necessary in order to obtain a stationary (finite variance) filter error. Therefore, our implicit assumption $|\Gamma(0) - \hat{\Gamma}_X(0)| > 0$ covers the interesting 'non-degenerate' case.

Let now

$$
\hat{\mathbf{\Gamma}}_{f:m}(\omega) := \begin{pmatrix} \hat{\Gamma}_{W_f}(\omega) \\ \hat{\Gamma}_{W_{f+1}}(\omega) \\ . \\ \hat{\Gamma}_{W_m}(\omega) \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{f:m} := \begin{pmatrix} \alpha_{1f} & \alpha_{2f} & ... & \alpha_{ff} \\ \alpha_{1,f+1} & \alpha_{2,f+1} & ... & \alpha_{f,f+1} \\ ::: & & & \\ \alpha_{1m} & \alpha_{2m} & ... & \alpha_{fm} \end{pmatrix}
$$

Then

$$
\begin{aligned}
\hat{\mathbf{\Gamma}}_{f:m}(0) &= (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{f:m}\mathbf{w} \\
&= (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{f:m}(\mathbf{W}\mathbf{w}_{f-1} + \mathbf{e}_f) \hspace{2cm} (87) \\
&= (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{f:m}\left(\mathbf{W}\left\{(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}\left(\frac{\hat{\mathbf{\Gamma}}_{1:f-1}(0)}{\Gamma(0) - \hat{\Gamma}_X(0)} - \mathbf{A}_{1:f-1}\mathbf{e}_f\right)\right\} + \mathbf{e}_f\right) \\
&= \mathbf{A}_{f:m}\left(\mathbf{W}\left\{(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}\left(\hat{\mathbf{\Gamma}}_{1:f-1}(0) - (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{A}_{1:f-1}\mathbf{e}_f\right)\right\} + (\Gamma(0) - \hat{\Gamma}_X(0))\mathbf{e}_f\right) \hspace{0.3cm} (88)
\end{aligned}
$$

where the next to last equation is obtained by inserting 86 into 87. For an arbitrary cointegration rank f, equation 88 links the coefficients of the filters in $\hat{\mathbf{\Gamma}}_{f:m}$ to the freely determined coefficients of $\left(\hat{\Gamma}_X, \hat{\mathbf{\Gamma}}_{1:f-1}\right)'$. Equation 76 is obtained as a special case when $f = 1$ by noting that $\mathbf{w}_{f-1} = 0$, $\mathbf{e}_f = 1/a_0$ and $\mathbf{A}'_{1:m} = (-a_1, ..., -a_m)$ in 87. We can now express 83 in terms of filter coefficients:

$$
b_0^u = -\sum_{k=-h, k \neq 0}^{L-h} b_k^u + F(\hat{\Gamma}_X(0), \hat{\mathbf{\Gamma}}_{1:f-1}(0)) \hspace{2cm} (89)
$$

for all $u \geq f$ (for $u < f$ no constraint is imposed), where $F(\cdot)$ is a linear function. We now disentangle $F(\cdot)$ in order to express the constraints in the familiar form

$$
\mathbf{b} = \mathbf{R}\mathbf{b_f} + \mathbf{c} \hspace{2cm} (90)
$$

where the $(f(L+1) + (m+1-f)L)$-dimensional vector $\mathbf{b}'_\mathbf{f} = (\mathbf{b}^0, \mathbf{b}^1, \mathbf{b}^{f-1}, \mathbf{b}^f_{k\neq0}, \mathbf{b}^{f+1}_{k\neq0}, ..., \mathbf{b}^m_{k\neq0})'$ collects all freely determined parameters: the vectors $\mathbf{b}^u$ are of full length $L+1$ for $0 \leq u \leq f-1$ (no restriction is imposed to the coefficients of $\left(\hat{\Gamma}_X, \hat{\mathbf{\Gamma}}_{1:f-1}\right)'$) but $\mathbf{b}^u_{k\neq0} = (b^u_{-h}, ..., b^u_{-1}, b^u_1, ..., b^u_{L-h})'$ are of length L 'only' for $u \geq f$ since the lag-zero coefficient b_0^u is 'deleted', see 89. The $((m+1)(L+1))$-dimensional vector $\mathbf{b}$ collects all coefficients. We first extract the constant from 88:

$$
\Gamma(0)\left(\mathbf{A}_{f:m}\mathbf{e}_f - \mathbf{A}_{f:m}\mathbf{W}(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}\mathbf{A}_{1:f-1}\mathbf{e}_f\right)
$$

41

Define $\mathbf{d} := \left(\mathbf{A}_{f:m}\mathbf{e}_f - \mathbf{A}_{f:m}\mathbf{W}(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}\mathbf{A}_{1:f-1}\mathbf{e}_f\right)$. One can verify that the dimension of $\mathbf{d}$ is $m+1-f$. The component pertaining to $\hat{\Gamma}_X(0)$ in 88 is then:

$$-\hat{\Gamma}_X(0)\mathbf{d}$$

Next we extract $\hat{\mathbf{\Gamma}}_{1:f-1}(0)$:

$$\mathbf{A}_{f:m}\mathbf{W}(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}\hat{\mathbf{\Gamma}}_{1:f-1}(0)$$

One can verify that $\mathbf{A}_{f:m}\mathbf{W}(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}$ is a $(m+1-f)*(f-1)$ matrix. Define

$$\mathbf{A} := \mathbf{A}_{f:m}\mathbf{W}(\mathbf{A}_{1:f-1}\mathbf{W})^{-1}$$

$$\mathbf{C}_0 := \begin{pmatrix} 0 & 0 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ 1 & 1 & 1 & ... & 1 & 1 \\ 0 & 0 & 0 & ... & 0 & 0 \\ ::: & & & & & \\ 0 & 0 & 0 & ... & 0 & 0 \end{pmatrix}$$

$$\mathbf{C}_1 := \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 & 0 \\ ::: & & & & & & \\ -1 & -1 & -1 & ... & -1 & -1 & -1 \\ 0 & 0 & 0 & ... & 1 & ... & 0 \\ ::: & & & & & & \\ 0 & 0 & 0 & ... & ... & 0 & 1 \end{pmatrix}$$

where $\mathbf{C}_0$, $\mathbf{C}_1$ are $L*(L+1)$ matrices with the rows of 1 and -1 vectors in position $h+1$. Then

$$\mathbf{R} = \begin{pmatrix} \mathbf{Id}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Id}_2 & \mathbf{0} \\ -\mathbf{d}\otimes\mathbf{C}_0 & \mathbf{A}\otimes\mathbf{C}_0 & \mathbf{Id}_3\otimes\mathbf{C}_1 \end{pmatrix} \tag{91}$$

where $\mathbf{Id}_1$ is a $(L+1)*(L+1)$ Identity, $\mathbf{Id}_2$ is a $(L+1)(f-1)*(L+1)(f-1)$ Identity and $\mathbf{Id}_3$ is an $(m+1-f)*(m+1-f)$ identity and where the symbol $\otimes$ means the Kronecker-product[29]. The first row-block in $\mathbf{R}$ is made of an $(L+1)*(L+1)$ identity, $\mathbf{Id}_1$ and two zero-blocks of dimension $(L+1)*(L+1)(f-1)$ and $(L+1)*(m+1-f)(L+1)$; the second row-block has a first zero-block of

---

[29] $A\otimes B = \begin{pmatrix} a_{11}B & a_{12}B & ... & a_{1n}B \\ ::: & & & \\ a_{m1}B & a_{m2}B & ... & a_{mn}B \end{pmatrix}$ where A is $m*n$. If B is $k*l$ then $A\otimes B$ is $mk*nl$ dimensional.

dimension $(L+1)(f-1)*(L+1)$, a $(L+1)(f-1)*(L+1)(f-1)$ Identity $\mathbf{Id}_2$ and a second zero-block of dimension $(L+1)(f-1)*(m+1-f)(L+1)$; finally the third row block has a $\mathbf{c}\otimes\mathbf{C}_0$-matrix of dimension $(m+1-f)L*(L+1)$, a second matrix $\mathbf{A}\otimes\mathbf{C}_0$ of dimension $(m+1-f)L*(L+1)(f-1)$ and a last matrix $\mathbf{Id}_3\otimes\mathbf{C}_1$ of dimension $(m+1-f)L*(m+1-f)(L+1)$. The matrix $\mathbf{R}$ is $((m+1)(L+1))*(f(L+1)+(m+1-f)L)$. Finally, the vector $\mathbf{c}$ in 90 becomes

$$\mathbf{c} = \begin{pmatrix} \mathbf{0} \\ \mathbf{d}\otimes\mathbf{i} \end{pmatrix} \tag{92}$$

where the upper zero-vector is of length $f*(L+1)$ and $\mathbf{i}=(0,...,0,1,0,...,0)'$ has dimension $L+1$ with the 1 in position $h+1$. The Kronecker-product $\mathbf{d}\otimes\mathbf{i}$ is a vector of length $(m+1-f)(L+1)$.

We now briefly discuss the 'extreme' cases $f=0$ and $f=m+1$. For the first one there does not exist a stationary linear combination of the data in levels which means that $\alpha_i=0, i=1,...,m$ (rank zero) i.e. $\Gamma(0)-\hat{\Gamma}(0)=0, \hat{\Gamma}_{W_i}(0)=0, i=1,...,m$, as briefly discussed in the previous section 9.1. The only possibility for obtaining a stationary (finite variance) filter error r_t in this situation ($f=0$) is to track the signal Y_t by imposing a perfect fit of Γ by $\hat{\Gamma}_X$ in the unit root frequency and to squash the unrelated trends of W_{it} by imposing a zero in the unit root frequency of the corresponding transfer functions. The case $f=m+1$ means that any linear combination of the data series is stationary i.e. $\mathbf{b_f}=\mathbf{b}$: no constraint is imposed. Note that all series must be stationary in this case which would conflict with the (model-based) I(1)-assumption. In order to avoid this inconsistency we can simply ignore the case, $f=m+1$, since unconstrained estimation was treated in full length in previous sections.

Let us briefly establish a link between the cointegration constraints proposed in this section and the amplitude and time-shift constraints proposed in the previous section 8. Amplitude and cointegration constraints both emphasize 'level' performances (we did not discuss I(2)-cointegration which is likely to be irrelevant in an economic context, see McElroy and Wildi (2013) for the general case). A time-shift constraint as proposed in section 8 can be imposed independently of I(1)-cointegration constraints: we can apply any combination of both types. In contrast, amplitude and cointegration constraints cannot be mixed, in general: whereas the former emphasize completely *independent* 'series-specific' constraints, the latter emphasize *cross-sectional* links *between* filters of different series. The special case $f=0$ (cointegration rank zero i.e. unrelated trends) is an example which can be replicated by amplitude constraints, specifically: $\hat{A}_X(0)=\Gamma(0)$ (assuming a positive symmetric target filter) and $\hat{A}_{W_i}(0)=0, i=1,...,m$. But in general both sets of level-constraints are essentially different and the user has to decide on a preference, depending on a particular application as well as a given research priority.

To conclude, we would like to emphasize that the cointegration vectors have to be supplied in order to apply the above restrictions. These could be obtained by classical linear regression applied to the cross-section of a set of time series (there are more sophisticated full-maximum likelihood alternatives available) or through a DFM-model, see McElroy and Wildi (2013). In any case, we avoid mixing real-time signalextraction (by MDFA) and structural (economic) modeling: the latter must be performed prior to MDFA in order to derive the relevant cointegration equations.

10 Summary

We here list all optimization criteria in a hierarchical way, going from the simplest to the most general and complex.

10.1 Mean-Square

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

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

10.2 Customization

Let

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

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

Then

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

Accordingly, the customized coefficient estimate is obtained as

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

10.3 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. The first case is the zero-shrink estimate:

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

$$+ \quad \lambda_{\text{smooth}} \mathbf{b}' \mathbf{Q_{smooth}} \mathbf{b} + \lambda_{\text{cross}} \mathbf{b}' \mathbf{Q_{cross}} \mathbf{b} + \lambda_{\text{decay}} \mathbf{b}' \mathbf{Q_{decay}} \mathbf{b} \rightarrow \min$$

$$\hat{\mathbf{b}}^{\text{Cust-Reg}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, \mathbf{b_0} = \mathbf{0}) =$$
$$\left(\Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}})' \mathbf{X}_{\text{rot}}^{\text{Cust}} \right] + \lambda_{\text{smooth}} \mathbf{Q_{smooth}} + \lambda_{\text{cross}} \mathbf{Q_{cross}} + \lambda_{\text{decay}} \mathbf{Q_{decay}} \right)^{-1} \Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})' \mathbf{Y}_{\text{rot}}^{\text{Cust}}$$

The general H_0-case is given by

$$\hat{\mathbf{b}}_c^{\text{Cust-Reg}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, \mathbf{b_0})$$
$$= \left\{ \Re\left[(\mathbf{X}_{\text{rot}}^{\text{Cust}})' \mathbf{X}_{\text{rot}}^{\text{Cust}} \right] + \lambda_{\text{smooth}} \mathbf{Q_{smooth}} + \lambda_{\text{cross}} \mathbf{Q_{cross}} + \lambda_{\text{decay}} \mathbf{Q_{decay}} \right\}^{-1}$$
$$\left((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}))' \mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const} \right)$$

with

$$\mathbf{Const} = \left\{ \lambda_{\text{smooth}} \mathbf{Q_{smooth}} + \lambda_{\text{cross}} \mathbf{Q_{cross}} + \lambda_{\text{decay}} \mathbf{Q_{decay}} \right\} \mathbf{b_0}$$

10.4 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 68 and $\mathbf{R}$ is specified by 67.

- i1=F, i2=T (time-shift without level-constraint): then $\mathbf{C}, \mathbf{c}$ are based on 69 and $\mathbf{R}$ is specified by 67.

- i1= i2=T (both constraints imposed): then $\mathbf{C}, \mathbf{c}$ are based on 70 and $\mathbf{R}$ is specified by 67.

Under zero-shrinkage the estimate in each of these cases is:

$$\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2, \mathbf{b_0} = \mathbf{0})$$
$$= \left\{ \Re\left[\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})' \mathbf{X}_{\text{rot}}^{\text{Cust}} \mathbf{R} \right] + \lambda_{\text{smooth}} \mathbf{R}' \mathbf{Q_{smooth}} \mathbf{R} + \lambda_{\text{cross}} \mathbf{R}' \mathbf{Q_{cross}} \mathbf{R} + \lambda_{\text{decay}} \mathbf{R}' \mathbf{Q_{decay}} \mathbf{R} \right\}^{-1}$$
$$\left(\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}} \mathbf{R})' \mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const} \right)$$

where

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

The general H_0-shrinkage case is given by

$$\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2, \mathbf{b_0})$$
$$= \left\{\Re\left[\mathbf{R}'(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{R}\right] + \lambda_{\text{smooth}}\mathbf{R}'\mathbf{Q_{smooth}}\mathbf{R} + \lambda_{\text{cross}}\mathbf{R}'\mathbf{Q_{cross}}\mathbf{R} + \lambda_{\text{decay}}\mathbf{R}'\mathbf{Q_{decay}}\mathbf{R}\right\}^{-1}$$
$$\left(\left(\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}})\mathbf{R}\right)'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const}\right)$$

where

$$\mathbf{Const} = -\mathbf{R}'\left\{\Re\left[\mathbf{X}_{\text{rot}}^{\text{Cust}\prime}\mathbf{X}_{\text{rot}}^{\text{Cust}}\right]\mathbf{c} + \left[\lambda_{\text{smooth}}\mathbf{Q}_{\text{smooth}} + \lambda_{\text{cross}}\mathbf{Q}_{\text{cross}} + \lambda_{\text{decay}}\mathbf{Q}_{\text{decay}}\right](\mathbf{c} - \mathbf{b_0})\right\}$$

The sought-after parameter vector is then obtained as

$$\hat{\mathbf{b}} = \mathbf{R}\hat{\mathbf{b}}_f^{\text{Cust-Reg-Const}}(\lambda, \eta, \lambda_{\text{smooth}}, \lambda_{\text{cross}}, \lambda_{\text{decay}}, i1, i2, \mathbf{b_0}) + \mathbf{c}$$

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.

10.5   Cointegration Constraints

This setting is analogous to the previous one but we rely on the parametrizations of $\mathbf{R}, \mathbf{b_f}$ and $\mathbf{c}$ as proposed in sections 9.1 and 9.2, instead. Time-shift constraints and cointegration constraints can be mixed whereas amplitude constraints and cointegration constraints cannot.

10.6   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.

11 Non-Stationary Integrated Series

Some analysts rely on cointegration for 'convenience' rather than for formal 'integrity': they know pretty well that the explicit model-based assumptions are violated (for example because series are bounded i.e. not integrated) but they nevertheless rely on the conceptual cointegration framework in order to express a salient feature of the data, namely that (carefully selected) series 'move together'. This common movement can be expressed alternatively in terms of factors (DFM or Dynamic Factor Model) or in terms of cross-sectional constraints such as imposed by the cross-sectional term in the Regularization Troika. Therefore, cointegration is but one possible toolkit - an instrument - to express and to transpose the salient feature 'comovement' into the optimization criterion.

The material in this section starts in a universal -tautological - time-domain decomposition of the filter error whose exploration and transposition into the frequency-domain will lead to multiple potentially interesting ramifications of the multivariate signal extraction problem. From a theoretical point of view, our results allow for a formal extension of the previous 'stationary' (M)DFA-framework, including customization and regularization, to non-stationary (co)integrated processes. Therefore, the user can select either the *stationary* or a suitable *non-stationary* MDFA criterion (or both, in case of doubt...) depending on the data. As a crude rule of thumb, (my) experience suggests that the stationary MDFA works well with economic data in first differences; when working with trending data in levels (GDP, IPI, financial data) I'd strongly recommend the non-stationary MDFA, instead; unfortunately, some important economic data is 'between' levels and first differences: it is typically bounded (business-surveys, interest rates) but it typically shares some of the low-frequency dynamics (but not all...) of the fundamental data in levels, see Gyomai and Wildi (2013).

11.1 Universal Time-Domain Decompositions of the Filter Error

The following decomposition is universal in the sense that it is a 'number-identity' - a tautology - which holds for any time series irrespective of assumptions. We start our analysis in the time-domain decomposition 74 and 75 which for sake of reader's comfort are replicated here:

$$r_t = (\Gamma(0) - \hat{\Gamma}_X(0))X_{t+h} - \hat{\Gamma}_{W_1}(0)W_{1,t+h} - ... - \hat{\Gamma}_{W_m}(0)W_{m,t+h} \tag{93}$$

$$+ \sum_{k=-\infty}^{\infty}(\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) - \sum_{k=-h}^{L-h}b_k^1(W_{1,t-k} - W_{1,t+h}) - \sum_{k=-h}^{L-h}b_k^m(W_{m,t-k} - W_{m,t+h}) \tag{94}$$

Note that the split of the sums at lag h is to some extent arbitrary, though it will allow to separate causal from non-causal contributions to the filter error. As suggested in section 9, the first 'level'-term 93 will account for cointegration. It can be straightforwardly transposed into the frequency domain by relying on the DFT's of the data. However, the second expression 94 involves a non-

homogeneous difference operator $(1 - B^{k+h})$ whose span is increasing with lag k. We thus need to transform this term into an expression which can be transformed readily into the frequency-domain. Let's proceed by focusing on the first summand in 94:

$$\sum_{k=-\infty}^{\infty} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) = \sum_{k=-h}^{\infty} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h})$$

$$+ \sum_{k=-\infty}^{-h-1} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h})$$

$$= \sum_{k=-h}^{\infty} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) + \sum_{k=-\infty}^{-h-1} \gamma_k(X_{t-k} - X_{t+h}) \tag{95}$$

Recall that $b_k^0 = 0$ for $k < -h$ or $k > L - h$. Once again, the split of the bi-infinite sum around $-h$ is useful when interpreting the new non-stationary optimization criterion because the second sum in 95 is a purely non-causal contribution[30] to the filter error which does not depend on the MDFA-filter. Now

$$\sum_{k=-h}^{\infty} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) \tag{96}$$

$$= \sum_{k=-h}^{\infty} (\gamma_k - b_k^0) \sum_{j=0}^{k+h-1} (X_{t-k+j} - X_{t-k+j+1})$$

$$= \sum_{j=-h+1}^{\infty} \left[\sum_{k=j}^{\infty} (\gamma_k - b_k^0) \right] (X_{t-j} - X_{t-j+1}) \tag{97}$$

$$= \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty} (\gamma_k - b_k^0) \right] (X_{t-j} - X_{t-j+1})$$

$$- \left[\sum_{k=-h}^{\infty} (\gamma_k - b_k^0) \right] (X_{t+h} - X_{t+h+1})$$

$$= - \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty} (\gamma_k - b_k^0) \right] (X_{t+1-j} - X_{t-j}) \tag{98}$$

$$+ \left[\Gamma(0) - \hat{\Gamma}_X(0) \right]_{-h}^{+} (X_{t+1+h} - X_{t+h}) \tag{99}$$

Please note our convention according to which the summand $\sum_{j=0}^{k+h-1}(X_{t-k+j} - X_{t-k+j+1})$ on the second line is zero for $k = -h$ because the superscript is smaller than the subcript: formally this is because $X_{t-k} - X_{t+h}$ in 96 vanishes for $k = -h$. The '+' superscript with the h subscript in the transfer function of 99 specifies that we sum on all lags $k \geq -h$, as implied by 96. The above sequence of transformations has replaced the undesirable $(1 - B^{k+h})$-filter by a homogeneous first

---

[30]When estimating Y_{T-h} based on data up to T, then X_{T-h-k} is unobservable if $k < -h$. Therefore, $\sum_{k=-\infty}^{-h-1} \gamma_k(X_{t-k} - X_{t+h})$ depends on future data, exclusively ('non-causal').

difference operator $1 - B$, as desired. Proposition 2 in the appendix then implies that the weights $\gamma_j^{0+'} := \left[\sum_{k=j}^{\infty} (\gamma_k - b_k^0) \right]$, $j = -h, ,..., \infty$ in the sum 98 are the coefficients of the transfer function

$$\Delta\Gamma_{-h}^{0+'}(\omega) := \frac{\exp(ih\omega)\Delta\Gamma_{-h}^{0+}(0) - \exp(-i\omega)\Delta\Gamma_{-h}^{0+}(\omega)}{1 - \exp(-i\omega)} \tag{100}$$

where $\Delta\Gamma_{-h}^{0+}(\omega) := \sum_{k=-h}^{\infty} (\gamma_k - b_k^0) \exp(-ik\omega)$ involves all coefficients with lag $k \geq -h$. We are thus able to transpose 96 formally into the frequency-domain, see below for details. We now proceed to the second term in 95 ($k < -h$) and obtain analogously:

$$\sum_{k=-\infty}^{-h-1} (\gamma_k - b_k^0)(X_{t-k} - X_{t+h}) \tag{101}$$

$$= \sum_{k=-\infty}^{-h-1} \gamma_k(X_{t-k} - X_{t+h}) \tag{102}$$

$$= \sum_{k=-\infty}^{-h-1} \gamma_k \sum_{j=k+h+1}^{0} (X_{t-k+j} - X_{t-k+j-1}) \tag{}$$

$$= \sum_{j=-\infty}^{-h-1} \left[\sum_{k=-\infty}^{j} \gamma_k \right] (X_{t-j} - X_{t-j-1}) \tag{103}$$

Note, again, that $b_k^0 = 0$ for $k < -h$. Proposition 2 shows that the coefficients $\gamma_j^{0-'} := \sum_{k=-\infty}^{j} \gamma_k$, $j = -\infty, ..., -h-1$ in 103 belong to a filter with transfer function

$$\Gamma_{-h}^{0-'}(\omega) = \Delta\Gamma_{-h}^{0-'}(\omega) := \frac{\exp(-i(-h-1)\omega)\Delta\Gamma_{-h}^{0-}(0) - \exp(i\omega)\Delta\Gamma_{-h}^{0-}(\omega)}{1 - \exp(i\omega)} \tag{104}$$

where $\Gamma_{-h}^{0-'}(\omega) = \Delta\Gamma_{-h}^{0-}(\omega) := \sum_{k=-\infty}^{-h-1} \gamma_k \exp(-ik\omega)$[31]. Note that the signs in the complex exponentials of the transfer function are flipped because we address the non-causal component, see proposition 2 for details. For the remaining terms in 94 we obtain analogously

$$\sum_{k=-h}^{L-h} b_k^n(W_{n,t-k} - W_{n,t+h}) = \sum_{k=-h}^{\infty} b_k^n(W_{n,t-k} - W_{n,t+h})$$

$$= \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty} b_k^n \right] (W_{n,t+1-j} - W_{n,t-j})$$

$$- \hat{\Gamma}_{W_n}(0)(W_{n,t+1+h} - W_{n,t+h}) \tag{105}$$

where $\gamma_j^{n'} := \sum_{k=j}^{\infty} b_k^n$ are the coefficients of the transfer function

$$\hat{\Gamma}'_{W_n}(\omega) := \frac{\exp(ih\omega)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega)\hat{\Gamma}_{W_n}(\omega)}{1 - \exp(-i\omega)} \tag{106}$$

---

[31]For sake of clarity we note that $\Delta\Gamma_{-h}^{0-}(\omega) := \sum_{k=-\infty}^{-h-1} (\gamma_k - b_k^0) \exp(-ik\omega) = \sum_{k=-\infty}^{-h-1} \gamma_k \exp(-ik\omega)$. Therefore $\Gamma_{-h}^{0-}(\omega) = \Delta\Gamma_{-h}^{0-}(\omega)$.

see proposition 2. Note that we allowed for infinite sums which are tackled by imposing $b_k^n = 0$ for $k \geq L - h$. Let

$$C'_{t+h} := (\Gamma(0) - \hat{\Gamma}_X(0))X_{t+h} - \hat{\Gamma}_{W_1}(0)W_{1,t+h} - \ldots - \hat{\Gamma}_{W_m}(0)W_{m,t+h} \tag{107}$$

designate the 'level'-term. We now insert 98, 99, 103, 105 and 107 into 93 and 94 to obtain:

$$
\begin{aligned}
r_t \;=\; & C'_{t+h} \\
& - \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty}(\gamma_k - b_k^0) \right] (X_{t+1-j} - X_{t-j}) \\
& + \Delta\Gamma_{-h}^{0+}(0)(X_{t+1+h} - X_{t+h}) \\
& + \sum_{j=-\infty}^{-h-1} \left[\sum_{k=-\infty}^{j} \gamma_k \right] (X_{t-j} - X_{t-j-1}) \\
& + \sum_{n=1}^{m} \left\{ \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty} b_k^i \right] (W_{n,t+1-j} - W_{n,t-j}) \right. \\
& \qquad\qquad \left. - \hat{\Gamma}_{W_n}(0)(W_{n,t+1+h} - W_{n,t+h}) \right\}
\end{aligned}
\tag{108}
$$

where $\Delta\Gamma_{-h}^{0+}(0) = \left[\Gamma(0) - \hat{\Gamma}_X(0) \right]_{-h}^{+}$. The distinguishing properties of this decomposition are:

- Universality: the decomposition relies on a 'number'- identity which holds for *all* time series irrespective of 'assumptions'.

- Interpretability: the decomposition can be split into two blocks, a level-block C'_{t+h} and a (first-) difference-block. The latter can be split additionally into causal and non-causal (anticipative) components too. This categorization will allow for easier interpretations of the (more complex) non-stationary MDFA optimization criteria, see below.

- Homogeneity: the expression for r_t involves homogeneous filters, only, and thus it can be transposed into the frequency-domain.

Consider now the case $f = 0$ (rank-zero). Then the cointegration space is a Null-space, $\alpha_i = 0, i = 1, \ldots, m$ (rank zero), and therefore $\Delta\Gamma(0) = \Gamma(0) - \hat{\Gamma}(0) = 0, \hat{\Gamma}_{W_i}(0) = 0, i = 1, \ldots, m$, as discussed in section 9.2: imposing these filter restrictions is necessary in order to ensure stationarity (fixed finite variance) of the filter error r_t. Moreover, these restrictions lead to a vanishing cointegration-term, $C'_t = 0$ in 107. Therefore, the above time-domain decomposition of the filter error r_t 'degenerates'

to

$$r_t = -\sum_{j=-h}^{\infty}\left[\sum_{k=j}^{\infty}(\gamma_k - b_k^0)\right](X_{t+1-j} - X_{t-j}) \tag{109}$$
$$+\Delta\Gamma_{-h}^{0+}(0)(X_{t+1+h} - X_{t+h})$$
$$+\sum_{j=-\infty}^{-h-1}\left[\sum_{k=-\infty}^{j}\gamma_k\right](X_{t-j} - X_{t-j-1})$$
$$+\sum_{n=1}^{m}\left\{\sum_{j=-h}^{\infty}\left[\sum_{k=j}^{\infty}b_k^n\right](W_{n,t+1-j} - W_{n,t-j})\right.$$

Note that $\Delta\Gamma(0) = 0$ does not imply $\Delta\Gamma_{-h}^{0+}(0) = 0$ and therefore the term on the second line is generally non-vanishing. The third term $\sum_{j=-\infty}^{-h-1}\left[\sum_{k=-\infty}^{j}\gamma_k\right](X_{t-j} - X_{t-j-1})$ collects the non-causal contribution to the filter error: it does not depend on the multivariate MDFA-filter.

11.2 Frequency-Domain Decomposition of the Filter Error

We here aim at a decomposition of the DFT of the filter error r_t, denoted by $\Xi_{Tr}(\omega_k)$, according to the time-domain decomposition 108: this will enable to express the cointegration structure in the frequency-domain by means of the DFT of C_t', see section 11.3. Note, however, that r_t and therefore $\Xi_{Tr}(\omega_k)$ are not observed because of the non-causal error-components. We thus seek an observable approximation, denoted by $\Xi_{Tr}'(\omega_k)$, which inherits the advantageous features of the decomposition 108. As we shall see, such a statistic exists and, (very) interestingly, the approximation error is such, that the resulting optimization criterion is a (very good) *superconsistent* estimate of the 'true' unobserved criterion, see section 11.3. So let's dive into the topic.

Since the DFT is linear, we can apply it to each summand on the right of 108. For the first term, C_{t+h}', we obtain

$$\exp(ih\omega_k)\Xi_{TC'}(\omega)$$

where $\Xi_{TC'}(\omega) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}C_t'\exp(-it\omega)$ is the DFT of C_t'. The exponential term $\exp(ih\omega)$ accounts for the time-shift by h time-units of C_{t+h}' in 108[32]. For the second term

$$-\sum_{j=-h}^{\infty}\left[\sum_{k=j}^{\infty}(\gamma_k - b_k^0)\right](X_{t+1-j} - X_{t-j}) \tag{110}$$

---

[32]Indeed $\frac{1}{\sqrt{2\pi T}}\sum_{t=-h+1}^{T-h}C_{t+h}'\exp(-it\omega) = \exp(ih\omega)\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}C_t'\exp(-it\omega)$. Strictly speaking, if $h \neq 0$ then the DFT of C_{t+h}' would assume observations outside the sample $t = 1, ..., T$. Therefore, we can approximate the corresponding DFT by $\exp(ih\omega_k)\Xi_{TC'}(\omega_k)$: the resulting error will be negligible by proposition 5. Alternatively, we could have split 93 and 94 about lag 0 (instead of h) such that this problem would not appear at all. However, all expressions would have become more cumbersome (than they already are) and therefore we preferred the presentation 'as is'. In summary, this discrepancy can be neglected i.e. $\exp(ih\omega_k)\Xi_{TC'}(\omega)$ is the correct expression.

we use 100 to obtain

$$-\frac{\exp(ih\omega)\Delta\Gamma_{-h}^{0+}(0) - \exp(-i\omega)\Delta\Gamma_{-h}^{0+}(\omega)}{1 - \exp(-i\omega)}\exp(i\omega)\Xi_{T\Delta X}(\omega) \tag{111}$$

where $\Xi_{T\Delta X}(\omega) = \frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}(X_t - X_{t-1})\exp(-it\omega)$ is the DFT of the first differences ΔX_t of X_t[33] and where the additional exponential term $\exp(i\omega)$ of $\exp(i\omega)\Xi_{T\Delta X}(\omega)$ accounts for the fact that the (first) differences in 110 are shifted forward by one time unit i.e. we have $(X_{t+1-j} - X_{t-j})$ instead of $(X_{t-j} - X_{t-j-1})$. For the third term

$$\Delta\Gamma_{-h}^{0+}(0)(X_{t+1+h} - X_{t+h})$$

we obtain

$$\Delta\Gamma_{-h}^{0+}(0)\exp(i(h+1)\omega)\Xi_{T\Delta X}(\omega)$$

where $\exp(i(h+1)\omega)$ accounts for the time-shift by $h+1$ time units of $(X_{t+1+h} - X_{t+h})$. The fourth term

$$\sum_{j=-\infty}^{-h-1}\left[\sum_{k=-\infty}^{j}\gamma_k\right](X_{t-j} - X_{t-j-1})$$

becomes

$$\left[\frac{\exp(-i(-h-1)\omega)\Gamma_{-h}^{0-}(0) - \exp(i\omega)\Gamma_{-h}^{0-}(\omega)}{1 - \exp(i\omega)}\right]\Xi_{T\Delta X}(\omega)$$

see 104 (recall that $\Gamma_{-h}^{0-}(\omega) = \Delta\Gamma_{-h}^{0-}(\omega)$). Since no time-shift is applied to the differences $(X_{t-j} - X_{t-j-1})$ we don't need an exponential phase-term here. Finally, the DFT's of the fifth and sixth terms in the decomposition 108 are obtained as

$$\sum_{n=1}^{m}\frac{\exp(ih\omega)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega)\hat{\Gamma}_{W_n}(\omega)}{1 - \exp(-i\omega)}\exp(i\omega)\Xi_{T\Delta W_n}(\omega)$$

(see 106) and

$$-\sum_{n=1}^{m}\hat{\Gamma}_{W_n}(0)\exp(i(h+1)\omega)\Xi_{T\Delta W_n}(\omega)$$

where the exponential phase-terms on the left of the DFT's pick-up the relevant time-shifts in the (first) differences. Summarizing, rearranging terms and setting $\omega = \omega_k$, we obtain the following

---

[33]We here neglect the fact that X_0 is not available: we can either run all DFT's on the sample $t = 2, ..., T$ and retain X_1 for intialization or we may plug any reasonable estimate $\hat{X}_0$ of X_0 into the DFT (say $\hat{X}_0 = X_1$ for example).

decomposition of the filter error r_t in the frequency-domain:

$$\begin{aligned}
\Xi'_{Tr}(\omega_k) \quad := \quad & \exp(ih\omega_k)\Xi_{TC'}(\omega_k) \\
& - \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0) - \exp(-i\omega_k)\Delta\Gamma^{0+}_{-h}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0)\right]\Xi_{T\Delta X}(\omega_k) \\
& + \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma^{0-}_{-h}(0) - \exp(i\omega_k)\Gamma^{0-}_{-h}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k) \\
& + \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\right]\Xi_{T\Delta W_n}(\omega_k)
\end{aligned} \tag{112}$$

where the third term collects the non-causal contribution to the filter error. This statistic is endowed with remarkable properties:

1. Unlike the 'true' DFT $\Xi_{Tr}(\omega_k)$ of r_t, the new expression $\Xi'_{Tr}(\omega_k)$ is *observable*.

2. Therefore, obviously, $\Xi'_{Tr}(\omega_k) \neq \Xi_{Tr}(\omega_k)$ but one can show that the approximation error $\Xi'_{Tr}(\omega_k) - \Xi_{Tr}(\omega_k)$ is particular in the sense that the resulting criterion, based on $\Xi'_{Tr}(\omega_k)$, is a superconsistent estimate of the criterion based on the unobserved $\Xi_{Tr}(\omega_k)$, see section 11.3.

3. The decomposition $\Xi'_{Tr}(\omega_k)$ of the filter-error in the frequency-domain isolates the interesting level-term C'_t from the stationary differenced data and thus we'll be able to express cointegration relations by imposing stationarity of C'_t.

In the rank-zero case, $f = 0$, we rely on the 'degenerate' time-domain decomposition 109 to obtain

$$\begin{aligned}
\Xi'_{Tr}(\omega_k) \quad := \quad & -\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0) - \exp(-i\omega_k)\Delta\Gamma^{0+}_{-h}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0)\right]\Xi_{T\Delta X}(\omega_k) \\
& + \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma^{0-}_{-h}(0) - \exp(i\omega_k)\Gamma^{0-}_{-h}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k) \\
& - \sum_{n=1}^{m}\frac{\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)}\Xi_{T\Delta W_n}(\omega_k)
\end{aligned} \tag{113}$$

where we inserted $C'_t = 0$ and $\hat{\Gamma}_{W_n}(0) = 0$ in 112 and where the expression for the last term, corresponding to W_{nt}, $n = 1, ..., m$, was simplified accordingly. Note that 113 is a special case of the general expression 112 when $f = 0$: we therefore do not distinguish both expressions i.e. we use the same symbol $\Xi'_{Tr}(\omega_k)$ to address the (approximated) DFT of the filter error. The second term in the above expression collects the non-causal contribution to the filter error: it does not depend on the MDFA-filter. A distinguishing feature of 112 (and thus of 113) is the singularity of the difference-terms in the unit-root frequency zero: as we shall see, the cointegration filter-constraints proposed in section 9 cancel the unit-root effect, see section 11.4.

Before entering resolutely into the domain of non-stationarity, let us briefly compare $\Xi'_{Tr}(\omega_k)$ in 112 with the previous expression for the DFT of the filter-error as obtained in the stationary case 26

$$\Xi''_{Tr}(\omega_k) = \Gamma(\omega_k)\Xi_{TX}(\omega_k) - \hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) - \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k) \qquad (114)$$

This expression was rewritten in vector notation as $\mathbf{Y} - \mathbf{Xb}$ in 33 (a similar transcription in the I(1)-case is proposed in section 11.5 below). To begin with, we note that $\Xi''_{Tr}(\omega_k) \neq \Xi'_{Tr}(\omega_k)$ because both DFT's rely on different decompositions which lead to different convolutions and, accordingly, to different approximation errors. The question is: which frequency-domain representation of the filter error is 'better': $\Xi''_{Tr}(\omega_k)$ or $\Xi'_{Tr}(\omega_k)$? Exact finite-sample overdifferencing results in Wildi (2005), chapter 4, suggest that the latter is *less* well-suited in the case of stationary processes when no filter restrictions are imposed, (figs 4.5 and 4.6 in the cited literature illustrate undesirable issues). If cointegration restrictions are imposed, both statistics are almost equivalent in terms of signal extraction performances. But of course, imposing unnecessary restrictions conflicts with efficiency. Therefore, we recommend $\Xi''_{Tr}(\omega_k)$ in 26 when working with stationary data, such as, for example, differenced economic data.

11.3 I(1)-Mean-Square Criterion

We here list a set of assumptions which allow to distinguish our new optimization criterion as a superconsistent estimate of the 'true' unobserved criterion. Specifically, we assume that all series are non-stationary I(1) with a single unit-root in frequency zero. A general framework addressing unit-roots of any order and at any frequencies is proposed in McElroy and Wildi (2013): here we typically address the case of an analyst working with fundamental and/or financial economic data in levels (macro-indicators, trading rules) so that the above I(1)-assumption is 'sufficient'. Furthermore we assume that the filter constraints $\mathbf{b} = \mathbf{Rb_f} + \mathbf{c}$ as defined in section 9.2 are imposed. We assume, also, that the (stationary) differenced series admit absolutely summable wold-decompositions; as an example, this assumption is met by stationary ARMA-processes whose (MA-coefficients of the) Wold-decompositions converge at an exponential rate towards zero. Finally, we require the co-efficients γ_k of the bi-infinite filter $\Gamma(\cdot)$ to converge sufficiently rapidly towards zero in the sense that $\sum_{k=-\infty}^{\infty} |\gamma_k||k|^{1.5} < \infty$ (formally, we require $\Gamma(\cdot) \in C^{1.5}$, see propositions 2, 3 and 5 in the appendix[34]). This last restriction is trivially satisfied if $\Gamma(\cdot)$ is of *finite* length which can invariably be assumed in typical economic applications[35].

---

[34]Proposition 2 ensures that the coefficients of the filters for the data in first difference are in $C_f^{1.5-1} = C_f^{0.5}$. Therefore, proposition 5 ensures super-consistency of the resulting criterion.

[35]It is generally admitted that data in the remote past (say one recession back in time) does not contribute significantly to a better understanding of current activity. Therefore the length of the symmetric filter can be truncated to ± 10 years, say.

Let's now proceed step-by-step: the discrete spectral factorization theorem warrants:

$$\frac{1}{T}\sum_{t=1}^{T} r_t^2 \;=\; \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k \left|\Xi_{Tr}(\omega_k)\right|^2 \tag{115}$$

where the weights w_k (not to be confounded with ω_k) are defined by 12 and where $\left|\Xi_{Tr}(\omega_k)\right|^2$ is the periodogram of the filter error r_t. The above equality is a universal number-identity - a tautology - which applies to all data, irrespective of assumptions, see for example proposition 10.4 in Wildi (2008). Our intention here will be to minimize this expression as a function of filter coefficients $\mathbf{b_f}$ i.e. we want to minimize the *empirical* MSE. Besides its intuitive appeal, one can show that the empirical MSE is an asymptotically efficient estimate of the true (unobservable) mean-square error $E[r_t^2]$ under additional 'mild' assumptions on the DGP, see proposition 6 in the appendix. In this sense, we could invoke statistical *efficiency* for the resulting MDFA-criterion. However, we neglect this technical background here and focus on the empirical MSE 115 instead. Since $\Xi_{Tr}(\omega_k)$ is not observable directly, we consider the following natural approximation

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k \left|\Xi_{Tr}(\omega_k)\right|^2 \approx \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} \left|\Xi'_{Tr}(\omega_k)\right|^2 \tag{116}$$

where $\Xi'_{Tr}(\omega_k)$ is defined by 112 or by the 'degenerate' rank-zero form 113. For notational simplicity we ignored the weights w_k in the second summand because their contribution to the approximation error is negligible (order $\mathrm{O}(1/T)$) . Under the above assumptions, all series entering 112 or 113 (C'_t as well as the data in first differences) are stationary. Then, proposition 5 can be invoked to distinguish $\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} \left|\Xi'_{Tr}(\omega_k)\right|^2$ on the right-hand side of 116 as a (very) good *superconsistent* estimate of the (unobservable) sample mean-square criterion $\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} w_k \left|\Xi_{Tr}(\omega_k)\right|^2 = \frac{1}{T}\sum_{t=1}^{T} r_t^2$ on the left-hand side. Given the unconventional tightness of the approximation 116 we propose the following optimization criterion

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} \left|\Xi'_{Tr}(\omega_k)\right|^2 \rightarrow \min_{\mathbf{b}=\mathbf{Rb_f}+\mathbf{c},\hat{\mathbf{\Gamma}}\in C_f^{1.5}} \tag{117}$$

where $\Xi'_{Tr}(\omega_k)$ is defined by 112 (113) and where the restrictions 89, $\mathbf{b} = \mathbf{Rb_f} + \mathbf{c}$, are imposed. Note that the regularity requirement ,$\hat{\mathbf{\Gamma}} \in C_f^{1.5}$ means that all sub-filters of the multivariate solution must be in $C_f^{1.5}$ (see definition 1 in the appendix) i.e. their coefficients must decay sufficiently rapidly towards zero. This assumption is automatically satisfied for *finite* filters. But if we allow the filter length $L+1$ to increases with the sample size T (which will invariably be the case in practice) then the finiteness assumption is potentially violated. Fortunately, we can link this (somehow abstract regularity) requirement in a symbiotic way with the Regularization Troika and, more specifically, with the decay-term in section 7.2.2. Therefore, from a practical point of view,

the Regularization Troika provides a very natural operationalization of the regularity-requirement in criterion 117. Note also, that the singularity in the unit-root frequency $\omega_k = 0$ can be tackled exactly, see the next section 11.4.

Criterion 117 is a natural extension of the mean-square MDFA criterion 25 to the case of (co)integrated data. As we did in the previous section, we might ask which criterion 'works' better in this case: 25 or 117? The desirable superconsistency-argument in proposition 5 depends critically on the fact that the coefficients of the Wold-decomposition of the time series are absolutely summable. Otherwise, the errors of the discrete finite sample convolutions would not be negligible anymore when compared to the natural sampling error (precise results are derived in Wildi (2005) and (2008)). The distinguishing feature of the decomposition 112, and therefore of criterion 117, lies in its ability to comply with the requirements necessary to ensure superconsistency under the postulated (co)integration framework. Therefore, we advice in favor of 117 and against 25 when working with series whose levels appear to be non-stationary.

11.4 Unveiling the Unit-root Singularity

Components in the unit-root frequency $\omega_k = 0$ potentially contribute the most to the filter error r_t. Unfortunately, effects are masked by a singularity. Therefore, in order to get a better understanding of the optimization criterion 117 we here rely on proposition 4 in order to analyze the singular unit-root frequency. Since the DFT 112 and thus the resulting criterion 117 have a rich structure we here simplify the discussion by focusing on the relevant real-time filter $h = 0$ (though the proposition applies to arbitrary h). Also we assume that $f = 1$. According to section 9.1 we can now define a new cointegration term C_t

$$
\begin{aligned}
C_t &= \frac{C_t'}{\Delta\Gamma(0)} \\
&= X_{t+h} - \frac{\hat{\Gamma}_{W_1}(0)}{\Delta\Gamma(0)}W_{1,t+h} - ... - \frac{\hat{\Gamma}_{W_m}(0)}{\Delta\Gamma(0)}W_{m,t+h} \\
&= X_{t+h} - \alpha_1 W_{1,t+h} - ... - \alpha_m W_{m,t+h}
\end{aligned}
$$

The new cointegration term C_t does not depend on filters anymore: it is a property of the DGP, as specified by the cointegration vector $(1, -\alpha_1, ..., -\alpha_m)$. We then rewrite 112 to obtain

$$
\begin{aligned}
\Xi_{Tr}'(\omega_k) \quad := \quad & \Delta\Gamma(0)\Xi_{TC}(\omega_k) && (118) \\
& - \exp(i\omega_k)\left[\frac{\Delta\Gamma_0^{0+}(0) - \exp(-i\omega_k)\Delta\Gamma_0^{0+}(\omega_k)}{1 - \exp(-i\omega_k)} - \Delta\Gamma_0^{0+}(0)\right]\Xi_{T\Delta X}(\omega_k) \\
& + \left[\frac{\exp(i\omega_k)\Gamma_0^{0-}(0) - \exp(i\omega_k)\Gamma_0^{0-}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k) \\
& + \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \hat{\Gamma}_{W_n}(0)\right]\Xi_{T\Delta W_n}(\omega_k)
\end{aligned}
$$

where we inserted $h = 0$ and $\Delta\Gamma(0)\Xi_{TC}(\omega_k) = \Xi_{TC'}(\omega_k)$ (using linearity of the DFT). We deduce that $\Delta\Gamma(0)$ modulates the contribution of the cointegration equilibrium C_t to the filter error r_t; interestingly, the effect depends solely on the value of $\Delta\Gamma(\cdot)$ in the unit-root frequency $\omega_0 = 0$: no smoothing/filtering is operated on C_t. In frequency zero the above DFT becomes

$$\begin{aligned}
\Xi'_{Tr}(0) & := & \Delta\Gamma(0)\Xi_{TC}(0) + \left(\sum_{k=-h}^{L-h} b_k^0 k\right)\Xi_{T\Delta X}(0) + \sum_{n=1}^{m}\left(\sum_{k=-h}^{L-h} b_k^n k\right)\Xi_{T\Delta W_n}(0) \\
& = & \Delta\Gamma(0)\Xi_{TC}(0) + \hat{A}_X(0)\hat{\phi}_X(0)\Xi_{T\Delta X}(0) + \sum_{n=1}^{m}\hat{A}_{W_n}(0)\hat{\phi}_{W_n}(0)\Xi_{T\Delta W_n}(0)
\end{aligned}$$

if the filter restrictions 76 or, equivalently, 77 are applied, see proposition 4. We infer that the exact contribution by the singularities corresponding to the differenced data are modulated by the product of amplitude and time-shift functions. Thus, whenever one (or both) of these vanishes, then the corresponding contribution vanishes. We can recognize the time-shift constraint proposed in section 8: imposing a vanishing time-shift corresponds to a second order zero in the unit-root frequency which would allow, in principle, to work with I(2)-data. Note the time-shifts can be varied freely whereas the amplitude functions are cross-sectionally linked through the cointegration-constraint: thus 'optimizing' the amplitude for *one* of the components will affect MSE-contributions by the *others*. Also, modifying the time-shifts of the filters in frequency zero will invariably affect filter performances at all other frequencies $\omega_k > 0$: no free-lunch.

In the rank-zero case ($f = 0$) we know that $\Delta\Gamma(0) = \hat{\Gamma}_{W_n}(0) = 0$ are necessary (and sufficient) in order to impose stationarity (finite fixed variance) of the filter error r_t. The above results then imply

$$\begin{aligned}
\Xi'_{Tr}(0) & := & \Delta\Gamma(0)\Xi_{TC}(0) + \hat{A}_X(0)\hat{\phi}_X(0)\Xi_{T\Delta X}(0) + \sum_{n=1}^{m}\hat{A}_{W_n}(0)\hat{\phi}_{W_n}(0)\Xi_{T\Delta W_n}(0) \\
& = & \Gamma(0)\hat{\phi}_X(0)\Xi_{T\Delta X}(0)
\end{aligned}$$

where we inserted $\hat{A}_X(0) = \hat{\Gamma}(0) = \Gamma(0)$ (first order constraint[36]). In the unit-root frequency $\omega_k = 0$, all contributions by W_{nt} $n = 1, ..., m$ vanish. The filter error is entirely determined by $\hat{\Gamma}_X(\cdot)$ or, more precisely, by $\hat{\phi}_X(0)$ (since $\hat{A}_X(0)$ is already determined by the constraint). Imposing a vanishing time-shift to $\hat{\Gamma}_X(\cdot)$ is sufficient to cancel the MSE-contribution in the unit-root frequency in the rank-zero case: a hypotehtical I(2)-constraint would thus involve $\hat{\Gamma}_X(\cdot)$ alone.

The richness of the DFT 112 and of the resulting criterion 117 could be illustrated further by analyzing the DFT when $h \neq 0$ and/or when $f > 1$ but we leave this as an 'exercise' and proceed a step further towards customization.

---

[36]We here assume $\Gamma(0) \geq 0$ which is straightforward in most (known to me) applications.

11.5 Matrix Notation (Frequency Domain)

In section 5 we introduced a convenient matrix notation of the (stationary) optimization criterion 25 which allowed to introduce customization, regularization and filter constraints in a tractable way. Accordingly, we here derive the target vector $\mathbf{Y}$, of length $T/2+1$, and the data-matrix $\mathbf{X}$, of dimension $(T/2+1)*(L+1)(m+1)$, in the frequency-domain, corresponding to the (non-stationary) criterion 117:

$$\frac{2\pi}{T}(\mathbf{Y} - \mathbf{X}(\mathbf{Rb_f} + \mathbf{c}))'(\mathbf{Y} - \mathbf{X}(\mathbf{Rb_f} + \mathbf{c})) = \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]} \left|\Xi'_{Tr}(\omega_k)\right|^2$$

where $\mathbf{b} = \mathbf{Rb_f} + \mathbf{c}$ satisfies the cointegration constraints in section 9. For reader's comfort lets paste the richly-structured DFT $\Xi'_{Tr}(\omega_k)$ defined in 112:

$$
\begin{aligned}
\Xi'_{Tr}(\omega_k) \quad := \quad & \exp(ih\omega_k)\left(\Delta\Gamma(0)\Xi_{TX}(\omega_k) - \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(0)\Xi_{TW_n}(\omega_k)\right) \qquad\qquad (119)\\
& - \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0) - \exp(-i\omega_k)\Delta\Gamma^{0+}_{-h}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\Delta\Gamma^{0+}_{-h}(0)\right]\Xi_{T\Delta X}(\omega_k)\\
& + \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma^{0-}_{-h}(0) - \exp(i\omega_k)\Gamma^{0-}_{-h}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k)\\
& + \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\right]\Xi_{T\Delta W_n}(\omega_k)
\end{aligned}
$$

where we exploded $\Xi_{TC'}(\omega_k)$, in the first line, into $\Delta\Gamma(0)\Xi_{TX}(\omega_k) - \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(0)\Xi_{TW_n}(\omega_k)$, see 93. The target $\mathbf{Y}$ can now be extracted by focusing on

$$\sum_{k=-\infty}^{\infty} \gamma_k X_{t-k} \quad = \quad \Gamma(0)X_{t+h} + \sum_{k=-\infty}^{\infty} \gamma_k(X_{t-k} - X_{t+h})$$

in 93. Grouping the corresponding components in $\Xi'_{Tr}(\omega_k)$ under the common $\mathbf{Y}$-banner leads to:

$$
\begin{aligned}
y_k \quad := \quad & \exp(ih\omega_k)\Gamma(0)\Xi_{TX}(\omega_k) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (120)\\
& - \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Gamma^{+}_{-h}(0) - \exp(-i\omega_k)\Gamma^{+}_{-h}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\Gamma^{+}_{0}(0)\right]\Xi_{T\Delta X}(\omega_k)\\
& + \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma^{-}_{-h}(0) - \exp(i\omega_k)\Gamma^{-}_{-h}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k)
\end{aligned}
$$

where y_k are the 'coordinates' of $\mathbf{Y} = (y_0, y_1, ..., y_{T/2})'$ and where $\Delta\Gamma(\omega_k) = \Gamma(\omega_k) - \hat{\Gamma}(\omega_k)$ in $\Xi'_{Tr}(\omega_k)$ has been replaced by $\Gamma(\omega_k) = \Gamma^{+}_{-h}(\omega_k) + \Gamma^{-}_{-h}(\omega_k)$[37].

---

[37]Note that $\tilde{y}_k := \Gamma(\omega_k)\Xi_{TX}(\omega_k)$, as used in section 5.1 (stationary processes), would be misleading here because the convolution $\Gamma(\omega_k)\Xi_{TX}(\omega_k)$ cannot claim for super-consistency of the criterion i.e. the approximation of the

In order to derive $\mathbf{X}$ we now define a new DFT corresponding to the output $\hat{Y}_t$ of the multivariate filter: in fact this is $y_k - \Xi'_{Tr}(\omega_k)$:

$$
\begin{aligned}
\Xi'_{\hat{Y}}(\omega_k) \; := \; & \exp(ih\omega_k)\left(\hat{\Gamma}_X(0)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(0)\Xi_{TW_n}(\omega_k)\right) \\
& - \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_X(0) - \exp(-i\omega_k)\hat{\Gamma}_X(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_X(0)\right]\Xi_{T\Delta X}(\omega_k) \\
& - \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\right]\Xi_{T\Delta W_n}(\omega_k)
\end{aligned}
\tag{121}
$$

Here, $\Xi'_{\hat{Y}}(\omega_k)$ are the coordinates of the vector $\hat{\mathbf{Y}}$. Some care is needed when 'extracting' $\mathbf{X}$ from $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$. Specifically we decompose $\mathbf{X}$ into three terms

$$
\mathbf{X} = \boldsymbol{\Sigma}_{\mathbf{XW}} + \boldsymbol{\Delta}_{\mathbf{XW}}(\mathbf{1}) + \boldsymbol{\Delta}_{\mathbf{XW}}(\mathbf{2})
\tag{122}
$$

The cointegration term (first line in 121) leads to $\boldsymbol{\Sigma}_{\mathbf{XW}}$ whose k-th row, $k = 0, ..., T/2$, is defined by:

$$
\boldsymbol{\Sigma}_{\mathbf{XW},k} = (1 + I_{k>0})\mathrm{Vec}_{\mathrm{row}}\begin{pmatrix}
\exp(ih\omega_k)\Xi_{TX}(\omega_k) & \exp(ih\omega_k)\Xi_{TX}(\omega_k) & ... & \exp(ih\omega_k)\Xi_{TX}(\omega_k) \\
\exp(ih\omega_k)\Xi_{TW_1}(\omega_k) & \exp(ih\omega_k)\Xi_{TW_1}(\omega_k) & ... & \exp(ih\omega_k)\Xi_{TW_1}(\omega_k) \\
\exp(ih\omega_k)\Xi_{TW_2}(\omega_k) & \exp(ih\omega_k)\Xi_{TW_2}(\omega_k) & ... & \exp(ih\omega_k)\Xi_{TW_2}(\omega_k) \\
... & ... & ... & ... \\
\exp(ih\omega_k)\Xi_{TW_m}(\omega_k) & \exp(ih\omega_k)\Xi_{TW_m}(\omega_k & ... & \exp(ih\omega_k)\Xi_{TW_m}(\omega_k)
\end{pmatrix}
$$

where the $\mathrm{Vec}_{\mathrm{row}}$-operator signifies that $\boldsymbol{\Sigma}_{\mathbf{XW},k}$ is a (long) row-vector, made of juxtaposition of the above $m + 1$ rows: the subscript k refers to the k-th component of $\hat{\mathbf{Y}}$ corresponding to ω_k. The indicator function $(1 + I_{k>0}) = \begin{cases} 1 & k = 0 \\ 2 & k > 0 \end{cases}$ accounts for the fact that $|\Xi'_{Tr}(\omega_k)|^2$ is symmetric in ω_k such that all strictly positive frequencies are doubled up. Note that the DFT's in $\boldsymbol{\Sigma}_{\mathbf{XW},k}$ correspond to the *data in level*. One can easily verify that $\boldsymbol{\Sigma}_{\mathbf{XW}}\mathbf{b}$ replicates the first line in 121. We now look at the second line. Proposition 2 implies that the coefficients c_k^0 of the transferfunction $\dfrac{\exp(ih\omega_k)\hat{\Gamma}_X(0) - \exp(-i\omega_k)\hat{\Gamma}_X(\omega_k)}{1 - \exp(-i\omega_k)}$ are given by

$$
c_k^0 = \sum_{j \geq k} b_j^0
$$

unobserved DFT of the target, $\Xi_{TY}(\omega_k)$, by $\tilde{y}_k$ would be inefficient: the optimization criterion would be biased. The power of the decomposition leading to $\Xi'_{Tr}(\omega_k)$ in 119 is to avoid a direct convolution of $\Gamma(\omega_k)$ and $\Xi_{TX}(\omega_k)$, which is achieved by splitting the expression into the level term $\Gamma(0)\Xi_{TX}(\omega_k)$, which is fine by linearity of the DFT (note that this is not a convolution!), and a convolution-term involving stationary first differences only: therefore proposition 5 can be invoked to derive superconsistency of the resulting optimization-criterion (-statistic).

and similarly for the transfer functions corresponding to $\Delta W_{nt}, n = 1, ..., m$. Let $\mathbf{c}$ denote the corresponding vector of length $(L+1)(m+1)$. We now define the following $(L+1)(m+1) * (L+1)(m+1)$ integration matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{S} & 0 & 0 & ... & 0 \\ 0 & \mathbf{S} & 0 & ... & 0 \\ ::: & & & & \\ 0 & 0 & ... & 0 & \mathbf{S} \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & 1 & 1 & ... & 1 \\ ::: & & & & \\ 0 & ... & 0 & 1 & 1 \\ 0 & 0 & ... & 0 & 1 \end{pmatrix}$$

where $\mathbf{S}$ is $(L+1) * (L+1)$. One verifies readily that

$$\mathbf{c} = \mathbf{\Sigma b}$$

We now define a matrix $\mathbf{Z}$ with the following rows

$$\mathbf{Z}_k = -\exp(i\omega_k)(1 + I_{k>0})\mathrm{Vec}_{\mathrm{row}} \begin{pmatrix} \exp(ih\omega_k)\Xi_{T\Delta X}(\omega_k) & \exp(i(h-1)\omega_k)\Xi_{T\Delta X}(\omega_k) & ... & \exp(i(h-L)\omega_k)\Xi \\ \exp(ih\omega_k)\Xi_{T\Delta W_1}(\omega_k) & \exp(i(h-1)\omega_k)\Xi_{T\Delta W_1}(\omega_k) & ... & \exp(i(h-L)\omega_k)\Xi \\ ... & ... & ... & ... \\ \exp(ih\omega_k)\Xi_{T\Delta W_m}(\omega_k) & \exp(i(h-1)\omega_k)\Xi_{T\Delta W_m}(\omega_k & ... & \exp(i(h-L)\omega_k)\Xi \end{pmatrix}$$

where the $\mathrm{Vec}_{\mathrm{row}}$-operator signifies that $\mathbf{Z}_k$ is a (long) row-vector, made of juxtaposition of the above $m+1$ rows: the subscript k refers to the k-th component of $\hat{\mathbf{Y}}$ corresponding to ω_k. Note that the data enters in *first differences* in the DFT's. Then the insert $\mathbf{\Delta_{XW}(1)}$ of the data-matrix corresponding to

$$-\exp(i\omega_k)\frac{\exp(ih\omega_k)\hat{\Gamma}_X(0) - \exp(-i\omega_k)\hat{\Gamma}_X(\omega_k)}{1 - \exp(-i\omega_k)}\Xi_{T\Delta X}(\omega_k)$$

$$-\sum_{n=1}^{m}\exp(i\omega_k)\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)}\Xi_{T\Delta W_n}(\omega_k)$$

becomes

$$\mathbf{\Delta_{XW}(1)} = \mathbf{Z\Sigma}$$

Finally, we look at

$$\exp(i\omega_k)\exp(ih\omega_k)\hat{\Gamma}_X(0)\Xi_{T\Delta X}(\omega_k) + \sum_{n=1}^{m}\exp(i\omega_k)\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\Xi_{T\Delta W_n}(\omega_k)$$

in 121. This term can be replicated by defining the k-th row of $\boldsymbol{\Delta}_{\mathbf{XW}}(\mathbf{2})$ as

$$\boldsymbol{\Delta}_{\mathbf{XW},\mathbf{k}}(\mathbf{2}) = \exp(i(h+1)\omega_k)(1+I_{k>0})\mathrm{Vec}_{\mathrm{row}}\begin{pmatrix} \Xi_{T\Delta X}(\omega_k) & \Xi_{T\Delta X}(\omega_k) & \dots & \Xi_{T\Delta X}(\omega_k) \\ \Xi_{T\Delta W_1}(\omega_k) & \Xi_{T\Delta W_1}(\omega_k) & \dots & \Xi_{T\Delta W_1}(\omega_k) \\ \dots & \dots & \dots & \dots \\ \Xi_{T\Delta W_m}(\omega_k) & \Xi_{T\Delta W_m}(\omega_k & \dots & \Xi_{T\Delta W_m}(\omega_k) \end{pmatrix}$$

where, once again, the $\mathrm{Vec}_{\mathrm{row}}$-operator signifies juxtaposition of the above $m+1$ rows in a long row $\boldsymbol{\Delta}_{\mathbf{XW},\mathbf{k}}(\mathbf{2})$, where the subscript k refers to the k-th component of $\hat{\mathbf{Y}}$.

We can now reformulate the frequency-domain criterion 117 in the more convenient (familiar) form

$$(\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) \to \min_{\mathbf{b}=\mathbf{Rb_f}+\mathbf{c}} \tag{123}$$

where $\mathbf{Y}$ and $\mathbf{X}$ are defined by 120 and 122 and the cointegration constraints 89 are imposed (we ignored the irrelevant scaling $\frac{2\pi}{T}$). In analogy to section 5.1 we now rotate target and data-matrix. Let $\mathbf{Y} = |\mathbf{Y}|\exp(-i\boldsymbol{\Phi})$ where the k-th component of the vector $\boldsymbol{\Phi}$ is $-\arg(y_k)$. Define $\mathbf{Y_{rot}} = \mathbf{Y} \cdot \exp(i\boldsymbol{\Phi}) = |\mathbf{Y}|$ where the operator $\cdot$ means an elementwise product. The rotated data-matrix $\mathbf{X_{rot}}$ is obtained accordingly by rotating the columns of $\mathbf{X}$ by $\boldsymbol{\Phi}$: we multiply elementwise each column with $\exp(i\boldsymbol{\Phi})$. Since the criterion 123 is insensitive to a simultaneous rotation of target and data matrix we may consider the new rotated criterion

$$(\mathbf{Y_{rot}} - \mathbf{X_{rot}b})'(\mathbf{Y_{rot}} - \mathbf{X_{rot}b}) \to \min_{\mathbf{b}=\mathbf{Rb_f}+\mathbf{c}}$$

instead. The mean-square solution in the case of I(1)-processes is then obtained as

$$\begin{aligned} \hat{\mathbf{b}}_{\mathbf{f}} = \quad & \left\{\Re\left[(\mathbf{X}_{\mathrm{rot}}\mathbf{R})'\mathbf{X}_{\mathrm{rot}}\mathbf{R}\right]\right\}^{-1} \\ & \left((\Re(\mathbf{X}_{\mathrm{rot}})\mathbf{R})'\mathbf{Y}_{\mathrm{rot}} - \Re\left\{(\mathbf{X}_{\mathrm{rot}}\mathbf{R})'\mathbf{X}_{\mathrm{rot}}\mathbf{c}\right\}\right) \end{aligned}$$

where $\mathbf{R}$ and $\mathbf{c}$ are defined in 78 and 79 (rank one case) or in 91 and 92 (general rank) and where the solution of the constrained optimization problem was obtained in section 8.3[38]. Note that the above rotation is useful when customizing the filter with respect to timeliness issues since we then emphasize the Imaginary part (the I in the acronym I-MDFA) of $\hat{\mathbf{Y}}$: if the target isn't a real positive vector, then controlling the Imaginary part of the multivariate filter would be delusive[39].

---

[38]We here ignore customization, regularization as well as i1 and i2 constraints.

[39]From a numerical perspective, emphasizing the imaginary part allows to keep the criterion quadratic: the optimum is unique and a closed-form solution exists.

11.6 Customization

Given that we derived $\mathbf{Y_{rot}}$ and $\mathbf{X_{rot}}$ in the previous section, we can proceed according to section 5.3. The corresponding optimization criterion was proposed in section 8.3: one just needs to plug-in the modified $\mathbf{Y_{rot}}$ and $\mathbf{X_{rot}}$ (I(1)-case) as well as the cointegration constraints.

11.7 Regularization

We may proceed according to section 7. The general optimization criterion proposed in section 8.3 applies.

11.8 The Model-Based Approach

Rank-1 vs. Rank $m + 1$. To be filled-in when the multivariate paper is ready.

11.9 An Application of Cointegration to Data Revisions

The topic of data revisions and real-time filtering is addressed in Wildi (2011). The problem is tackled by considering a sequence of releases[40] (not vintages) and by assigning optimal filter weights across releases. Whereas Wildi (2011) emphasizes stationary series (because, trivially, the new cointegration feature was missing) we are now able to extend these results to non-stationary I(1) data. In this framework, cointegration is a natural assumption since we expect revisions linking releases to be 'small' (stationary). Assume that we work with a total of $m + 1$ releases $re_{1t}, ..., re_{m+1,t}$, where $re_{m+1,t}$ is considered to be 'final'[41]. Then

$$re_{1t} - re_{2t}$$

$$re_{1t} - re_{3t}$$

$$:::$$

$$re_{1t} - re_{m+1,t}$$

are stationary linear combinations of the data in levels. Therefore, the cointegration rank is $f = m$ and the cointegration-matrix $\mathbf{A}_{1:m}$ becomes

$$\mathbf{A}_{1:m} = \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 0 & ... & 1 \end{pmatrix}$$

---

[40]The first release is a time series which collects the sequence of initial publications of a time series, say GDP. Since the initial release is subject to uncertainty, the data will be revised in subsequent releases. The k-th release is a time series which collects published data subject to the k-th revision.

[41]In general revisions after the first year become small. From a practical perspective we may thus consider the corresponding release to be 'final'.

The whole cointegration framework can then be applied to tackle the data-revision problem in a formal way.

11.10 Consistency and Efficiency: a Tale of Two Philosophies

Consistency means that an estimate, here $\mathbf{b}$ (or $\mathbf{b_f}$ in the case of constrained designs), converges to the *true* unknown parameters (filter coefficients) when sample size T augments. Efficiency assigns preference to a particular estimate which is closest possible - in some metric (for example MSE) - to the *truth*. So let's talk about 'truth'.

11.10.1 Knowing the Truth: Omniscience

McElroy and Wildi (2013) tackle the multivariate real-time signal extraction problem by analysing the structure of

$$\int_{-\pi}^{\pi} |\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)|'\mathbf{h}(\omega)|\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)|d\omega \quad = \quad E[r_t^2]$$

in the case of (co)integrated processes. Here, $\mathbf{\Gamma}(\omega) = (\Gamma_X(\omega), \Gamma_{W_1}(\omega, ..., \Gamma_{W_m}(\omega))'$ and $\hat{\mathbf{\Gamma}}(\omega) = (\hat{\Gamma}_X(\omega), \hat{\Gamma}_{W_1}(\omega, ..., \hat{\Gamma}_{W_m}(\omega))'$ are vectors of transferfunctions and $\mathbf{h}(\omega)$ is the (unknown) spectral matrix[42]. We allow for arbitrary unit-roots (order, location) and we consider a multivariate target signal $\mathbf{\Gamma}(\omega)$ which allows to replicate DFM's, for example[43]. In this example, the spectral matrix $\mathbf{h}(\omega)$ embodies the 'truth' and we (first) assume that $\mathbf{h}(\omega)$ is known. We then analyse the solution of the 'true' criterion

$$\int_{-\pi}^{\pi} |\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)|'\mathbf{h}(\omega)|\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)|d\omega \to \min_{\hat{\mathbf{\Gamma}}}$$

and derive some nice closed-form solution β. Unsurprisingly, this approach coincides with the classical one-step ahead mean-square paradigm, under suitable model assumptions i.e. we can replicate traditional VARMA or DFM approaches. Being ambitious but cautious, we call the solution of the above criterion a 'pseudo-true' value. Of course, this criterion is impracticable: it is a scheme, summarizing some of our primary intentions in a reduced form.

11.10.2 Believing in Truth: Faith and Fatalism

Since $\mathbf{h}(\omega)$ is generally unknown we look for a smart substitute of $\mathbf{h}(\omega)$ such that the solution of the resulting real-world criterion would be as close as possible to the pseudo-true coefficients[44]. It

---

[42]Care is needed in order to ensure stationarity of all terms such that $\mathbf{h}(\omega)$ exists. Unsurprisingly, in the I(1) case, we just need to impose the cointegration constraints proposed in section 9.

[43]The case treated here corresponds to I(1)-processes with a single unit-root in frequency zero and $\mathbf{\Gamma}(\omega) = (\Gamma(\omega), 0, ..., 0)'$. Experience suggests that this framework is sufficiently general to tackle some of the most interesting applications (economic indicators, financial trading).

[44]Note that we do not look for a good estimate of $\mathbf{h}(\omega)$; instead we want a good estimate of the pseudo-true value.

should not come as a surprise that we (Tucker and I) insert the DFT of the data:

$$\int_{-\pi}^{\pi} |\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)|' \mathbf{I}_{T\mathbf{X}}(\omega) |\mathbf{\Gamma}(\omega) - \hat{\mathbf{\Gamma}}(\omega)| d\omega \to \min_{\hat{\mathbf{\Gamma}}}$$

where $\mathbf{I}_{T\mathbf{X}}(\omega)$ is the outer-product of the DFT's. Under particular (model) assumptions, the (asymptotic) distribution of the resulting empirical estimate $\mathbf{b}$ can be derived and one can show that the estimate converges to the pseudo-true value i.e. it is consistent. Efficiency can be derived from the fact that we address the Whittle-likelihood, see McElroy and Wildi (2013).

The above approach shares the inherent Platonic perspective of model-based approaches where it is assumed that a true latent DGP exists but noise hinders a conclusive identification of its structure[45]. An interesting implication of this perspective is that (RTSE) 'performances' are subordinated to the model: if the empirical model is the 'chosen one' - Faith - then performances will follow 'automatically'. Otherwise, Fatalism is part of the design.

The original derivation and justification of (M)DFA, proposed in Wildi (1998) and formalized in Wildi (2005), was different, though.

11.10.3 From Truth to Effectiveness: Emphasizing Performances

Let's assume that we don't care about a vain quest of a dissimulating truth. Instead, we worship the user. Users are driven by mercantile incentives. Therefore, a good statistical design should address 'preferences' and 'priorities'. Users don't care about pseudo-true coefficients; instead they are in pursuit of outperformance. How can we deliver? Obviously, we should address performances 'directly':

$$\frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 = \frac{1}{T} \sum_{t=1}^{T} r_t^2 \to \min_{\mathbf{b_f}}$$

where Y_t is a target, specified by the user, and $\hat{Y}_t$ is the estimate: if Y_t is the one-step ahead target i.e. if $Y_t = X_{t+1}$, then we replicate the one-steap ahead paradigm. Whatever solution comes out: we are not interested in an exegesis about the true latent process. Consistency misses the point. We just want the filter to perform well.

A shortcoming of the above 'idealized' criterion is that Y_t is generally unobserved (it can be the output of a bi-infinite filter). Interestingly, this argument can be wipped off by entering the

---

[45]The link is intentional: Tucker and I want to show that model-based approaches can be replicated perfectly by (M)DFA, when inserting a suitable estimate of $\mathbf{h}$. Therefore, (M)DFA can be firmly anchored into model-based territory.

frequency-domain. Specifically, we showed that the argument $\mathbf{b_f}$ of the constrained minimization problem

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \left| \Xi'_{Tr}(\omega_k) \right|^2 \to \min_{\mathbf{b_f}}$$

minimizes

$$\frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2$$

up to a (very) small approximation error. In order to illustrate the magnitude of this error, we showed that it would be masked by the natural sampling error (assuming we could sample the DGP). The 'D' in the acronym DFA stands for the unusually tight (direct) link between the criterion and the objective of the analysis.

We note two important implications of this DFApproach. First, the user exerts direct control on performances through the criterion. Customization allows to implement further priorities. Fatalism is whipped-off. Second, efficiency and consistency do not apply to 'coefficients' anymore but to 'performances' instead. An approach is consistent if its solution maximizes performances up to an asymptotically vanishing error; it is efficient if this error is smallest possible. Super-consistency of the (M)DFA-criterion 117 warrants both requirements[46]. Interestingly, conditions establishing super-consistency are related to mild regularity assumptions, see proposition 5[47]. This inherent simplicity confers the whole approach a robust basement to operate on the edge of the ATS-trilemma (McElroy and Wildi (2013)).

12   Adaptive Filtering

In practice users often work with data which is not strictly trending, though level and possibly autocorrelation structures are changing over time. Typical examples are financial applications (when working in log-returns) or economic (real-time) indicators (when working on 'growth'). Also, some indicators cannot grow unboundedly by construction (business surveys) and therefore the difference operator would be a misspecification – although the series are felt to be 'non-stationary' –.

---

[46]Proposition 6 shows that the solution $\mathbf{b_f}$ of the above (M)DFA-criterion minimizes the 'true' unobserved mean-square filter error $E[r_t^2]$ up to a *smallest possible* approximation error if some additional distributional (i.e. model-based) assumptions are met.

[47]The only distributional assumption we need in the technical proposition 3 underlying this result is that second order moments of the stationary (differenced) processes exist. We don't need the typical iid assumption (independent identically distributed) nor do we require Gaussianity. Of course, if the Gaussian-iid assumption is violated, then our solution supports efficiency in the family of linear estimators only. But experience suggests that this family is rich enough for many economic forecast problems, see Gyomai and Wildi (2013) for example; also, past forecasting competitions (for example M3, NN3, NN5) seem to support *consistently* this statement.

This rich family of 'bounded' time-varying DGP's represents a challenge either for modelling or for filtering. We here propose a series of techniques which are felt to be useful in this general context. Let us emphasize that the 'boundedness' assumption enables to ignore filter constraints (i1 and/or i2 and/or cointegration restrictions) in our formal development which simplifies exposition. This simplification is merely a convenience and does not preclude application of these techniques to asymptotically unbounded time series (assuming suitably constrained designs).

The general proceeding here will be to address 'non-stationarity' by overemphasizing recent data or, equivalently, by down-weighting the importance of the remote past[48]. Various approaches can be distinguished according to the way, a particular weighting scheme is defined. The first method that we analyze is based on a very crude weighting scheme as defined by applying a window to the data: the data in the window is felt to be relevant - without reference to time i.e. recent and old data are considered equally informative - whereas the data on the outside is discarded. After reviewing univariate and multivariate approaches based on this 'coarse' concept, we then propose more refined - gradual - weighting schemes.

12.1 Up-Dating Filter Coefficients through 'Windowing'

We here assume, for simplicity, that windows of fixed length - a sliding span - are applied to the data[49] and we assume that a particular multivariate design - an I-MDFA filter (possibly regularized and/or customized) - has been estimated in a previous period: this will be referred to as the 'old' filter. We can now apply this 'old' filter to recent data (out-of-sample) and hope that its performances do not fade over time. However, many users feel uncomfortable when a particular design is not 'refreshed' periodically – I do not belong to this group! –. After a deep recession, a structural-shift might have affected the DGP, say (at least we are sure that GDP has been affected...). Therefore, the 'old' filter might have to be re-calibrated. We here distinguish two forms of re-calibration, namely uni and multivariate up-dating.

12.1.1 Univariate Up-Dating

One could 'correct' the failing output of the *'old'* filter by applying an univariate *'correction'* filter: the resulting convolution of (multivariate) *'old'* and (univariate new) *'correction'* filter is called the *'up-dated'* filter. This strategy could be implemented very easily withing I-MDFA[50]: formally, we

---

[48]This approach has be contrasted with difference stationary series: for the latter, the remote past is as important - when estimating parameters - as the most recent data.

[49]As an illustration we use a time-span of eight years when tracking Euro-Area GDP by a real-time indicator (EURI) on SEFBlog: the width of eight years corresponds roughly to the length of a business-cycle. See http://blog.zhaw.ch/idp/sefblog/index.php?/archives/291-Univariate-Filter-Up-Dating-Examples-2-8.html

[50]It would not require a lot of 'fresh' degrees of freedom i.e. overfitting could be controlled quite well even on shorter windows: therefore structural-shifts could be tracked faster (bias-variance dilemma). On the other hand, the

solve the following estimation problem

$$\frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} |\Gamma(\omega_k)\Xi_{T_{\text{span}}X}(\omega_k) - \hat{\Gamma}_{\text{cor}}(\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)|^2 \tag{124}$$

where $\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)$ is the DFT of the output $\hat{Y}_{\text{old},t}$ of the 'old' (multivariate) filter and where the DFT's are sampled on the relevant estimation window (up-date span) of length $T_{\text{span}} << T$. In principle, we now can apply customization and/or regularization and/or filter constraints to $\hat{\Gamma}_{\text{cor}}(\omega_k)$ in 124 : a comprehensive tutorial on this topic is provided on SEFBlog[51] where it is shown that particular design-settings can address various user-priorities (small up-dating revisions, fast and smooth filter outputs) in an effective way. Note, however, that some care is needed when relying on regularization (trivially there is no cross-sectional regularization since the design is univariate); to see this, assume that our Null-hypothesis H_0 is: "the 'old' filter is the best possible solution" (for example because the data is stationary and because we know the 'true' coefficients). Then $\hat{\Gamma}_{\text{cor}}(\cdot) \equiv 1$ must be an identity which implies that the coefficient vector $\mathbf{b}_{\text{cor}}$ of the 'correction' filter must satisfy: $b_{0,\text{cor}} = 1$ and $b_{j,\text{cor}} = 0$, $j = 1, ..., L_{\text{cor}} - 1$. The sharp decay of filter weights in this case complies with the decay-regularization but it conflicts with the smoothness-term. In order to solve this conflict we can rely on the general proceeding proposed in section 7.3. Specifically, we decompose the vector $\mathbf{b}_{\text{cor}}$ according to

$$\mathbf{b}_{\text{cor}} = \mathbf{b}_{\mathbf{0},\text{cor}} + \mathbf{b}_{\text{res,cor}} \tag{125}$$

where $\mathbf{b}_{\mathbf{0},\text{cor}}' = (1, 0, ..., 0)$. We can then estimate $\mathbf{b}_{\text{res,cor}}$ by relying on 55: in this case regularization (in particular the smoothness requirement) does no more conflict with our Null-hypothesis.

Expressed in terms of criterion 124 we obtain:

$$\frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} |\Gamma(\omega_k)\Xi_{T_{\text{span}}X}(\omega_k) - \left(\sum_{j=0}^{L_{\text{cor}}-1} b_{j,\text{cor}}\exp(-ij\omega_k)\right)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)|^2$$

$$= \frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} |\Gamma(\omega_k)\Xi_{T_{\text{span}}X}(\omega_k) - \left((1 + b_{0\text{res}}) + \sum_{j=1}^{L_{\text{cor}}-1} b_{j,\text{res}}\exp(-ij\omega_k)\right)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)|^2$$

$$= \frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} |\Gamma(\omega_k)\Xi_{T_{\text{span}}X}(\omega_k) - \Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k) - \hat{\Gamma}_{\text{res}}(\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)|^2$$

$$= \frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} |\Delta_{\text{old}}(\omega_k) - \hat{\Gamma}_{\text{res}}(\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}}(\omega_k)|^2$$

'univariate' correction filter is blind to structural shifts in the cross-section (of the data).

[51] http://blog.zhaw.ch/idp/sefblog/index.php?/archives/291-Univariate-Filter-Up-Dating-Examples-2-8.html.

where $\hat{\Gamma}_{\mathrm{res}}(\omega_k) = \sum_{j=0}^{L_{\mathrm{cor}}-1} b_{j,\mathrm{res}} \exp(-ij\omega_k)$ is the 'residual' transfer function whose unknown parameter vector $\mathbf{b}_{\mathrm{res,cor}}$ must be estimated; also

$$\Delta_{\mathrm{old}}(\omega_k) := \Gamma(\omega_k)\Xi_{T_{\mathrm{span}}X}(\omega_k) - \Xi_{T_{\mathrm{span}}\hat{Y}_{\mathrm{old}}}(\omega_k) \tag{126}$$

is the DFT of the error obtained by the 'old' filter in the up-date span (prior to correction). Under the above 'Null' we must have $\hat{\Gamma}_{\mathrm{res}}(\cdot) \equiv 0$. In this case, zero-shrinkage by the Regularization Troika is in accordance with the assumption of a well-behaved (out-of-sample) 'old' filter, as desired.

12.1.2 Multivariate Up-Dating

A multivariate up-dating strategy of an 'old' filter relies on multiple time series which implies, at least in principle, that we have more options at disposal than in the previous section. Let us first shortly describe what we're aiming at in order to restrict our choice in some meaningful way. We (I do) assume that the 'old' filter is more or less 'well-behaved' out-of-sample, due to careful design (regularization, customization, filter constraints are all at our disposal) and due to moderate non-stationarity of the DGP. Therefore, re-estimation of *all* coefficients of the 'old' filter is not required. Our preference here goes to simpler, shorter, faster - smarter - multivariate 'correction' filters which adapt the 'old' filter to new data by relying on few coefficients (short filter lengths): this way, narrower data windows (short up-date spans) can be used for estimation purpose and the corrected 'old' filter can track structural changes faster (if needed). Relying on the fundamental bias-variance dilemma (decomposition of the MSE into squared bias and variance), we may claim, schematically, that the new 'correction' filter draws attention to the 'bias' term whereas the 'old' filter emphasizes the 'variance' part (rather).

A natural generalization of the previous *univariate* design would be to apply specific 'correction' filters to the individual output-series of the 'old' filter (see criterion 127 below): the 'old' sub-filters provide the task of optimal *pre-filtering* of the raw data whereas the 'correction' sub-filters focus on 'adaptivity'. Under our standard Null-hypothesis – namely that the 'old' filter performs well out-of-sample – the 'correction' filter should degenerate to an identity; but even if this null-hypothesis does not apply strictly, we may nevertheless entertain some hope that the proper correction-task remains 'simple'[52].

---

[52]In contrast, the *pre-filtering* task sustained by the 'old' filter is likely to be (much) more challenging, since an effective noise suppression may require implementation of complex lag-distributions.

Formally, we propose to minimize the following (MSE) expression

$$\frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} \left| \Gamma(\omega_k) \Xi_{T_{\text{span}}X}(\omega_k) - \sum_{l=0}^{m} \hat{\Gamma}_{\text{cor}}^l(\omega_k) \hat{\Gamma}_{\text{old}}^l(\omega_k) \Xi_{T_{\text{span}}W^l}(\omega_k) \right|^2$$

$$= \frac{2\pi}{T_{\text{span}}} \sum_{k=-[T_{\text{span}}/2]}^{[T_{\text{span}}/2]} \left| \Gamma(\omega_k) \Xi_{T_{\text{span}}X}(\omega_k) - \sum_{l=0}^{m} \hat{\Gamma}_{\text{cor}}^l(\omega_k) \Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^l}(\omega_k) \right|^2 \to \min_{\hat{\mathbf{b}}_{\text{cor}}} \quad (127)$$

where $m+1$ denotes the number of explanatory time series[53] (the dimension of the multivariate design), $\hat{\Gamma}_{\text{cor}}^l(\omega_k)\hat{\Gamma}_{\text{old}}^l(\omega_k)$ is a convolution between the series-specific 'correction' filters – whose coefficients are estimated – and the series-specific 'old' filters – which are fixed – . Also, $\Xi_{T_{\text{span}}W^l}(\omega_k)$ are the DFT's of the explanatory series where, for $l = 0$, we have $\Xi_{T_{\text{span}}W^0}(\omega_k) = \Xi_{T_{\text{span}}X}(\omega_k)$ see footnote 53. The above expression is thus minimzed with respect to the coefficients $\hat{\mathbf{b}}_{\text{cor}}$ of the 'correction' filters . The DFT's of the series-specific 'old' filter outputs are denoted by $\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^l}(\omega_k)$ in 127. All DFT's are sampled on the relevant *up-date* span whose length $T_{\text{span}} << T$ is (much) smaller than the full data length.

As discussed in the previous section, we could easily generalize the MSE-criterion 127 by applying customization and/or regularization. For this purpose we decompose the coefficient vector according to 125

$$\mathbf{b}_{\text{cor}} = \mathbf{b}_{\mathbf{0},\text{cor}} + \mathbf{b}_{\text{res},\text{cor}}$$

where $\mathbf{b}_{\mathbf{0},\text{cor}}' = (1, 0, ..., 0|1, 0...., 0|...|1, 0, ..., 0)$. The $m+1$ identity-filters stacked into $\mathbf{b}_{\mathbf{0},\text{cor}}$ reflect our Null-hypothesis: if this hypothesis applies, then we expect the estimate $\mathbf{b}_{\text{res},\text{cor}}$ to be small (ideally vanishing). We can then estimate $\mathbf{b}_{\text{res},\text{cor}}$ by relying on 55. Specifically, the k-th component of the residual target $\mathbf{Y}_{\text{res}}$ in 53 is defined as

$$\mathbf{Y}_{\text{res},k} := \Delta_{\text{old}}(\omega_k)$$

where $\Delta_{\text{old}}(\omega_k)$ is the DFT of the 'old' filter error 126. The design-matrix before rotation is obtained as

$$\mathbf{X}_{k,\text{res}} = (1 + I_{k>0})\text{Vec}_{\text{row}} \begin{pmatrix} \Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^0}(\omega_k) & \exp(-i\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^0}(\omega_k) & ... & \exp(-iL\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^0}(\omega_k) \\ ... & ... & ... & ... \\ \Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^m}(\omega_k) & \exp(-i\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^m}(\omega_k) & ... & \exp(-iL\omega_k)\Xi_{T_{\text{span}}\hat{Y}_{\text{old}}^m}(\omega_k) \end{pmatrix}$$

---

[53]Recall the convention $W_t^0 = X_t$ if $l = 0$. It is not always recommended to include X_t – the series to which the symmetric target filter is applied – in the set of explanatory variables. As an example, GDP is generally not very helpfull when targeting a GDP-cycle in real-time, because of data-revisions and publication lags. In such a case, the summands in the above expressions would start as $l = 1$ signifying that X_t is not an explanatory variable.

where the DFT's of the series-specific outputs of the 'old' filter $\Xi_{T_{\text{span}}\hat{Y}^l_{\text{old}}}(\omega_k)$ are computed on the *up-date* span (of length T_{span}). We can now proceed to rotation and obtain

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

and

$$\mathbf{X}_{k,\text{rot,res}} = \mathbf{X}_{k,\text{res}} \exp\left(-i * \arg\left(\mathbf{Y}_{k,\text{res}}\right)\right)$$

where $\mathbf{X}_{k,\text{rot, res}}$ designates the k-th row (of the new rotated design matrix). At this stage the whole I-MDFA formalism applies and we can rely on the general criterion 55 for estimating $\mathbf{b}_{\text{res}}$.

To conclude, we note that the proposed *multivariate* up-dating strategy is able to replicate the previous *univariate* design by imposing a strong cross-sectional regularization (to the former): we are thus proposing a natural parametric extension (of the latter).

13   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 **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).

- 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 57 possibly with constraints imposed through 66. 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):

$$\sum_k |\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k)) - i * \sqrt{1 + \lambda\Gamma(\omega_k)} Im(\hat{\Gamma}(\omega_k))|^2 I_{TX}(\omega_k)$$

$$= \sum_k \left(\left[\Gamma(\omega_k) - Re(\hat{\Gamma}(\omega_k))\right]^2 + (1 + \lambda\Gamma(\omega_k)) Im\left(\hat{\Gamma}(\omega_k)\right)^2 \right) I_{TX}(\omega_k)$$

We now differentiate this expression with respect to filter parameters:

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

Now $-d/db_j(Re(\hat{\Gamma}(\omega_k))) = -\cos(j\omega_k)$ and $d/db_j(Im(\hat{\Gamma}(\omega_k))) = \sin(j\omega_k)$[54]. Therefore we obtain

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

Or

$$\sum_k (\Gamma(\omega_k) \cos(j\omega_k)) I_{TX}(\omega_k) = \sum_k \left(Re(\hat{\Gamma}(\omega_k)) \cos(j\omega_k) + (1 + \lambda\Gamma(\omega_k)) Im(\hat{\Gamma}(\omega_k)) \sin(j\omega_k) \right) I_{TX}(\omega_k) \quad (128)$$

---

[54]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.

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 128:

$$b_0 \sum_k (\cos(j\omega_k)\cos(0\omega_k) + (1 + \lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(0\omega_k))I_{TX}(\omega_k) +$$

$$b_1 \sum_k (\cos(j\omega_k)\cos(1\omega_k) + (1 + \lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(1\omega_k))I_{TX}(\omega_k) +$$

$$... +$$

$$b_L \sum_k (\cos(j\omega_k)\cos(L\omega_k) + (1 + \lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(L\omega_k))I_{TX}(\omega_k)$$

Let $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$. Note that all expressions are real here, in contrast to 37, 40, 58, 59 and 71 which require usage of the $\Re$-operator. Then the right-hand side of 128 implies

$$\mathbf{X'X} = \left(\sum_k (\cos(j\omega_k)\cos(m\omega_k) + (1 + \lambda\Gamma(\omega_k))\sin(j\omega_k)\sin(m\omega_k))I_{TX}(\omega_k)\right)_{jm} \tag{129}$$

where both indices $0 \le j, m \le L$. Finally, the left-hand side of 128 implies that

$$\mathbf{X'Y} = \left(\sum_k (\Gamma(\omega_k)\cos(j\omega_k))I_{TX}(\omega_k)\right)_j \tag{130}$$

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

B   I-MDFA: Working out Trigonometric Terms

The criterion is:

$$\sum_k \left|\Gamma(\omega_k) - \tilde{\Gamma}(\omega_k)\right|^2 |\Xi_{TX}(\omega_k)|^2 \to \min_{\mathbf{B}} \tag{131}$$

where

$$\tilde{\Gamma}(\omega_k) := \hat{\Gamma}_X(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\frac{\Xi_{TW_n}(\omega_k)}{\Xi_{TX}(\omega_k)} \tag{132}$$

Instead of the numerically potentially unstable ratio of DFT's appearing in 132, 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):

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

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

72

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:

$$\sum_k \left| \Gamma(\omega)\Xi_{TX}(\omega_k) - \Re\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k) \right) \right. \tag{133}$$

$$\left. -i * \sqrt{1 + \lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k) + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k) \right) \right\} \right|^2 \tag{134}$$

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

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

$$+ \sum_k \left(\Gamma(\omega)\Im\left(\Xi_{TX}(\omega_k)\right) - \sqrt{1 + \lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\Xi_{TX}(\omega_k)\right) + \sum_{n=1}^{m} \Im\left(\hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\right) \right\} \right)^2 \tag{135}$$

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 135.

- 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 134 would not necessarily lead to the expected improvement because the target signal $\Gamma(\omega)\Xi_{TX}(\omega_k)$ in 133 is a complex number with a non-vanishing imaginary part too.

Both problems could be avoided in 131, 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

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

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

$$\left|\exp\left(i\arg\left(\Xi_{TX}(\omega_k)\right)\right)\right|^2$$

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:

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

$$\left. -i * \sqrt{1 + \lambda\Gamma(\omega_k)} \left\{ \Im\left(\hat{\Gamma}_X(\omega_k)\left|\Xi_{TX}(\omega_k)\right| + \sum_{n=1}^{m} \hat{\Gamma}_{W_n}(\omega_k)\Xi_{TW_n}(\omega_k)\exp\left(-i\arg\left(\Xi_{TX}(\omega_k)\right)\right) \right) \right\} \right|^2 \tag{136}$$

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 $\Gamma(\omega)|\Xi_{TX}(\omega_k)|$ 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 136 by:

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

We can now proceed to the formal solution by differentiating expression 136 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

$$
\begin{aligned}
&\sum_k \Gamma(\omega)\tilde{\Xi}_{TX}(\omega_k)\Re\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right) \\
&= \sum_k \left\{\Re\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) + \sum_{n=1}^m \Re\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right\}\Re\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right) \qquad (137) \\
&\quad + \sum_k \sqrt{(1+\lambda\Gamma(\omega_k))}\left\{\Im\left(\hat{\Gamma}_X(\omega_k)\tilde{\Xi}_{TX}(\omega_k)\right) + \sum_{n=1}^m \Im\left(\hat{\Gamma}_{W_n}(\omega_k)\tilde{\Xi}_{TW_n}(\omega_k)\right)\right\}\Im\left(\exp(ij\omega_k)\tilde{\Xi}_{TW_m}(\omega_k)\right)
\end{aligned}
$$

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 $\tilde{\Xi}_{TX}(\omega_k)$ is real after rotation)

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

which corresponds to 130. The right-hand side simplifies to

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

which corresponds to the data-matrix 129.

The right-hand side of equation 137 (the criterion differentiated with respect to b_j^m) attributes the following weight to the filter coefficient b_l^u:

$$\Re\Big(\exp(il\omega_k)\Xi_{TW_u}(\omega_k)\Big)\Re\Big(\exp(ij\omega_k)\Xi_{TW_m}(\omega_k)\Big)$$
$$+\sqrt{(1+\lambda\Gamma(\omega_k))}\Im\Big(\exp(il\omega_k)\Xi_{TW_u}(\omega_k)\Big)\Im\Big(\exp(ij\omega_k)\Xi_{TW_m}(\omega_k)\Big)$$

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 129 in the univariate case. This expression is used in the published R-code.

C Alternative Solution of Regularization Towards General H_0-Shrinkage

The solution of criterion 55, allowing for shrinkage towards arbitrary H_0, is obtained as follows:

$$
\begin{aligned}
d/d\mathbf{b}_{\text{res}} \text{ Criterion} \;=\;& d/d\mathbf{b}_{\text{res}}\; (\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{b_0}+\mathbf{b}_{\text{res}}))'(\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \mathbf{X}_{\text{rot}}^{\text{Cust}}(\mathbf{b_0}+\mathbf{b}_{\text{res}}))\\
+\;& d/d\mathbf{b}_{\text{res}}\; \lambda_{\text{smooth}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{smooth}}\mathbf{b}_{\text{res}}\\
+\;& d/d\mathbf{b}_{\text{res}}\; \lambda_{\text{cross}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{cross}}\mathbf{b}_{\text{res}}\\
+\;& d/d\mathbf{b}_{\text{res}}\; \lambda_{\text{decay}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{decay}}\mathbf{b}_{\text{res}}\\
=\;& -2(\mathbf{Y}_{\text{rot}}^{\text{Cust}})'\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}) - 2\Re\Big\{(\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_0})'(\mathbf{X}_{\text{rot}}^{\text{Cust}})\Big\} + 2\mathbf{b}_{\text{res}}{}'\Re\Big\{(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\Big\}\\
+\;& 2\lambda_{\text{smooth}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{smooth}}\\
+\;& 2\lambda_{\text{cross}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{cross}}\\
+\;& 2\lambda_{\text{decay}}\mathbf{b}_{\text{res}}{}'\mathbf{Q_{decay}}
\end{aligned}
$$

where we used the fact that bilinear forms are symmetric. Equating this expression to zero, we obtain

$$
\begin{aligned}
&\hat{\mathbf{b}}^{\text{Cust-Reg}}(\lambda,\eta,\lambda_{\text{smooth}},\lambda_{\text{cross}},\lambda_{\text{decay}},\mathbf{b_0}) =\\
=\;& \mathbf{b_0} + \Big\{\Re\big[(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\big] + \lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\Big\}^{-1}\\
& \Big((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}))'\mathbf{Y}_{\text{rot}}^{\text{Cust}} - \Re\Big\{(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_0}\Big\}\Big)\\
=\;& \mathbf{b_0} + \Big\{\Re\big[(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\big] + \lambda_{\text{smooth}}\mathbf{Q_{smooth}} + \lambda_{\text{cross}}\mathbf{Q_{cross}} + \lambda_{\text{decay}}\mathbf{Q_{decay}}\Big\}^{-1}\\
& \Big((\Re(\mathbf{X}_{\text{rot}}^{\text{Cust}}))'\mathbf{Y}_{\text{rot}}^{\text{Cust}} + \mathbf{Const}\Big) \qquad\qquad (138)
\end{aligned}
$$

where the constant

$$\mathbf{Const} = \Re\Big\{(\mathbf{X}_{\text{rot}}^{\text{Cust}})'\mathbf{X}_{\text{rot}}^{\text{Cust}}\mathbf{b_0}\Big\}$$

depends on the null-hypothesis H_0. In particular, if $H_0 : \mathbf{b_0} = \mathbf{0}$, then the constant vanishes and 138 simplifies to 58. Note that we need to add $\mathbf{b_0}$ in these expressions because the derivative was taken with respect to $\mathbf{b}_{\text{res}}$ (and not $\mathbf{b_c}$).

D   Propositions

We first state a fundamental set of regularity assumptions necessary in deriving efficiency and super-consistency.

Definition 1 *A general (not necessarily symmetric) filter $\Gamma(\cdot)$ is said to belong to the class C_f^u, $u \in I\!\!R$, if*

$$\sum_{k=-\infty}^{\infty} |\gamma_k||k|^u < \infty$$

By analogy, a stationary process

$$X_t = \mu_x + \sum_{k=-\infty}^{\infty} b_k \epsilon_{t-k} \tag{139}$$

where ϵ_t is a white noise process, is said to belong to the class C_f^u if

$$\sum_{k=-\infty}^{\infty} |b_k||k|^u < \infty$$

The above definition does not make a distinction between processes and filters. Instead, the rate of decay of the weights is addressed. Note that a stationary ARMA-process or a stable ARMA-filter are in C_f^u for all $u > 0$ which is subsequently denoted by $X_t \in C_f^\infty$. Also, every finite (MA-)filter or MA-process is in C_f^∞.

We first propose a result ensuring inheritance of regularity properties through convolution.

Proposition 1 *If a_j, b_k are sequences satisfying $\sum_{j=-\infty}^{\infty} |a_j||j|^u < \infty$, $\sum_{k=-\infty}^{\infty} |b_k||k|^v < \infty$ where $u, v \geq 0$, then*

$$\sum_{j=-\infty}^{\infty} \left| \sum_{k=-\infty}^{\infty} a_k b_{j\pm k} \right| |j|^{\min(u,v)} < \infty$$

Proof

76

Let $w := \min(u, v) \geq 0$. Then

$$\sum_{j=-\infty}^{\infty} \left| \sum_{k=-\infty}^{\infty} a_k b_{j \pm k} \right| |j|^w \leq \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_k b_{j \pm k}| \, |j|^w$$

$$\leq \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_k b_{j \pm k}| \max(1, 2^{w-1})(|j \pm k|^w + |k|^w)$$

$$= \max(1, 2^{w-1}) \left(\sum_{k=-\infty}^{\infty} |a_k| \sum_{j=-\infty}^{\infty} |b_j| |j|^w + \sum_{k=-\infty}^{\infty} |a_k| |k|^w \sum_{j=-\infty}^{\infty} |b_j| \right)$$

$$< \infty$$

which completes the proof of the proposition.

We next specify the form of the centered integrated transfer functions in section 11.2.

Proposition 2 *Let* $\Gamma(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k \exp(-ik\omega) = \Gamma_h^+(\omega) + \Gamma_h^-(\omega)$ *where* $\Gamma_h^+(\omega) = \sum_{k \geq h} \gamma_k \exp(-ik\omega)$ *and* $\Gamma_h^-(\omega) = \sum_{k=-\infty}^{h-1} \gamma_k \exp(-ik\omega)$ *and where* h *is any (positive or negative) integer.*

- *Then* $\gamma_k^{+'} := \sum_{j=k}^{\infty} \gamma_j, k \geq h$, *are the coefficients of a filter with transfer function*

$$\Gamma_h^{+'}(\omega) := \frac{\exp(-ih\omega)\Gamma_h^+(0) - \exp(-i\omega)\Gamma_h^+(\omega)}{1 - \exp(-i\omega)} \tag{140}$$

- *The coefficients* $\gamma_k^{-'} := \sum_{j=-\infty}^{k} \gamma_j, k < h$ *pertain to a filter with transfer function*

$$\Gamma_h^{-'}(\omega) := -\frac{\exp(-ih\omega)\Gamma_h^-(0) - \Gamma_h^-(\omega)}{1 - \exp(-i\omega)} \tag{141}$$

or, equivalently

$$\Gamma_h^{-'}(\omega) := \frac{\exp(-i(h-1)\omega)\Gamma_h^-(0) - \exp(i\omega)\Gamma_h^-(\omega)}{1 - \exp(i\omega)} \tag{142}$$

- *If* $\Gamma(\cdot) \in C_f^{1+\delta}$, $\delta \geq 0$, *then* $\Gamma_{k_0}^{+'}(\cdot), \Gamma_{k_0}^{-'}(\cdot) \in C_f^{\delta}$.

Proof

We consider the coefficients $\tilde{\gamma}_k$ of the filter with transfer function

$$\tilde{\Gamma}(\omega) := (1 - \exp(-i\omega))\Gamma_h^{+'}(\omega)$$

$$= (1 - \exp(-i\omega)) \sum_{j \geq h} \gamma_j' \exp(-ij\omega)$$

$$\tilde{\gamma}_h = \gamma_h' = \Gamma_h^+(0)$$

$$\tilde{\gamma}_{h+1} = \gamma_{h+1}' - \gamma_h' = -\gamma_h$$

$$\tilde{\gamma}_{h+2} = \gamma_{h+2}' - \gamma_{h+1}' = -\gamma_{h+1}$$

$$\tilde{\gamma}_{h+k} = \gamma_{h+k}' - \gamma_{h+k-1}' = -\gamma_{h+k-1}$$

Therefore

$$(1 - \exp(-i\omega)) \sum_{j \geq h} \gamma'_j \exp(-ij\omega) \;=\; \exp(-ih\omega)\Gamma_h^+(0) - \exp(-i\omega)\Gamma_h^+(\omega)$$

which achieves the proof of the first assertion. Note that 140 is well-defined in frequency zero if $\Gamma(\omega)$ and thus $\Gamma_h^+(\omega)$ are sufficiently regular. For the second assertion we consider the coefficients $\tilde{\gamma}_k$ of the filter with transfer function

$$\tilde{\Gamma}(\omega) = (1 - \exp(-i\omega)) \sum_{j=-\infty}^{h-1} \gamma'_j \exp(-ij\omega)$$

$$
\begin{aligned}
\tilde{\gamma}_h &= -\gamma'_{h-1} = -\Gamma_h^-(0) \\
\tilde{\gamma}_{h-1} &= \gamma'_{h-1} - \gamma'_{h-2} = \gamma_{h-1} \\
\tilde{\gamma}_{h-2} &= \gamma'_{h-2} - \gamma'_{h-3} = \gamma_{h-2} \\
\tilde{\gamma}_{h-k} &= \gamma'_{h-k} - \gamma'_{h-k-1} = \gamma_{h-k}
\end{aligned}
$$

Therefore

$$(1 - \exp(-i\omega)) \sum_{j=-\infty}^{h-1} \gamma'_j \exp(-ij\omega) \;=\; - \exp(-ih\omega)\Gamma_h^-(0) + \Gamma_h^-(\omega)$$

which shows 141. Equation 142 is achieved by looking at the coefficients $\tilde{\gamma}_k$ of the filter with transfer function

$$\tilde{\Gamma}(\omega) = (1 - \exp(i\omega)) \sum_{j=-\infty}^{h-1} \gamma'_j \exp(-ij\omega)$$

$$
\begin{aligned}
\tilde{\gamma}_{h-1} &= \gamma'_{h-1} = \Gamma_h^-(0) \\
\tilde{\gamma}_{h-2} &= \gamma'_{h-2} - \gamma'_{h-1} = -\gamma_{h-1} \\
\tilde{\gamma}_{h-3} &= \gamma'_{h-3} - \gamma'_{h-2} = -\gamma_{h-2} \\
\tilde{\gamma}_{h-k} &= \gamma'_{h-k} - \gamma'_{h-k+1} = -\gamma_{h-k}
\end{aligned}
$$

Therefore

$$(1 - \exp(i\omega)) \sum_{j=-\infty}^{h-1} \gamma'_j \exp(-ij\omega) \;=\; \exp(-i(h-1)\omega)\Gamma_h^-(0) - \exp(i\omega)\Gamma_h^-(\omega)$$

which achieves the proof of the second assertion. Note that 141 is well-defined in frequency zero if $\Gamma(\omega)$ and thus $\Gamma_h^-(\omega)$ are sufficiently regular. For the last assertion we consider

$$\Gamma_{k_0}^{+'}(\omega) \;=\; \sum_{k \geq k_0} \gamma_k^{+'} \exp(-ik\omega)$$

78

Now

$$\sum_{k \geq k_0} |\gamma_k^{+'}||k|^\delta \leq \sum_{k \geq k_0} \left(\sum_{j \geq k} |\gamma_j| \right) |k|^\delta$$

$$= \sum_{j \geq k_0} |\gamma_j| \sum_{k=k_0}^{j} |k|^\delta$$

$$\leq \sum_{j \geq k_0} |\gamma_j||j - k_0||j|^\delta$$

$$\leq \sum_{j \geq k_0} |\gamma_j||j|^{1+\delta} + \sum_{j \geq k_0} |\gamma_j||k_0||j|^\delta$$

Since $\Gamma(\cdot) \in C_f^{1+\delta}$ both sums must be finite and therefore $\Gamma_{k_0}^{+'}(\cdot) \in C_f^\delta$ as was to be shown. A similar proof applies to $\Gamma_{k_0}^{-'}$.

The following proposition proposes a *discrete finite sample* convolution result which is at the origin of the super-consistency argument. It is a subtle technical - boring - result.

Proposition 3 *Let $Y_{it} = \sum_{k=-\infty}^{\infty} \gamma_{ik} X_{i,t-k}, i = 1, 2$ be the outputs of arbitrary (not necessarily symmetric) filters with transfer functions $\Gamma_i(\cdot), i = 1, 2$ and input series $X_{it}, i = 1, 2$. Define*

$$r_T := \frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} \Xi_{TY_1}(\omega_j)\overline{\Xi_{TY_2}(\omega_j)}$$

$$- \frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} \Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)} \tag{143}$$

If $X_{it} \in C_f^0, i = 1, 2$ then

$$E[|r_T|] = \begin{cases} o(1/\sqrt{T}) & if \ \Gamma_i(\cdot) \in C_f^{1/2}, i = 1, 2 \\ O(1/T) & if \ \Gamma_i(\cdot) \in C_f^1, i = 1, 2 \end{cases} \tag{144}$$

Proof:

Recall that we assume the sample size T to be an odd integer for notational simplicity (similar though slightly modified developments apply in the case of even T). Let

$$R_j(\omega) := \Xi_{TY_j}(\omega) - \Gamma(\omega)\Xi_{TX_j}(\omega)$$

For stationary $X_t \in C_f^0$ theorem 4.8 in Wildi (2005) shows that

$$E[|R_j(\omega)|^2] = O(1/T) \tag{145}$$

uniformly in j. Consider now the relevant error term

$$
\begin{aligned}
r_T &= \frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} \Xi_{TY_1}(\omega_j)\overline{\Xi_{TY_2}(\omega_j)} \\
&\quad -\frac{2\pi}{T} \sum_{j=(T-1)/2}^{(T-1)/2} \Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)} \\
&= \frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} (R_1(\omega_j)+\Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)) \\
&\quad \overline{(R_2(\omega_j)+\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j))} \\
&\quad -\frac{2\pi}{T} \sum_{j=(T-1)/2}^{(T-1)/2} \Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)} \\
&= \frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} R_1(\omega_j)\overline{R_2(\omega_j)} \\
&\quad +\frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} R_1(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)} \\
&\quad +\frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} \Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)\overline{R_2(\omega)}
\end{aligned}
\tag{146}
$$

Using 145 and the Cauchy-Schwartz-inequality, the first summand is of order $O(1/T)$ and can be neglected. We now show that the second summand is of order $o(1/\sqrt{T})$ (in absolute mean). A similar reasoning would apply to the third one by symmetry.

From

$$
\begin{aligned}
\Xi_{TY_j}(\omega) &= \frac{\sqrt{2\pi}}{\sqrt{T}} \sum_{t=1}^{T} Y_{jt}\exp(-it\omega) \\
&= \frac{\sqrt{2\pi}}{\sqrt{T}} \sum_{t=1}^{T} \left(\sum_{k=-\infty}^{\infty} \gamma_{jk}X_{j,t-k}\right)\exp(-it\omega) \\
&= \sum_{k=-\infty}^{\infty} \gamma_{jk}\exp(-ik\omega)\frac{\sqrt{2\pi}}{\sqrt{T}} \sum_{t=1}^{T} X_{j,t-k}\exp(-i(t-k)\omega) \\
&= \sum_{k=-\infty}^{\infty} \gamma_{jk}\exp(-ik\omega)\frac{\sqrt{2\pi}}{\sqrt{T}} \sum_{t=1-k}^{T-k} X_{jt}\exp(-it\omega)
\end{aligned}
$$

one deduces

$$
\begin{aligned}
R_j(\omega_n) \ := \ & \Xi_{TY_j}(\omega_n) - \Gamma(\omega_n)\Xi_{TX_j}(\omega_n) \\[4pt]
= \ & \sum_{k=-\infty}^{\infty} \gamma_{jk}\exp(-ik\omega_n) \\[4pt]
& \frac{\sqrt{2\pi}}{\sqrt{T}}\left(\sum_{t=1-k}^{T-k} X_{jt}\exp(-it\omega_n) - \sum_{t=1}^{T} X_{jt}\exp(-it\omega_n)\right) \\[4pt]
= \ & \sum_{k=-\infty}^{\infty} \gamma_{jk}\exp(-ik\omega_n)\frac{\sqrt{2\pi}}{\sqrt{T}} \\[4pt]
& \left[\sum_{t=1}^{\min(k,T)} \Big(X_{j,t-k}\exp(-i(t-k)\omega_n)\right. \\[4pt]
& \quad -X_{j,T+t-k}\exp(-i(T+t-k)\omega_n)\Big) \\[4pt]
& +\sum_{t=1}^{\min(-k,T)} \Big(X_{j,T+1-t-k}\exp(-i(T+1-t-k)\omega_n) \\[4pt]
& \quad \left.-X_{j,1-t-k}\exp(-i(1-t-k)\omega_n)\Big)\right] \\[4pt]
= \ & \sum_{k=-\infty}^{\infty} \gamma_{jk}\exp(-ik\omega_n)\frac{\sqrt{2\pi}}{\sqrt{T}} \\[4pt]
& \left[\sum_{t=1}^{\min(k,T)} \Big(X_{j,t-k}-X_{j,T+t-k}\Big)\exp(-i(t-k)\omega_n)\right. \\[4pt]
& \left.+\sum_{t=1}^{\min(-k,T)} \Big(X_{j,T+1-t-k}-X_{j,1-t-k}\Big)\exp(-i(1-t-k)\omega_n)\right]
\end{aligned}
$$

where the last equality follows from $T\omega_n$ being a multiple of 2π. We now analyze the error component 146.

$$\frac{2\pi}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} R_1(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)}$$

$$= \left(\frac{2\pi}{T}\right)^2 \sum_{j=-(T-1)/2}^{(T-1)/2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} I_{\{l>0\}}$$

$$\times \sum_{r=1}^{\min(l,T)} \sum_{t=1}^{T} \gamma_{2k}\gamma_{1l}X_{2t}\left(X_{1,r-l} - X_{1,T+r-l}\right)\exp(-i(r-l-k-t)\omega_j)$$

$$+ \left(\frac{2\pi}{T}\right)^2 \sum_{j=-(T-1)/2}^{(T-1)/2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} I_{\{l<0\}}$$

$$\times \sum_{r=1}^{\min(-l,T)} \sum_{t=1}^{T} \gamma_{2k}\gamma_{1l}X_{2t}\left(X_{1,T+1-r-l} - X_{1,1-r-l}\right)$$

$$\exp(-i(1-r-l-k-t)\omega_j)$$

where $I_{\{l>0\}}$ and $I_{\{l<0\}}$ are indicator functions. For k,l,r fixed let $s = r - l - k \mod(T)$ so that

$$\frac{1}{T} \sum_{j=-(T-1)/2}^{(T-1)/2} \exp(-i(r-l-k-t)\omega_j) = \begin{cases} 0 & t \neq s \\ 1 & \text{else} \end{cases}$$

Thus

$$E\left[\left|\frac{1}{T}\sum_{t=1}^{T} \gamma_{2k}\gamma_{1l}X_{2t}\left(X_{1,r-l} - X_{1,T+r-l}\right)\right.\right.$$

$$\left.\left.\sum_{j=-(T-1)/2}^{(T-1)/2} \exp(-i(r-l-k-t)\omega_j)\right|\right]$$

$$\leq 2|\gamma_{2k}\gamma_{1l}|\sqrt{E[X_{1t}^2]E[X_{2t}^2]}$$

and analogously

$$E\left[\left|\frac{1}{T}\sum_{t=1}^{T} \gamma_{2k}\gamma_{1l}X_{2t}\left(X_{1,T+1-r-l} - X_{1,1-r-l}\right)\right.\right.$$

$$\left.\left.\sum_{j=-(T-1)/2}^{(T-1)/2} \exp(-i(1-r-l-k-t)\omega_j)\right|\right]$$

$$\leq 2|\gamma_{2k}\gamma_{1l}|\sqrt{E[X_{1t}^2]E[X_{2t}^2]}$$

Therefore

$$
\begin{aligned}
E[|r_T|] \quad \leq \quad & E\left[\left|\frac{2\pi}{T}\sum_{j=-(T-1)/2}^{(T-1)/2} R_1(\omega_j)\overline{\Gamma_2(\omega_j)\Xi_{TX_2}(\omega_j)}\right|\right] \\
& +E\left[\left|\frac{2\pi}{T}\sum_{j=-(T-1)/2}^{(T-1)/2} \Gamma_1(\omega_j)\Xi_{TX_1}(\omega_j)\overline{R_2(\omega)}\right|\right] + O(1/T) \\
\leq \quad & \frac{4\pi^2}{T}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} 2|\gamma_{2k}\gamma_{1l}||\min(|l|,T)|\sqrt{E[X_{1t}^2]E[X_{2t}^2]} \\
& +\frac{4\pi^2}{T}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} 2|\gamma_{2k}\gamma_{1l}||\min(|k|,T)|\sqrt{E[X_{1t}^2]E[X_{2t}^2]} \\
& +O(1/T) \\
= \quad & \begin{cases} o(1/\sqrt{T}) & \text{if } \Gamma_i(\cdot) \in C_f^{1/2}, i=1,2 \\ O(1/T) & \text{if } \Gamma_i(\cdot) \in C_f^1, i=1,2 \end{cases}
\end{aligned}
$$

where the last equality - for $\Gamma_1(\cdot) \in C_f^{1/2}$ - follows from

$$
\begin{aligned}
& \frac{1}{T}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} |\gamma_{2k}\gamma_{1l}||\min(|l|,T)| \\
\leq \quad & \frac{1}{\sqrt{T}}\sum_{k=-\infty}^{\infty}|\gamma_{2k}|\sum_{l=-\infty}^{\infty}|\gamma_{1l}|\sqrt{|l|}\frac{\sqrt{\min(|l|,T)}}{\sqrt{T}} \\
= \quad & o(1/\sqrt{T})
\end{aligned}
$$

since $\Gamma_1(\cdot) \in C_f^{1/2}$ and $\lim_{T\to\infty}\frac{\sqrt{\min(|l|,T)}}{\sqrt{T}} = 0$ for each l (and analogously for $\Gamma_2(\cdot) \in C_f^{1/2}, C_f^1$).
This completes the proof of the proposition.

The following proposition analyzes the behaviour of the non-stationary optimization criterion 117 in the unit-root singularity.

Proposition 4

- *If $\Delta\Gamma(0) \neq 0$, then:*

$$
\begin{aligned}
& \lim_{\omega\to 0}\Bigg\{ -\exp(i\omega)\left[\frac{\exp(ih\omega)\Delta\Gamma_{-h}^{0+}(0) - \exp(-i\omega)\Delta\Gamma_{-h}^{0+}(\omega)}{1-\exp(-i\omega)} - \exp(ih\omega)\Delta\Gamma_{-h}^{0+}(0)\right] \\
& +\frac{\exp(-i(-h-1)\omega)\Gamma_{-h}^{0-}(0) - \exp(i\omega)\Gamma_{-h}^{0-}(\omega)}{1-\exp(i\omega)}\Bigg\} \\
= \quad & \sum_{k=-h}^{L-h} b_k^0 k - h\Delta\Gamma(0) \\
= \quad & \hat{A}_X(0)\Big(sign[\hat{\Gamma}_X(0)]h + \hat{\phi}_X(0)\Big) - h\Gamma(0) \quad \text{if } \hat{A}_X(0) > 0
\end{aligned}
$$

where $\hat{A}_X(0)$ and $\hat{\phi}_X(0)$ are the amplitude and the time-shift functions of $\hat{\Gamma}_X(\omega)$ in frequency zero and where sign$[\cdot]$ is the signum function (1 or -1 depending on the sign of its argument).

- If $\Delta\Gamma(0) = 0$, then:

$$\lim_{\omega \to 0} \left\{ -\exp(i\omega) \left[\frac{\exp(ih\omega)\Delta\Gamma^{0+}_{-h}(0) - \exp(-i\omega)\Delta\Gamma^{0+}_{-h}(\omega)}{1 - \exp(-i\omega)} - \exp(ih\omega)\Delta\Gamma^{0+}_{-h}(0) \right] \right.$$
$$\left. + \frac{\exp(-i(-h-1)\omega)\Gamma^{0-}_{-h}(0) - \exp(i\omega)\Gamma^{0-}_{-h}(\omega)}{1 - \exp(i\omega)} \right\}$$
$$= \sum_{k=-h}^{L-h} b^0_k k$$
$$= \hat{A}_X(0)\hat{\phi}_X(0) \quad \text{if } \hat{A}_X(0) > 0$$

- if $\hat{\Gamma}_{W_n}(0) \neq 0$ then

$$\lim_{\omega \to 0} \exp(i\omega_k) \left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) \right]$$
$$= \sum_{k=-h}^{L-h} b^n_k k + h\hat{\Gamma}_{W_n}(0)$$
$$= \hat{A}_{W_n}(0)\Big(sign\big[\hat{\Gamma}_{W_n}(0)\big]h + \hat{\phi}_{W_n}(0) \Big)$$

Proof:

$$\lim_{\omega \to 0} -\exp(i\omega) \left[\frac{\exp(ih\omega)\Delta\Gamma^{0+}_{-h}(0) - \exp(-i\omega)\Delta\Gamma^{0+}_{-h}(\omega)}{1 - \exp(-i\omega)} - \exp(ih\omega)\Delta\Gamma^{0+}_{-h}(0) \right]$$
$$= -\frac{ih\Delta\Gamma^{0+}_{-h}(0) + i\Delta\Gamma^{0+}_{-h}(0) - \Delta\dot{\Gamma}^{0+}_{-h}(0)}{i} + \Delta\Gamma^{0+}_{-h}(0)$$
$$= -h\Delta\Gamma^{0+}_{-h}(0) - \sum_{k=-h}^{\infty} (\gamma_k - b^0_k)k \tag{147}$$

where $\Delta\dot{\Gamma}^{0+}_{-h}(0) = -i\sum_{k \geq -h}(\gamma_k - b^0_k)k$ is the derivative of the transfer function in frequency zero. Note that we used de l'Hopital's rule to when passing to the second line (taking derivatives above and below the quotient line i.e. we use first order Taylor approximations to get the asymptotic true values). Analogously we obtain

$$\lim_{\omega \to 0} \frac{\exp(-i(-h-1)\omega)\Gamma^{0-}_{-h}(0) - \exp(i\omega)\Gamma^{0-}_{-h}(\omega)}{1 - \exp(i\omega)}$$
$$= \frac{i(h+1)\Delta\Gamma^{0-}_{-h}(0) - i\Delta\Gamma^{0-}_{-h}(0) - \Delta\dot{\Gamma}^{0-}_{-h}(0)}{-i}$$
$$= -h\Delta\Gamma^{0-}_{-h}(0) - \sum_{k=-\infty}^{-h-1} (\gamma_k - b^0_k)k \tag{148}$$

Summing 147 and 148 gives

$$-h\Delta\Gamma^0(0) - \sum_{k=-\infty}^{\infty} (\gamma_k - b_k^0)k = -h\Delta\Gamma^0(0) + \sum_{k=-h}^{L-h} b_k^0 k$$

where $b_k^0 = 0$ if $k < -h$ or $k > L - h$ and where we used $\sum_{k=-\infty}^{\infty} \gamma_k k = 0$ by symmetry of $\Gamma(\cdot)$. If $\hat{A}_X(0) > 0$ then

$$\begin{aligned}
\sum_{k=-h}^{L-h} b_k^0 k &= \frac{\frac{d}{d\omega}\hat{\Gamma}_X(\omega)\big|_{\omega=0}}{-i} \\
&= \frac{\dot{\hat{A}}_X(0)\exp^{-i\hat{\Phi}_X(0)} - i\hat{A}_X(0)\dot{\hat{\Phi}}_X(0)\exp^{-i\hat{\Phi}_X(0)}}{-i} \\
&= \hat{A}_X(0)\dot{\hat{\phi}}_X(0)
\end{aligned}$$

where the derivative of the phase in frequency zero is the time-shift of $\hat{\Gamma}_X(\cdot)$: $\dot{\hat{\Phi}}_X(0) = \dot{\hat{\phi}}_X(0)$. Note that the derivative of the amplitude in frequency zero vanishes, $\dot{\hat{A}}_X(0) = 0$, because the amplitude function is even and differentiable in $\omega = 0$, if $\hat{A}_X(0) > 0$. Now

$$\begin{aligned}
-h\Delta\Gamma^0(0) + \sum_{k=-\infty}^{\infty} b_k^0 k &= -h\Gamma(0) + h\hat{\Gamma}_X(0) + \hat{A}_X(0)\dot{\hat{\phi}}_X(0) \\
&= -h\Gamma(0) + \hat{A}_X(0)\left(\text{sign}\big[\hat{\Gamma}_X(0)\big]h + \dot{\hat{\phi}}_X(0)\right)
\end{aligned}$$

which achieves the proof of the first two claims. The third assertion follows from

$$\lim_{\omega \to 0} \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\right]$$

$$= \frac{ih\hat{\Gamma}_{W_n}(0) + i\hat{\Gamma}_{W_n}(0) - \dot{\hat{\Gamma}}_{W_n}(0)}{i} - \hat{\Gamma}_{W_n}(0)$$

$$= h\hat{\Gamma}_{W_n}(0) + \sum_{k=-h}^{L-h} b_k^n k \tag{149}$$

If $\hat{A}_{W_n}(0) > 0$ then

$$\begin{aligned}
\sum_{k=-h}^{L-h} b_k^n k &= \frac{\frac{d}{d\omega}\hat{\Gamma}_{W_n}(\omega)\big|_{\omega=0}}{-i} \\
&= \frac{\dot{\hat{A}}_{W_n}(0)\exp^{-i\hat{\Phi}_{W_n}(0)} - i\hat{A}_{W_n}(0)\dot{\hat{\Phi}}_{W_n}(0)\exp^{-i\hat{\Phi}_{W_n}(0)}}{-i} \\
&= \hat{A}_{W_n}(0)\dot{\hat{\phi}}_{W_n}(0)
\end{aligned}$$

Therefore

$$\begin{aligned}
\sum_{k=-h}^{L-h} b_k^n k + h\hat{\Gamma}_{W_n}(0) &= h\hat{\Gamma}_{W_n}(0) + \hat{A}_{W_n}(0)\dot{\hat{\phi}}_{W_n}(0) \\
&= \hat{A}_{W_n}(0)\left(\text{sign}\big[\hat{\Gamma}_{W_n}(0)\big]h + \dot{\hat{\phi}}_{W_n}(0)\right)
\end{aligned}$$

which achieves the proof of the proposition.

The following proposition derives the desirable super-consistency argument.

Proposition 5 *Let $\Xi_{Tr}(\omega_k)$ designate the (generally unobserved) DFT of the filter error.*

- *Let $\Xi'_{Tr}(\omega_k)$ be defined by 112. If $\Gamma(\cdot), \hat{\Gamma}_X(\cdot), \hat{\Gamma}_{W_n}(\cdot) \in C_f^{1.5}$, $n = 1, ..., m$, and $X_t, W_{1t}, ..., W_{mt}$ are integrated I(1)-series and if $\Delta X_t, \Delta W_{nt} \in C_f^0$, $n = 1, ..., m$ (first differences have an absolutely summable Wold-domposition), then*

$$\frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} w_k \left|\Xi_{Tr}(\omega_k)\right|^2 = \frac{2\pi}{T} \sum_{k=-[T/2]}^{[T/2]} \left|\Xi'_{Tr}(\omega_k)\right|^2 + r_T$$

where the approximation error r_T is unusually small (smaller than the natural sampling error) in absolute mean:

$$E[|r_T|] = o(1/\sqrt{T})$$

- *If, instead, the data is stationary with absolutely summable Wold-decompositions, $X_t, W_{nt} \in C_f^0$, $n = 1, ..., m$, then the above holds true if the DFT 112 (non-stationary case) is replaced by the DFT $\Xi''_{Tr}(\omega_k)$ 114 (stationary case).*

Proof:

The universal (tautological) decomposition of the filter error r_t in section 11.1 led to the following time-domain expression:

$$
\begin{aligned}
r_t =\ & C'_{t+h} \\
& - \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty}(\gamma_k - b_k^0)\right](X_{t+1-j} - X_{t-j}) \\
& + \Delta\Gamma_h^+(0)(X_{t+1+h} - X_{t+h}) \\
& + \sum_{j=-\infty}^{-h-1} \left[\sum_{k=-\infty}^{j}(\gamma_k - b_k^0)\right](X_{t-j} - X_{t-j-1}) \\
& + \sum_{n=1}^{m} \left\{ \sum_{j=-h}^{\infty} \left[\sum_{k=j}^{\infty} b_k^i\right](W_{n,t+1-j} - W_{n,t-j}) \right. \\
& \left. - \hat{\Gamma}_{W_n}(0)(W_{n,t+1+h} - W_{n,t+h}) \right\}
\end{aligned}
$$

see 108. Since the DFT is linear we can apply it to each of the above components separately. Lets re-label the above components:

$$\Xi_0(\omega_k) := DFT\left(C'_{t+h} + \Delta\Gamma_h^+(0)(X_{t+1+h} - X_{t+h}) - \hat{\Gamma}_{W_n}(0)(W_{n,t+1+h} - W_{n,t+h})\right)(\omega_k)$$

$$\Xi_1(\omega_k) := DFT\left(-\sum_{j=-h}^{\infty}\left[\sum_{k=j}^{\infty}(\gamma_k - b_k^0)\right](X_{t+1-j} - X_{t-j})\right)(\omega_k)$$

$$\Xi_2(\omega_k) := DFT\left(\sum_{j=-\infty}^{-h-1}\left[\sum_{k=-\infty}^{j}(\gamma_k - b_k^0)\right](X_{t-j} - X_{t-j-1})\right)(\omega_k)$$

$$\Xi_3(\omega_k) := DFT\left(\sum_{n=1}^{m}\left\{\sum_{j=-h}^{\infty}\left[\sum_{k=j}^{\infty}b_k^i\right](W_{n,t+1-j} - W_{n,t-j})\right\}\right)(\omega_k)$$

Now consider 112

$$\Xi'_{Tr}(\omega_k) := \exp(ih\omega_k)\Xi_{TC'}(\omega_k)$$
$$- \exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Delta\Gamma_{-h}^{0+}(0) - \exp(-i\omega_k)\Delta\Gamma_{-h}^{0+}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\Delta\Gamma_{-h}^{0+}(0)\right]\Xi_{T\Delta X}(\omega_k)$$
$$+ \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma_{-h}^{0-}(0) - \exp(i\omega_k)\Gamma_{-h}^{0-}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k)$$
$$+ \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)} - \exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\right]\Xi_{T\Delta W_n}(\omega_k)$$

and let us relabel the corresponding DFT's:

$$\Xi'_0(\omega_k) := \exp(ih\omega_k)\Xi_{TC'}(\omega_k) + \exp(i\omega_k)\exp(ih\omega_k)\Delta\Gamma_{-h}^{0+}(0)\Xi_{T\Delta X}(\omega_k)$$
$$- \exp(i\omega_k)\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0)\Xi_{T\Delta W_n}(\omega_k)$$

$$\Xi'_1(\omega_k) := -\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\Delta\Gamma_{-h}^{0+}(0) - \exp(-i\omega_k)\Delta\Gamma_{-h}^{0+}(\omega_k)}{1 - \exp(-i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k)$$

$$\Xi'_2(\omega_k) := \left[\frac{\exp(-i(-h-1)\omega_k)\Gamma_{-h}^{0-}(0) - \exp(i\omega_k)\Gamma_{-h}^{0-}(\omega_k)}{1 - \exp(i\omega_k)}\right]\Xi_{T\Delta X}(\omega_k)$$

$$\Xi'_3(\omega_k) := \sum_{n=1}^{m}\exp(i\omega_k)\left[\frac{\exp(ih\omega_k)\hat{\Gamma}_{W_n}(0) - \exp(-i\omega_k)\hat{\Gamma}_{W_n}(\omega_k)}{1 - \exp(-i\omega_k)}\right]\Xi_{T\Delta W_n}(\omega_k)$$

Note that $\Xi'_j(\omega_k)$ corresponds to the previous $\Xi_j(\omega_k)$ for $j = 0, 1, 2, 3$. Then

$$\frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}w_k\,|\Xi_{Tr}(\omega_k)|^2 = \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}w_k\left(\sum_{j=0}^{3}\Xi_j(\omega_k)\overline{\sum_{j=0}^{3}\Xi_j(\omega_k)}\right)$$

$$= \frac{2\pi}{T}\sum_{k=-[T/2]}^{[T/2]}w_k\left(\sum_{j=0}^{3}\Xi'_j(\omega_k)\overline{\sum_{j=0}^{3}\Xi'_j(\omega_k)}\right) + r_T$$

If $\Gamma(\cdot), \hat{\Gamma}_X(\cdot), \hat{\Gamma}_{W_n}(\cdot) \in C_f^{1.5}$, as assumed, then proposition 2, third claim, implies that the transfer functions in the definitions of $\Xi_j'(\omega_k)$ are all in $C_f^{1/2}$. Then proposition 3 implies that

$$E[|r_T|] = o(1/\sqrt{T})$$

which show the first claim (note that we can neglect the effect by the weights w_k which is of smaller magnitude $O(1/T)$). The second claim follows along a similar argumentation, which achieves the proof of the proposition.

The following proposition shows that the sample-mean square filter error is an asymptotically efficient estimate of the unkown (true) mean-square filter error.

Proposition 6 *Assume either that*

1. *all processes are in C_f^0 (stationary) or that*

2. *the cointegration equilibrium C_t' as well as the differenced (stationary) processes $\Delta X_t, \Delta W_{ht}$, $h = 1, ..., m$ are in C_f^0.*

Furthermore we require $\Gamma(\cdot) \in C_f^{1+1/2}$. Assume also that the innovation-sequences $\epsilon_{Xt}, \epsilon_{W_k t}, k = 1, ..., m$ of the input series $X_t, W_{kt}, k = 1, ..., m$ are longitudinally iid with finite fourth order moments. In the cross-section we assume that there exist shifts $\delta_1, ..., \delta_m$ such that the shifted innovations $\epsilon_{Xt_0}, \epsilon_{W_k, t_k+\delta_k}, k = 1, ..., m$ may be (contemporaneously) correlated for $t_j = t_k$ and mutually iid for $t_j \neq t_k$, $j, k = 0, ..., m$. Assume also that the solutions $\hat{\Gamma}_X(\cdot), \hat{\Gamma}_{W_k}(\cdot), k = 1, ..., m$ of 137 define a uniformly stable vector-sequence i.e. that each univariate filter-sequence is uniformly stable as a function of the sample size T. Then the sample moment

$$\frac{1}{T} \sum_{t=1}^{T} r_t^2$$

is an asymptotically best linear unbiased estimate of the revision error variance $E[r_t^2]$.

Note that the above time shifts δ_k allow for leads or lags of the explaining variables W_{kt}. The cross-sectional independency requirements in the above proposition could be generalized if necessary. We require them here in order to make the proof more readable. All the independence requirements could be generalized to moment properties of order two, three and four which could be verified empirically (details are omitted here).

Proof

We here focus on the case of cointegration. A similar proof would apply to the stationary case. Consider Z_t as given in **??**

$$r_t = \Delta \Gamma_X(0) C_t + Z_t$$

The proof of theorem **??** below shows that $Z_t \in C_f^0$. Thus if $C_t \in C_f^0$ (by assumption) we conclude $r_t \in C_f^0$. Using the cross-sectional dependency structure of the multivariate innovation series we may then rearrange (shift) innovations such that r_t can be written as

$$
\begin{aligned}
r_t &= \sum_{k=-\infty}^{\infty} \left(\gamma'_{Xk} \epsilon_{X,t-k} + \sum_{j=1}^{m} \hat{\gamma}'_{W_j,k+\delta_j} \epsilon_{W_j,t-k} \right) \\
&= \sum_{k=-\infty}^{\infty} \tilde{\gamma}_k \tilde{\epsilon}_{t-k}
\end{aligned}
$$

where $\tilde{\epsilon}_{t-k}$ are iid innovations (which may assumed to be standardized). Together with $r_t \in C_f^0$ all assumptions of proposition E.2 in Wildi [**?**] are satisfied. The latter result states that the squared process r_t^2 has an absolutely summable autocovariance function under the given circumstances. Therefore its spectral density function exists and is continuous. Grenander and Rosenblatt [**?**], section 7.3, then show that $\overline{r_t^2} := \frac{1}{T} \sum_{t=1}^{T} r_t^2$ is an asymptotically best linear unbiased estimate of $E[r_t^2]$.

References

[1] Gyomai G. and Wildi, M. (2013) OECD Composite Leading Indicators for G7 Countries: a Comparison of the Hodrick-Prescott Filter and the Multivariate Direct Filter Approach. *OECD: STD working paper series 2012(5), published on 16th January 2013.* http://blog.zhaw.ch/idp/sefblog/index.php?/archives/340-I-MDFA-and-OECD-CLIs.html

[2] McElroy T and Wildi, M. (2013) On a Trilemma Between Accuracy, Timeliness and Smoothness in Real-Time Forecasting and Signal Extraction. *Soon on SEFBlog.*

[3] Taniguchi, M. and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

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

[5] 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.

[6] 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.

[7] 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

[8] Wildi, M. (2008.2) Efficient Multivariate Real-Time Filtering and Cointegration. *IDP-working paper, IDP-WP-08Sep-01.* http://blog.zhaw.ch/idp/sefblog/uploads/multivar_dfa1.pdf

[9] 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

[10] 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

[11] Wildi, M. (2011) A Unified Framework to real-Time Signal Extraction and Data Revisions. Working paper: http://blog.zhaw.ch/idp/sefblog/index.php?/archives/163-A-Unified-Framework-to-Real-Time-Signal-Extraction-RTSE-and-Data-Revisions.html