JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **36**: 209–227 (2015)
Published online 3 November 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12102

ORIGINAL ARTICLE

SIGNAL EXTRACTION FOR NON-STATIONARY MULTIVARIATE TIME SERIES WITH ILLUSTRATIONS FOR TREND INFLATION

TUCKER MCELROY[a] AND THOMAS TRIMBUR[b,c]*

[a] *US Census Bureau, Washington, DC, USA*
[b] *Federal Reserve Board, Washington, DC, USA*
[c] *Johns Hopkins University, Baltimore, MD, USA*

This article advances the theory and methodology of signal extraction by developing the optimal treatment of difference stationary multivariate time-series models. Using a flexible time-series structure that includes co-integrated processes, we derive and prove formulas for minimum mean square error estimation of signal vectors in multiple series, from both a finite sample and a bi-infinite sample. As an illustration, we present econometric measures of the trend in total inflation that make optimal use of the signal content in core inflation.

Received 7 September 2013; Revised 12 June 2014; Accepted 10 September 2014

Keywords: co-integration; common trends; filters; multivariate models; stochastic trends; unobserved components.

1. INTRODUCTION

In many scientific fields, research and analysis make widespread use of a signal extraction paradigm. Often, interest centres on underlying dynamics (such as trend, seasonal, and cyclical parts) of time series also subject to less regular components representing temporary fluctuations. In such cases, the resulting strategy involves the estimation of signals in the presence of noise. For instance, economists and policymakers routinely want to assess major price trends, cycles in real activity, and other pivotal indicators of economic performance. Another application is seasonal adjustment – a major undertaking for statistical agencies – wherein the noise is the seasonal component and the signal consists of the non-seasonal components, such as trend and cycle.

Here, we concentrate on developing the formal apparatus for detecting signals in a comprehensive statistical framework, motivated by two basic considerations. First, the signal extraction problems relevant to experience usually involve more than one variable; at central banks, for instance, staff use a range of available series to monitor prevailing inflationary conditions. Second, economic data often involve non-stationary movements, with (possibly close) statistical relationships among the stochastic trends (or seasonals) for a set of indicators. In this article, we generalize the existing theory and methodology of signal extraction to multivariate difference stationary time series, handling both the asymptotic and finite-sample cases and introducing the treatment of co-integrated systems.

While the co-integration literature expanded rapidly following initial advances (e.g. Engle and Granger, 1987; Johansen, 1988; Stock and Watson, 1988), the problem of designing optimal estimators of common trends has not yet been solved. Here, we show how the co-integrated structure has special implications for the characteristics of the signal estimators. For instance, it follows from our formulae (Theorem 1) that the trend estimator must satisfy the same integration relations as the true signal; such a finding, while plausible and aesthetically attractive, remained elusive in previous work, despite the immense effort devoted to studying common trend specifications.

* Correspondence to: Thomas Trimbur, Johns Hopkins University, Baltimore, MD, USA. E-mail: ttrimbu1@jhu.edu

Our main new result is the formula for the finite-sample signal extraction matrix, for difference stationary vector time series; this generalizes the univariate results of McElroy (2008) to the multivariate case. The bi-infinite-sample case was first treated by Wiener (1949) and Whittle (1963) for stationary series and extended by Bell (1984) to the difference stationary case. Bell (1984) focuses on univariate series and, although explicit multivariate expressions are not derived or proven, he does mention in a remark that the same methods can be extended to the multivariate case; however, collinearity of signal innovations is not explicitly mentioned or addressed therein. Our Theorem 1 extends existing results to difference stationary time series that are possibly non-invertible (corresponding to co-integration in the raw data). For finite samples, the difference stationary univariate case was treated by Bell and Hillmer (1988), Pollock (2007), and McElroy (2008). The stationary multivariate case, restricted to processes embeddable in a state space (SS) framework (excluding long-memory time series), was treated by Gómez (2006). Theorem 2 gives the general treatment of difference stationary multivariate processes, for a finite sample; in the case of a process that is embeddable in a linear SS framework, our direct formulae provide the same signal extraction results as obtained from a properly initialized (see Bell and Hillmer, 1991, for discussion) SS smoother.

Many of the economic applications involving the extraction of signals (e.g. the multivariate trend–cycle analyses of Harvey and Trimbur, 2003, and Basistha and Startz, 2008) have been undertaken without the rigorous foundation provided in this article and have relied on standard recursive filtering and smoothing algorithms. From a computational perspective, our framework is complementary to linear SS methods; our matrix expressions give a different, direct route for computing multivariate signals that is more straightforward conceptually but ultimately produces the same results. The focus of this article is methodological, presenting a mathematical basis for signal extraction applicable to widely used multivariate difference stationary models.

The rest of the article is arranged as follows. Section 2 develops the generalized Wiener–Kolmogorov (WK) formula for a set of difference stationary time series, expressing the optimal multivariate filters in both the frequency and time domains. Exact signal extraction results for a finite-length dataset are derived in Section 3. Then Section 4 reviews some major models for multivariate stochastic trends, and the methodology is illustrated by the statistical measurement of trend inflation using both core and total inflation data. Section 5 provides our conclusions, and mathematical proofs are in the Appendix.

2. MULTIVARIATE SIGNAL EXTRACTION FROM A BI-INFINITE SAMPLE

This section gives the solution to the signal extraction problem for multiple difference stationary series (i.e. series that are stationary after the application of a differencing polynomial) and generalizes the WK formula. Theorem 1 provides a generalization of the results of Bell (1984), which solves the univariate problem and goes on to claim that the same formulae apply verbatim to the multivariate case as well. We also explicitly address the presence of collinearity through Proposition 1 and the subsequent discussion.

Consider a vector-valued difference stationary time series denoted $\{\mathbf{y}_t\} = \{\mathbf{y}_t, t = 0, \pm 1, \pm 2, \ldots\}$, with each $\mathbf{y}_t$ of dimension N. A multivariate filter for a set of N series has the expression

$$\boldsymbol{W}(L) = \sum_{h=-\infty}^{\infty} \boldsymbol{W}_h L^h, \tag{1}$$

where L is the standard lag operator and $\boldsymbol{W}_h$ is the $N \times N$ matrix of coefficients for lag h. The filter produces output $\{\mathbf{z}_t\}$ as follows:

$$\mathbf{z}_t = \boldsymbol{W}(L)\mathbf{y}_t = \sum_{h=-\infty}^{\infty} \boldsymbol{W}_h L^h \mathbf{y}_t = \sum_{h=-\infty}^{\infty} \boldsymbol{W}_h \mathbf{y}_{t-h}. \tag{2}$$

Therefore, the weight matrix $\boldsymbol{W}_h$ at lag h is applied to the lagged series $\mathbf{y}_{t-h}$. In the case that $\{\mathbf{y}_t\}$ is stationary, recall that its spectral density $\boldsymbol{F}$ is related to the spectral density of the output process $\{\mathbf{z}_t\}$ via the formula

DOI: 10.1111/jtsa.12102

$\boldsymbol{W}(\mathrm{e}^{-i\lambda})\boldsymbol{F}(\lambda)\boldsymbol{W}'(\mathrm{e}^{i\lambda})$. When $\boldsymbol{W}_{-h} = \boldsymbol{W}_h$ for all h, the frequency response function (FRF) $\boldsymbol{W}(\mathrm{e}^{-i\lambda})$ is real valued and is identical with the gain function, denoted $\boldsymbol{G}$; that is, $\boldsymbol{G}(\lambda) = \boldsymbol{W}(\mathrm{e}^{-i\lambda})$.

Recall that the spectral density $\boldsymbol{F}$ is related to the autocovariances $\boldsymbol{\Gamma}_h = E(\boldsymbol{y}_t \boldsymbol{y}'_{t-h})$ of stationary $\{\mathbf{y}_t\}$ via $\boldsymbol{F}(\lambda) = \sum_h \boldsymbol{\Gamma}_h \mathrm{e}^{-i\lambda h}$ (a Hermitian matrix), and the multivariate autocovariance generating function (ACGF) is given by substituting L for $\mathrm{e}^{-i\lambda}$, so that $\boldsymbol{\Gamma}_{\boldsymbol{y}}(L) = \sum_h \boldsymbol{\Gamma}_h L^h$. The basic aim of signal extraction is to estimate a target signal $\{\boldsymbol{s}_t\}$ from a series $\{\mathbf{y}_t\}$ of interest, where for all integer t

$$\mathbf{y}_t = \boldsymbol{s}_t + \boldsymbol{n}_t, \tag{3}$$

and $\{\boldsymbol{s}_t\}$ and $\{\boldsymbol{n}_t\}$ have dimension $N \times 1$ and are uncorrelated with one another. In the two-component decomposition (3), the signal $\{\boldsymbol{s}_t\}$ can represent either a single component of interest or a combination of components, with the noise $\{\boldsymbol{n}_t\}$ absorbing the remaining components of $\mathbf{y}_t$. For example, a signal could refer to the sum of latent trend and cycle components in $\{\mathbf{y}_t\}$, while the noise might be the latent seasonal and idiosyncratic components. We emphasize that the noise processes considered here are not 'white' but could be non-stationary (e.g. a seasonal component).

In this article, we compute the Gaussian conditional expectation $\mathbb{E}[\mathbf{s}_t|\{\mathbf{y}_t\}]$, which will also be the minimum mean squared error (MSE) *linear* estimator of $\mathbf{s}_t$ in the non-Gaussian case. This is our notion of optimality, and we denote the estimate by $\hat{\mathbf{s}}_t$. For stationary series $\boldsymbol{s}_t$ and $\boldsymbol{n}_t$, the optimal estimate is given by $\boldsymbol{W}_{WK}(L)\mathbf{y}_t$, where the WK filter (Wiener, 1949) has FRF

$$\boldsymbol{W}_{WK}(e^{-i\lambda}) = \boldsymbol{\Gamma}_{\boldsymbol{s}}(\mathrm{e}^{-i\lambda})\left[\boldsymbol{\Gamma}_{\boldsymbol{s}}(\mathrm{e}^{-i\lambda}) + \boldsymbol{\Gamma}_{\boldsymbol{n}}(\mathrm{e}^{-i\lambda})\right]^{-1}. \tag{4}$$

Here, $\boldsymbol{\Gamma}_{\boldsymbol{s}}(\mathrm{e}^{-i\lambda})$ and $\boldsymbol{\Gamma}_{\boldsymbol{n}}(\mathrm{e}^{-i\lambda})$ are the multivariate spectral densities of signal and noise respectively.

For the multivariate signal and noise processes, we consider all difference stationary processes, including vector autoregressive integrated moving average as a special case. For the jth observed process, $\{\mathbf{y}_t^{(j)}\}$, we assume there exists a unit root polynomial $\delta^{(j)}(L)$ of minimal degree d^j, such that $\{\mathbf{w}_t^{(j)}\} = \{\delta^{(j)}(L)\mathbf{y}_t^{(j)}\}$ is covariance stationary. Similarly, $\{\mathbf{u}_t^{(j)}\} = \{\delta_{\mathbf{s}}^{(j)}(L)\mathbf{s}_t^{(j)}\}$ and $\{\mathbf{v}_t^{(j)}\} = \{\delta_{\mathbf{n}}^{(j)}(L)\mathbf{n}_t^{(j)}\}$ for appropriately chosen differencing polynomials for the jth signal and noise. We assume that $\delta^{(j)} = \delta_{\mathbf{n}}^{(j)}\delta_{\mathbf{s}}^{(j)}$ for each j and that no unit roots are shared between $\delta_{\mathbf{n}}^{(j)}$ and $\delta_{\mathbf{s}}^{(j)}$ (this assumption is not restrictive to the bulk of applications). As a special and leading case, it may occur that the signal and noise differencing operators do not depend on j, so that they are the same for each series (although they still differ for signal vs noise); we refer to this situation as 'uniform differencing operators'.

The identification of these differencing polynomials is taken as a given in all applications, in the sense that practitioner selection of these differencing polynomials constitutes a choice of signals that are of interest. The only practical restriction is that $\delta^{(j)}(L)$ applied to $\{\mathbf{y}_t^{(j)}\}$ be stationary. For example, if one is interested in a random-walk trend extraction application from a trend–seasonal–irregular decomposition, then $\delta_{\mathbf{s}}^{(j)}(L) = 1 - L$ for each j, while $\delta_{\mathbf{n}}^{(j)}(L) = 1 + L + L^2 + L^3$ (for quarterly data). If, on the other hand, one wants to capture the 'first seasonal' (i.e. seasonal effects that occur only once a year, not twice), then $\delta_{\mathbf{s}}^{(j)}(L) = 1 + L^2$ and the noise consists of a random-walk trend plus irregular plus the 'second seasonal', so that $\delta_{\mathbf{n}}^{(j)}(L) = (1 - L)(1 + L) = 1 - L^2$.

Let $\boldsymbol{F}_{\mathbf{u}}$, $\boldsymbol{F}_{\mathbf{v}}$, and $\boldsymbol{F}_{\mathbf{w}}$ denote the matrix spectral density functions for the differenced signal, noise, and observed processes respectively. We suppose that $\boldsymbol{F}_{\mathbf{w}}$ is invertible almost everywhere, that is, the set Λ of frequencies where $\boldsymbol{F}_{\mathbf{w}}$ is non-invertible has Lebesgue measure zero. Note that if the data process is co-integrated, then $\boldsymbol{F}_{\mathbf{w}}(0)$ is singular, although $\boldsymbol{F}_{\mathbf{w}}(\lambda)$ remains invertible for $\lambda \neq 0$. Next, define the 'over-differenced' processes:

$$\partial\mathbf{u}_t^{(j)} = \delta^{(j)}(L)\mathbf{s}_t^{(j)} = \delta_{\mathbf{n}}^{(j)}(L)\mathbf{u}_t^{(j)}$$

$$\partial\mathbf{v}_t^{(j)} = \delta^{(j)}(L)\mathbf{n}_t^{(j)} = \delta_{\mathbf{s}}^{(j)}(L)\mathbf{v}_t^{(j)}.$$

DOI: 10.1111/jtsa.12102

These occur when the data differencing polynomial $\delta^{(j)}(L)$ is applied to signal and noise respectively, resulting in covariance stationary processes that may have zeroes in their spectral densities.

As part of the model specification and identification, $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$ are assumed to be uncorrelated with one another, which is a reasonable condition when signal and noise are driven by distinct kinds of factors; for instance, a long-run trend might be little affected by temporary forces behind noisy fluctuations. This correlation structure is commonplace in the literature, and it has the attractive and intuitive property that tight model linkages, like common trends, give rise to symmetric filter structures.

Note that each difference stationary process $\{\mathbf{y}_t^{(j)}\}$ can be generated from d^j stochastic initial values $\mathbf{y}_*^{(j)}$ together with the disturbance process $\{\mathbf{w}_t^{(j)}\}$, for each j. The information contained in $\{\mathbf{y}_t^{(j)}\}$ is equivalent to that in $\{\mathbf{w}_t^{(j)}\} \cup \mathbf{y}_*^{(j)}$ for the purposes of linear projection, since the former is expressible as a linear transformation of the latter, for each j. In model fitting and forecasting applications, a working assumption for vector time series is that these initial values $\mathbf{y}_*^{(j)}$ are uncorrelated with the disturbance process $\{\mathbf{w}_t^{(j)}\}$; we will assume a stronger condition that actually implies this assumption.

Assumption M_∞. Suppose that, for each $j = 1, 2, \ldots, N$, the initial values $\mathbf{y}_*^{(j)}$ are uncorrelated with the differenced signal and noise processes $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$.

This assumption generalizes the univariate Assumption A of Bell (1984) to a multivariate framework – each set of initial values $\mathbf{y}_*^{(j)}$ is orthogonal not only to the differenced signal and noise processes for the jth series but also for all N series. (Other types of assumptions are possible, but as argued by Bell, 1984, the framework of Assumption A is the most appealing.) Set $z = \mathrm{e}^{-i\lambda}$ and $\overline{z} = \mathrm{e}^{i\lambda}$. For any matrix A, the matrix consisting of only the diagonal entries is denoted by $\tilde{A}$. Then for difference stationary (and possibly co-integrated) multivariate time series, the optimal estimator of the signal, conditional on the observations $\{\mathbf{y}_t\}$, for each j and at each time t, is given by a multivariate filter $\boldsymbol{W}(L)$ described in the following.

Theorem 1. Assume that $\boldsymbol{F}_{\mathbf{w}}(\lambda)$ is invertible for each λ in a subset $\Lambda \subset [-\pi, \pi]$ of full Lebesgue measure. Also suppose $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$ are uncorrelated with one another and Assumption M_∞ holds. Let $\widetilde{\boldsymbol{\delta}}(z)$ denote the diagonal matrix with entries $\delta^{(j)}(z)$, with similar notation for $\widetilde{\boldsymbol{\delta}}_{\mathbf{s}}(z)$ and $\tilde{\boldsymbol{\delta}}_{\mathbf{n}}(z)$. With the quantity

$$\mathbf{B}(z) = \tilde{\boldsymbol{\Gamma}}_{\mathbf{v}}(z)\, \tilde{\boldsymbol{\delta}}_{\mathbf{s}}(\overline{z})\, \boldsymbol{\Gamma}_{\mathbf{u}}(z)\, \tilde{\boldsymbol{\delta}}_{\mathbf{n}}(\overline{z}) - \tilde{\boldsymbol{\Gamma}}_{\mathbf{u}}(z)\, \tilde{\boldsymbol{\delta}}_{\mathbf{n}}(\overline{z})\, \boldsymbol{\Gamma}_{\mathbf{v}}(z)\, \tilde{\boldsymbol{\delta}}_{\mathbf{s}}(\overline{z})$$

defined, consider the filter $\boldsymbol{W}(L)$ defined so that its FRF is given for all $\lambda \in \Lambda$ via the formula

$$\boldsymbol{W}(z) = \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(z) \left[\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z) + \mathbf{B}(z)\, \boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z)\, \tilde{\boldsymbol{\delta}}(z) \right]. \tag{5}$$

Moreover, suppose this formula can be continuously extended to $\lambda \notin \Lambda$ and refer to this extension by $\boldsymbol{W}(z)$ as well. Then the optimal estimate of the signal at time t is given by $\hat{\mathbf{s}}_t = \boldsymbol{W}(L)\mathbf{y}_t$. Furthermore, the spectral density of the signal extraction error process is

$$\tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(z) \left[\tilde{\boldsymbol{\Gamma}}_{\mathbf{u}}(z)\, \boldsymbol{\Gamma}_{\partial\mathbf{u}}(z)\, \tilde{\boldsymbol{\Gamma}}_{\mathbf{u}}(\overline{z}) + \tilde{\boldsymbol{\Gamma}}_{\mathbf{v}}(z)\, \boldsymbol{\Gamma}_{\partial\mathbf{v}}(z)\, \tilde{\boldsymbol{\Gamma}}_{\mathbf{v}}(\overline{z}) \right] \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(\overline{z})$$
$$- \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(z)\, \mathbf{B}(z)\, \boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z)\, \mathbf{B}'(\overline{z})\, \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(\overline{z}).$$

When the differencing operators are uniform, a compact matrix formula for $\boldsymbol{W}(z)$ is given by

$$\boldsymbol{W}(z) = \boldsymbol{\Gamma}_{\partial\mathbf{u}}(z) \boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z) = \boldsymbol{F}_{\partial\mathbf{u}}(\lambda)\, \boldsymbol{F}_{\mathbf{w}}^{-1}(\lambda) \tag{6}$$

for $\lambda \in \Lambda$, and by the limit of such for $\lambda \notin \Lambda$; the error spectral density is $\Gamma_{\mathbf{u}}(z) \boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z) \boldsymbol{\Gamma}_{\mathbf{v}}(z)$.

  J. Time. Ser. Anal. **36**: 209–227 (2015)
DOI: 10.1111/jtsa.12102

Remark 1. The univariate case of formula (6) is given by Bell (1984), who remarks at the end of the article that the method of proof can be extended to the multivariate case. However, this proposed expansion in the scope of Bell's (1984) exposition would rely on the assumption that matrix differencing operators [like $\tilde{\boldsymbol{\delta}}_{\mathbf{n}}(\overline{z})$] commute with spectral density matrices. Only with uniform differencing operators is this assumption guaranteed to hold.

Remark 2. Defining the pseudo-ACGFs of signal and noise as

$$\boldsymbol{\Gamma}_{\mathbf{s}}(L) = \boldsymbol{\Gamma}_{\mathbf{u}}(L)\left[\tilde{\delta}_{\mathbf{s}}(L)\tilde{\delta}_{\mathbf{s}}(L^{-1})\right]^{-1} \qquad \boldsymbol{\Gamma}_{\mathbf{n}}(L) = \boldsymbol{\Gamma}_{\mathbf{v}}(L)\left[\tilde{\delta}_{\mathbf{n}}(L)\tilde{\delta}_{\mathbf{n}}(L^{-1})\right]^{-1},$$

and substituting L for $z = \mathrm{e}^{-i\lambda}$, together with (6) yields the same WK formula as the stationary case, namely (4).

We now show that the key assumptions of Theorem 1 hold for a very wide class of co-integrated processes. Discussion is framed in the context of uniform differencing operators, so the vector signal and noise processes satisfy $\delta_{\mathbf{s}}(L)\mathbf{s}_t = \mathbf{u}_t$ and $\delta_{\mathbf{n}}(L)\mathbf{n}_t = \mathbf{v}_t$. Suppose that the Wold decompositions exist:

$$\begin{aligned} \mathbf{u}_t &= \Xi(L)\boldsymbol{\zeta}_t \qquad \{\boldsymbol{\zeta}_t\} \sim WN(0, \boldsymbol{\Sigma}_{\zeta}), \\ \mathbf{v}_t &= \Omega(L)\boldsymbol{\kappa}_t, \qquad \{\boldsymbol{\kappa}_t\} \sim WN(0, \boldsymbol{\Sigma}_{\kappa}), \end{aligned} \tag{7}$$

where $WN(0, \boldsymbol{\Sigma})$ denotes vector white noise, that is, serially uncorrelated process with zero mean vector and $N \times N$ non-negative definite covariance matrix $\boldsymbol{\Sigma}$. It is assumed that $\{\boldsymbol{\zeta}_t\}$ and $\{\boldsymbol{\kappa}_t\}$ are mutually uncorrelated, and the moving-average filters $\Xi(z)$ and $\Omega(z)$ are linear and causal. As $\Sigma_{\boldsymbol{\zeta}}$ and $\Sigma_{\boldsymbol{\kappa}}$ are non-negative definite, $\boldsymbol{F}_{\mathbf{u}}(\lambda) = \Xi(z)\Sigma_{\boldsymbol{\zeta}}\Xi'(\overline{z})$ is singular if and only if either $|\Sigma_{\boldsymbol{\zeta}}| = 0$ or there is a matrix A and a λ where $A'\Xi(z) = 0$. The former case is described as collinearity of the innovations or in particular is referred to as 'common trends' when the signal is a trend (Stock and Watson, 1988). The latter case is a type of co-integration (with A containing the co-integrating relations); co-integration in the sense of trends applies when there is a singularity at the $\lambda = 0$ frequency. Singularities in the spectrum of a differenced process can only arise in one of these two ways, as collinearity or co-integration (in our generalized sense). Moreover, in the former case, it must follow that the spectrum is singular at all frequencies, whereas in the latter case, there could be an isolated number of singular matrices.

To ensure that $\boldsymbol{F}_{\mathbf{w}}$ is invertible almost everywhere, we assume that $\boldsymbol{F}_{\mathbf{v}}$ is positive definite at all frequencies – as a sufficient condition. In such a case, the process $\{\mathbf{v}_t\}$ is called invertible.

Proposition 1. Suppose that the differencing operators are uniform and that the differenced processes follow (7). Also suppose that $\{\mathbf{v}_t\}$ is invertible. Then $\boldsymbol{F}_{\mathbf{w}}$ is invertible except at a finite set of frequencies, and $\boldsymbol{W}(z)$ can be continuously extended from its natural domain Λ to the entire interval $[-\pi, \pi]$.

3. MULTIVARIATE SIGNAL EXTRACTION FROM A FINITE SAMPLE

Here, we present exact matrix formulae for the solution to multivariate signal extraction problems for a finite series. The set-up is somewhat broader than in the previous section, including processes with heteroscedasticity of known form. Let (3) hold for $t = 1, 2, \ldots, T$. Each series is now a length-T vector, $\mathbf{y}^{(j)} = [y_1^{(j)}, y_2^{(j)}, \ldots, y_T^{(j)}]'$, and similarly for signal, $\mathbf{s}^{(j)}$, and noise, $\mathbf{n}^{(j)}$. For each j, the optimal estimate is the conditional expectation $\mathbb{E}\left[\mathbf{s}^{(j)} \middle| \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(N)}\right]$. As in the previous section, the definition of optimality used here is the minimum linear MSE estimator. The estimate for the jth signal is

$$\hat{\mathbf{s}}^{(j)} = \sum_{k=1}^{N} F^{jk}\mathbf{y}^{(k)}.$$

DOI: 10.1111/jtsa.12102

Each matrix F^{jk} is $T \times T$ dimensional. The superscript j refers to the output, whereas the superscript k refers to the input. It is sufficient to determine the entries of F^{jk} such that error process $\hat{\mathbf{s}}^{(j)} - \mathbf{s}^{(j)}$ is uncorrelated with the observed data.

We may express the specification of the finite series in matrix notation with $\Delta^{(j)}\mathbf{y}^{(j)}$ being a stationary vector, where $\Delta^{(j)}$ is a $T - d^j \times T$-dimensional matrix whose rows consist of the coefficients of $\delta^{(j)}$, appropriately shifted. The application of each $\Delta^{(j)}$ yields a stationary vector, called $\mathbf{w}^{(j)}$, which has length $T - d^j$ (so $\mathbf{w}^{(j)} = [\mathbf{w}^{(j)}_{d^j+1}, \ldots, \mathbf{w}^{(j)}_T]'$). Let $\Sigma^{jk}_{\mathbf{w}} = \mathbb{E}[\mathbf{w}^{(j)}\mathbf{w}^{(k)\prime}]$ denote the associated covariance structure. We further suppose that the differencing is taken such that all random vectors have mean zero (this presupposes that fixed effects have been removed via regression).

This discussion can also be extended to the signal and noise components as follows. Now, form the matrices $\Delta^{(j)}_{\mathbf{s}}$ and $\Delta^{(j)}_{\mathbf{n}}$ corresponding to $\delta^{(j)}_{\mathbf{s}}$ and $\delta^{(j)}_{\mathbf{n}}$. Let $\mathbf{u}^{(j)} = \Delta^{(j)}_{\mathbf{s}}\mathbf{s}^{(j)}$ and $\mathbf{v}^{(j)} = \Delta^{(j)}_{\mathbf{n}}\mathbf{n}^{(j)}$, with cross-covariance matrices denoted $\Sigma^{jk}_{\mathbf{u}}$ and $\Sigma^{jk}_{\mathbf{v}}$. Now assume there are no common roots among $\delta^{(j)}_{\mathbf{s}}$ and $\delta^{(j)}_{\mathbf{n}}$, so that $\delta^{(j)}(L) = \delta^{(j)}_{\mathbf{s}}(L)\delta^{(j)}_{\mathbf{n}}(L)$. Then as in the univariate case (McElroy, 2008)

$$\Delta^{(j)} = \underline{\Delta}^{(j)}_{\mathbf{n}}\Delta^{(j)}_{\mathbf{s}} = \underline{\Delta}^{(j)}_{\mathbf{s}}\Delta^{(j)}_{\mathbf{n}}, \tag{8}$$

where $\underline{\Delta}^{(j)}_{\mathbf{n}}$ and $\underline{\Delta}^{(j)}_{\mathbf{s}}$ are similar differencing matrices of reduced dimension, having $T - d^j$ rows. It follows that

$$\mathbf{w}^{(j)} = \Delta^{(j)}\mathbf{y}^{(j)} = \underline{\Delta}^{(j)}_{\mathbf{n}}\mathbf{u}^{(j)} + \underline{\Delta}^{(j)}_{\mathbf{s}}\mathbf{v}^{(j)}, \tag{9}$$

and hence – if $\mathbf{u}^{(j)}$ and $\mathbf{v}^{(k)}$ are uncorrelated for all j, k –

$$\Sigma^{jk}_{\mathbf{w}} = \underline{\Delta}^{(j)}_{\mathbf{n}}\Sigma^{jk}_{\mathbf{u}}\underline{\Delta}^{(k)\prime}_{\mathbf{n}} + \underline{\Delta}^{(j)}_{\mathbf{s}}\Sigma^{jk}_{\mathbf{v}}\underline{\Delta}^{(k)\prime}_{\mathbf{s}}.$$

We can splice all these $\Sigma^{jk}_{\mathbf{w}}$ matrices together as block matrices in one large matrix $\Sigma_{\mathbf{w}}$, which is also the covariance matrix of $\mathbf{w}$, the vector composed by stacking all the $\mathbf{w}^{(j)}$. A key condition for optimal filtering is the invertibility of $\Sigma_{\mathbf{w}}$; in this case, the Gaussian log likelihood is

$$\log L(\mathbf{w}, \Sigma_{\mathbf{w}}) = -\left(\mathbf{w}'\Sigma^{-1}_{\mathbf{w}}\mathbf{w} + \log|\Sigma_{\mathbf{w}}|\right)/2, \tag{10}$$

up to a constant (initial value vectors are factored out using Assumption M_T), and can be efficiently computed using the innovations algorithm (Brockwell and Davis, 1991). An extended version of the article (McElroy and Trimbur, 2012) shows that for possibly collinear signal and noise innovations, the invertibility of $\Sigma_{\mathbf{w}}$ is guaranteed.

The following assumption, which is the analogue of Assumption M_∞ to the finite-sample case, is given as follows.

Assumption M_T. Suppose that, for each $j = 1, 2, \ldots, N$, the initial values of $\mathbf{y}^{(j)}$ (the first d^j observations) are uncorrelated with $\mathbf{u}$ and $\mathbf{v}$.

Expressing the optimal estimate of the signal as $\widehat{\mathbf{s}} = F\mathbf{y}$, we now present a compact formula for F. The following notation is needed: if G is a block matrix partitioned into sub-matrices G^{jk}, then $\tilde{G}$ denotes a block matrix consisting of only the diagonal sub-matrices G^{jj}, being zero elsewhere.

Theorem 2. Assume $\Sigma_{\mathbf{w}}$ and all $\Sigma^{jj}_{\mathbf{u}}$, $\Sigma^{jj}_{\mathbf{v}}$ are invertible, and $\mathbf{u}^{(j)}$, $\mathbf{v}^{(k)}$ are mutually uncorrelated for all j, k. Also suppose Assumption M_T holds. Define block-matrices A, B, C, and D with jkth block entries given by

$$A^{jk} = \Delta^{(j)\prime}_{\mathbf{s}}\Sigma^{jj\,-1}_{\mathbf{u}}\Sigma^{jk}_{\mathbf{u}}\Sigma^{kk\,-1}_{\mathbf{u}}\Delta^{(k)}_{\mathbf{s}}, \qquad B^{jk} = \Delta^{(j)\prime}_{\mathbf{n}}\Sigma^{jj\,-1}_{\mathbf{v}}\Sigma^{jk}_{\mathbf{v}}\Sigma^{kk\,-1}_{\mathbf{v}}\Delta^{(k)}_{\mathbf{n}},$$
$$C^{jk} = \Delta^{(j)\prime}_{\mathbf{s}}\Sigma^{jj\,-1}_{\mathbf{u}}\Sigma^{jk}_{\mathbf{u}}\underline{\Delta}^{(k)\prime}_{\mathbf{n}}, \qquad D^{jk} = \Delta^{(j)\prime}_{\mathbf{n}}\Sigma^{jj\,-1}_{\mathbf{v}}\Sigma^{jk}_{\mathbf{v}}\underline{\Delta}^{(k)\prime}_{\mathbf{s}}.$$

  J. Time. Ser. Anal. **36**: 209–227 (2015)
DOI: 10.1111/jtsa.12102

Then the matrix $M = \tilde{A} + \tilde{B}$ is invertible. Letting $\tilde{\Delta}$ denote a block diagonal matrix with the matrix $\Delta^{(j)}$ in the jth diagonal,

$$F = M^{-1}\left[\tilde{B} + (C - D)\,\Sigma_{\mathbf{w}}^{-1}\tilde{\Delta}\right].$$

The covariance matrix of the error vector $\hat{\mathbf{s}} - \mathbf{s}$ is $M^{-1}VM^{-1}$, where

$$V = A + B - (C - D)\Sigma_{\mathbf{w}}^{-1}(C - D)'.$$

As an application, consider as the ultimate target of estimation some linear function of the signal, written as $H\mathbf{s}$ for suitable H. (This could be a growth rate, if H acts on $\mathbf{s}$ by temporal differencing, or could be a cross-panel aggregation of signals for a total.) Then the corresponding error covariance matrix is $HM^{-1}VM^{-1}H'$, which requires M and V as given earlier; thus having a knowledge of *all* entries of V, not just the diagonals (which are the MSEs for estimating $\mathbf{s}$), is an advantage of utilizing Theorem 2, as opposed to an SS approach.

4. SIGNAL EXTRACTION FOR RELATED AND COMMON TRENDS

We now treat some stochastic trend models widely used in econometrics. After developing the signal extraction filters, we discuss key aspects of the multivariate difference stationary case that warrant special attention, such as the case of co-integration, or common trends, as in Stock and Watson (1988) and Harvey (1989). In the case of common trends, the gain functions for signal extraction have a collective structure at the frequency origin; this result, although not obvious *a priori,* is aesthetically appealing, and we give a formal proof in the Appendix.

4.1. Discussion of Models

We follow the treatment of Harvey (1989), detailed here so as to set out our notation. Define the vector process $\boldsymbol{\mu}_t = (\mu_t^{(1)}, \ldots, \mu_t^{(N)})'$ as the trend, $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{(1)}, \ldots, \varepsilon_t^{(N)})'$ as the irregular, and $\mathbf{y}_t = (y_t^{(1)}, \ldots, y_t^{(N)})'$ as the observed series. The multivariate local-level model (LLM) is given by

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, &\quad \boldsymbol{\varepsilon}_t &\sim WN(0, \boldsymbol{\Sigma}_{\varepsilon}), \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t, &\quad \boldsymbol{\eta}_t &\sim WN(0, \boldsymbol{\Sigma}_{\eta}), \end{aligned} \tag{11}$$

where we may allow $|\boldsymbol{\Sigma}_{\varepsilon}| = 0$, that is, collinear trend innovations. The irregular $\{\boldsymbol{\varepsilon}_t\}$ accounts for transient factors, for instance, short-run movements due to weather, and is assumed to have invertible covariance matrix. We parametrize each Σ with a Cholesky factorization $\Theta D \Theta'$, where Θ is unit lower triangular and D is a non-negative diagonal matrix (Pinheiro and Bates, 1996). When the trend innovations $\{\boldsymbol{\eta}_t\}$ are collinear (i.e. the process has common trends present), $\boldsymbol{\Sigma}_{\eta}$ has reduced rank $K < N$, and we set the last $N - K$ entries of D to be zero, so that Θ becomes $N \times K$ dimensional.

As an I(2) process, the smooth trend model (STM) specification accounts for a time-varying slope:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, &\quad \boldsymbol{\varepsilon}_t &\sim WN(0, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), \quad t = 1, \ldots, T, \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_t, && \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t, &\quad \boldsymbol{\zeta}_t &\sim WN(0, \boldsymbol{\Sigma}_{\zeta}). \end{aligned} \tag{12}$$

This formulation tends to produce a visibly smooth trend when estimated. As with the LLM, $\boldsymbol{\Sigma}_{\zeta}$ can have a reduced rank (corresponding to common trends).

DOI: 10.1111/jtsa.12102

4.2. Gain Functions and Finite-sample Filters

Here, we present expressions for gain functions in the bi-infinite case (and for exact filters for finite-length series later). Let m denote an integer, where $m = 1$ for the LLM and $m = 2$ for the STM. Since

$$\boldsymbol{F}_{\mathbf{u}}(\lambda) = \Sigma_{\zeta} \qquad \boldsymbol{F}_{\mathbf{v}}(\lambda) = \Sigma_{\boldsymbol{\varepsilon}} \qquad \boldsymbol{F}_{\mathbf{w}}(\lambda) = \Sigma_{\zeta} + |1-z|^{2m}\Sigma_{\boldsymbol{\varepsilon}},$$

the quantities in Theorem 1 are $\boldsymbol{\Gamma}_{\partial\mathbf{u}}(z) = \Sigma_{\zeta}$ and $\boldsymbol{\Gamma}_{\partial\mathbf{v}}(z) = |1-z|^{2m}\Sigma_{\boldsymbol{\varepsilon}}$. Then the multivariate FRF, equivalent to the gain, is

$$\boldsymbol{W}(z) = \Sigma_{\zeta}\left(\Sigma_{\zeta} + |1-z|^{2m}\Sigma_{\boldsymbol{\varepsilon}}\right)^{-1}. \tag{13}$$

The time-domain expression for the filter follows by replacing z by L in (13).

The component gain functions are tied closely to the covariance matrices. Consider $\boldsymbol{W}(1)$, which relates to the lowest, limiting frequency. When Σ_{ζ} is invertible, $\boldsymbol{W}(1) = 1_N$, and there is a separation of gains at the zero frequency – even though the FRF is non-diagonal everywhere else. So, in the absence of co-integration, the filters eventually become specific to each series.

For the common trend case, with non-invertible Σ_{ζ}, a different pattern emerges. Using (A2) in the proof of Proposition 1, it follows that

$$\boldsymbol{W}(1) = \lim_{\lambda\to 0}\boldsymbol{W}(z) = \Theta\left(\Theta'\Sigma_{\boldsymbol{\varepsilon}}^{-1}\Theta\right)^{-1}\Theta'\Sigma_{\boldsymbol{\varepsilon}}^{-1}.$$

In the special case of one common trend with $\Theta = \iota$ – where ι is defined to be the column vector of 1s – and $\Sigma_{\boldsymbol{\varepsilon}}$ is a multiple of the identity matrix, we obtain $\boldsymbol{W}(1) = \iota\iota'/N$, which equally weights the contribution of each input series. In general, when Σ_{ζ} has less than a full rank, there exists β' such that $\beta'\Theta = 0$, which implies that $\beta'\hat{\mathbf{s}}_t = \beta'\boldsymbol{W}(L)\mathbf{y}_t = 0$. Hence, the signal estimate also has a co-integration property. We note in passing that a basis for the space of co-integrating vectors is given by the column space of the inverse of the upper $K \times K$ block of Θ.

For related trends, the determinant of Σ_{ζ} is close to zero; the filter FRF for common trends is similar away from frequency zero, and (13) gives

$$\Psi(z) = \Theta D\Theta'\left(\Theta D\Theta' + |1-z|^{2m}\Sigma_{\boldsymbol{\varepsilon}}^{-1}\right)^{-1}.$$

Now suppose we continuously change a related trend model to common trend form, by letting $N-K$ of the entries of D tend to zero. The limit of $\Theta D\Theta'$ becomes the reduced-rank version, with the last K columns of Θ eliminated. Hence, low frequencies – apart from frequency zero – are treated similarly by the signal extraction FRFs, when the correlations are high. But the treatment of frequency zero remains distinct; given the discontinuity in the FRFs' behaviour at the lower bound of the spectrum – the longest periodicity – it is important to clearly differentiate between the cases of related and common trends.

The same analysis also shows that signal extraction MSE can differ between the two cases. The error spectral density is $\Sigma_{\zeta}\left(\Sigma_{\zeta} + |1-z|^{2m}\Sigma_{\boldsymbol{\varepsilon}}\right)^{-1}\Sigma_{\boldsymbol{\varepsilon}}$; its average integral equals the signal extraction MSE matrix. Since the values at $\lambda = 0$ can be quite different for the common and related trend cases, the resulting MSEs need not be the same. Because of continuity of these functions in λ, no matter how close the related trends' eigenvalues are to zero, it will differ from the common trends' FRF in a neighbourhood of $\lambda = 0$, yielding a discrepancy in their integrals (as also verified numerically). Therefore, it is important to keep a clear distinction between the common versus related trend formulations when deriving gain functions or signal extraction MSE.

  J. Time. Ser. Anal. **36**: 209–227 (2015)
DOI: 10.1111/jtsa.12102

Moving to analytical finite-length filters, the covariance matrices needed in Theorem 2 are given by

$$\Sigma_{\mathbf{u}} = \Sigma_{\zeta} \otimes 1_{T-m} \qquad \Sigma_{\mathbf{v}} = \Sigma_{\boldsymbol{\varepsilon}} \otimes 1_{T}, \tag{14}$$

where $\otimes$ denotes the Kronecker product. Given white noise processes, we obtain partitioned matrices, each $T \times T$ block of which is a scalar times the identity. In the case of common trends, recall that $\Sigma_{\zeta} = \Theta D \Theta'$ has a reduced rank. It follows from (14) that

$$\Sigma_{\mathbf{w}} = \Sigma_{\zeta} \otimes 1_{T-m} + \Sigma_{\boldsymbol{\varepsilon}} \otimes \Delta\, \Delta', \tag{15}$$

where Δ is $(T-m) \times T$, with row entries given by the coefficients of $(1-L)^m$. Observe that (15) can be used as the basis for an explicit Gaussian likelihood for the observed data, given Assumption M_T.

Also, (14) and (15) allow us to compute the signal extraction quantities of Theorem 2. Details are omitted here, but the R code that provides both the exact Gaussian likelihood and the filter and error covariance matrices F, M, and V is available from the authors. Table I provides average run times (in seconds) for various algorithms by sample size, for the LLM (with the related trend parameter specification discussed later). The Gaussian likelihood is computed using both the innovations algorithm and the Kalman filter, with the former being slightly superior in speed. The matrices F and V of Theorem 2 are computed as well (third column of Table I), and when applied to the data to yield signal extraction estimates, the additional cost is given in the fourth column of Table I. If the goal is just to obtain signal estimates and their MSEs, the sum of columns 3 and 4 could be compared with the Kalman smoother column (the sixth), which is substantially faster. However, if one needs the entries of F or the full error covariance matrix, then we should compare the third column with the seventh column, which obtains F via passing impulse responses (i.e. unit vectors) through the Kalman smoother; then it is clear that the matrix-based approach is substantially faster.

An important facet in real applications is to remove outliers and other fixed effects (e.g. trading day effects) before moving on to signal extraction. Often, this is accomplished at the model fitting stage, by adapting the likelihood to allow for a mean function parametrized through regressors – this amounts to utilizing generalized least squares estimation for the fixed effects. If stochastic seasonality is present in the data, the simple trend–irregular

Table I. Average run time (seconds) by sample size for various algorithms, computed for the local-level model

		Average run times for algorithms				
Sample size	Lik	SigMat	SigEx	KF	KS	Coeff
25	0.0314	0.0721	0.0010	0.0409	0.0271	0.9884
50	0.0628	0.2860	0.0012	0.0828	0.0511	3.8894
100	0.1191	0.6614	0.0044	0.1539	0.1064	15.6767
150	0.1975	1.3934	0.0076	0.2330	0.1602	34.6117
200	0.2755	3.6801	0.0146	0.3129	0.1965	61.7130
250	0.3142	5.3545	0.0213	0.3858	0.2483	96.3984
300	0.4032	9.6289	0.0329	0.4646	0.3059	141.3734
350	0.4873	16.4432	0.0438	0.5710	0.3695	198.0243
400	0.5635	24.6435	0.0606	0.6447	0.4197	259.3929
450	0.6560	30.7412	0.0831	0.7296	0.4801	336.5640
500	0.7180	40.1444	0.0885	0.8042	0.5215	414.5795

Lik is the Gaussian likelihood, computed using the innovations algorithm; SigMat computes the matrices F and V of Theorem 2; SigEx applies F to the data to generate the signal extraction estimates; KF is the Kalman filter, together with a computation of the log determinant term needed in the Gaussian likelihood; KS is the Kalman smoother, which yields the signal extraction estimates; Coeff yields the coefficients of the Kalman smoother, that is, the entries of F.

DOI: 10.1111/jtsa.12102

Table II. Local-level model (LLM) parameter estimates for core–total inflation (1986Q1–2010Q4), based on separate univariate LLMs and the related trend bivariate LLM

Series	σ_η^2	σ_ϵ^2	q	ν_η	ν_ϵ	R_D^2	LB
			Parameter estimates for LLM univariate				
Core	5.82e−006	1.77e−005	0.329	NA	NA	0.164	13.42
Total	7.41e−006	0.000172	0.0431	NA	NA	0.261	13.26
			Related trend bivariate				
Core	5.18946e−006	1.84607e−005	0.281	1.000	0.49832	0.179	11.08
Total	3.66532e−006	0.000177859	0.0206	1.000	0.49832	0.355	12.77

See (11) for definitions. LB is the Ljung–Box statistic based on the first 10 residual autocorrelations, and R_D^2 is the coefficient of determination with respect to first differences. For the common trend LLM, the estimate of θ was 0.840.
NA, not applicable.

paradigm described here must be extended by including another stochastic latent component (or potentially several components, if modelling the individual seasonal frequencies); then Theorem 2 can still be applied, but the noise in the trend extraction problem will consist of seasonal plus irregular and will therefore be difference stationary (this is being studied further by the first author in a separate article). A simulation study of signal extraction accuracy can be found in the Supporting Information to this article.

4.3. Inflation Co-movements and Trend Estimation

Our principal goal here is to illustrate the multivariate framework for signal extraction with integrated series. For macroeconomic data, the non-stationary part often has an interpretation or represents a point of reference, as in the case of trend inflation. Here, we consider the use of core inflation – that is, excluding food and energy items – to inform our estimate of the trend in overall inflation. The models considered are the LLM and STM, discussed in Section 4.1.

The base data are the quarterly indices for total and core personal consumption expenditure prices from 1986Q1 to 2010Q4 (source: Bureau of Economic Analysis). Inflation is defined as $4 \cdot \log(P_t/P_{t-1})$ for price index P_t. In terms of notation, we consider $K = 1$ common trends and let ν denote the correlation across series for a given vector component – so for example, ν_ϵ is the correlation between the core and total irregular components. The load matrix just takes the simple form $[1, \theta]'$ for a scalar θ. Kiley (2008) uses a bivariate common trend model with a random walk and with the loading factor constrained to unity. Here, we consider two possible trend specifications and relax the assumption of perfect correlation; whether there are correlated or common trends becomes an empirical question. Also, when we subsequently estimate the common trend form, we allow the loading factor to be unrestricted.

Although the formula for the likelihood is derived from the Gaussian distribution, the assumption of normality is actually not needed for the efficiency of the resulting maximum likelihood estimators (Taniguchi and Kakizawa, 2000). The likelihood function is evaluated using formula (10) for each value of the parameter vector, computed using the innovations algorithm,[1] with the code written in the R statistical programming language.

Local-level model results are given in Table II for the univariate case. The fit was adequate according to standard diagnostics: R_D^2 is the coefficient of determination with respect to first differences, while LB refers to the Ljung–Box based on the first 10 residual autocorrelations – its 5% critical value is 16.9. The resulting trends are shown

[1] As an aside, we independently estimated all the models using Kalman Filtering algorithms (using diffuse initial conditions for the level) and obtained nearly identical results for the MLEs. As shown in Bell and Hillmer (1991), the exact Gaussian likelihood in (10) will be precisely equal to the linear SS likelihood only when initial conditions for recursive computation are based on the transformation approach of Ansley and Kohn (1985).

  J. Time. Ser. Anal. **36**: 209–227 (2015)
DOI: 10.1111/jtsa.12102

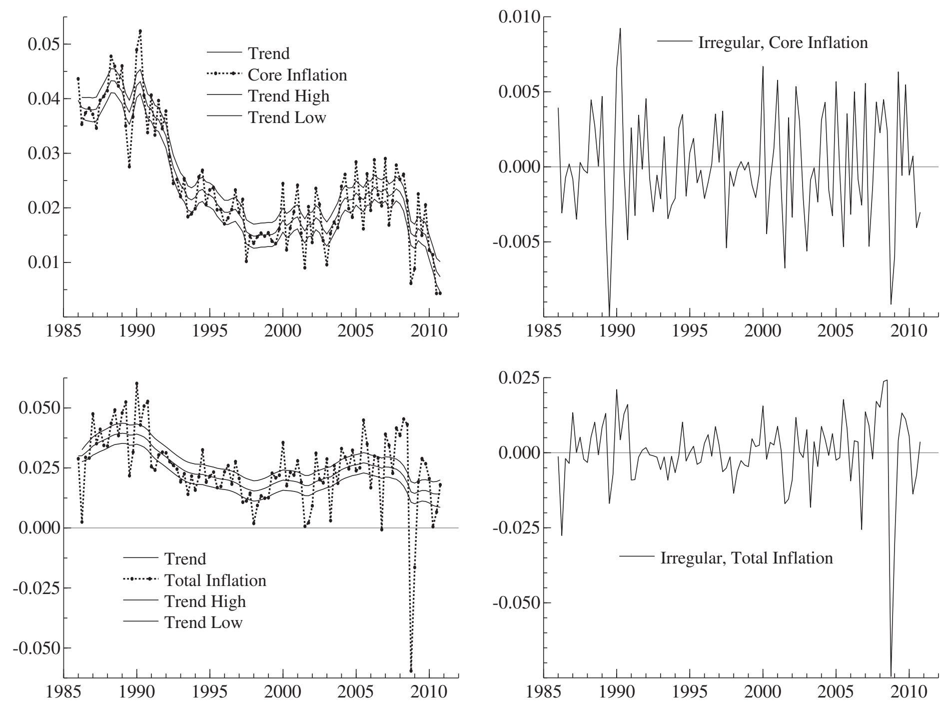

Figure 1. Trend estimates from the univariate local-level model for core–total inflation (1986Q1–2010Q4)

in Figure 1. The signal–noise ratio is defined as $q = \sigma_\eta^2/\sigma_\epsilon^2$, which indicates the relative variability of trend and noise (for a given model structure). The value of q reported in the table is much greater for the core; this quantifies the informal expression that 'core inflation has more signal'.

The results for the bivariate case are shown in Table II and also indicate a good fit. Relative to the univariate case, the R_D^2 rises substantially, now measuring over 35% for total inflation. This points to the utility of core inflation in the model. The cross-correlation for the irregulars takes on a smaller positive value of $\nu_\epsilon = 0.498$, while the correlation between the trend disturbances is numerically indistinguishable from unity. In reformulating the model with a common trend, for identification, we take the trend to represent that of core inflation; the loading factor $\theta = 0.84$ gives the coefficient for mapping the common trend into total inflation's trend. Fit and parsimony can be compared for the related trend and common trend models via the Akaike information criterion (AIC), defined via -2 times the log likelihood, plus a penalty equal to twice the number of model parameters; here, the AICs are -1349.25 for the related trend model and -1351.25 for the common trends, with lower values being preferred.

The left panel of Figure 6 shows the resulting trend in total inflation and compares it with the univariate output. The solid lines pertain to the bivariate estimates. There are noticeable differences in both the trajectory of the bivariate trend and in the substantially reduced degree of uncertainty associated with its estimation. Linked to our signal extraction formulas, the model-based estimator has a formulation as a bivariate filter; for each element as a 2×2 matrix of scalar filters of the usual form, Figure 2 plots the observation weight against the time separation from the signal location. The core-to-core weighting pattern in the upper-left box resembles a double exponential (with a slight offset). The weights for total-to-core pattern show that a weighted average of total is actually *subtracted*. So, overall, adjacent values of core inflation are somewhat overweighted and offset by subtracting weighted adjacent values of total inflation. The weights for estimating trend inflation, shown in the bottom half of the figure, have the same basic pattern (since there is a common trend).

DOI: 10.1111/jtsa.12102

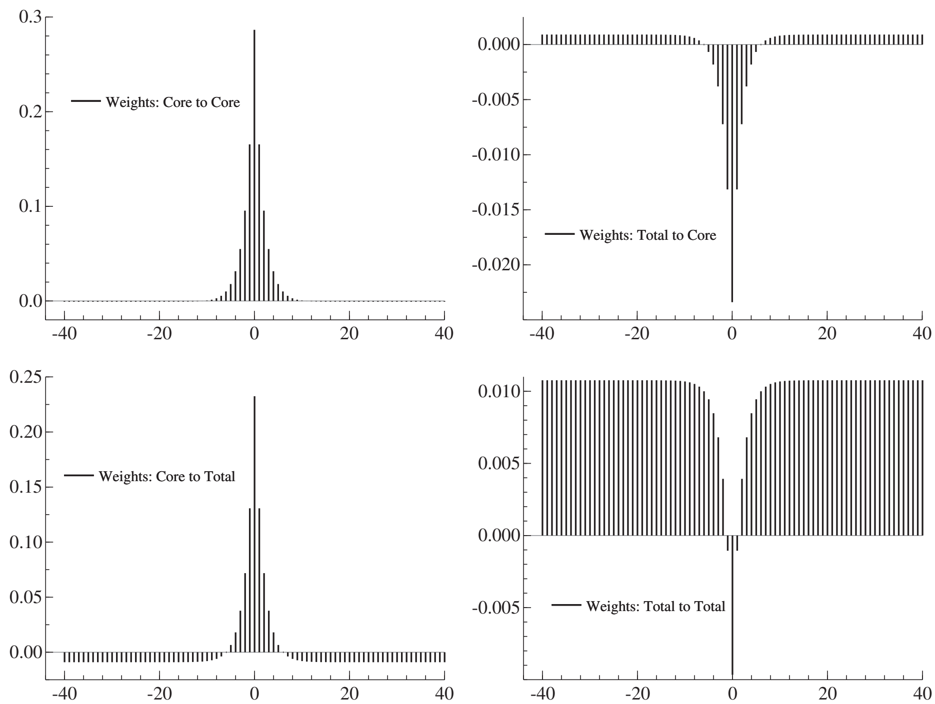

Figure 2. Observation weights for estimating trends in core and total inflation for the local-level model. The horizontal axis is the separation between the signal estimate and observation time points. Note that the y-axes have differing scales

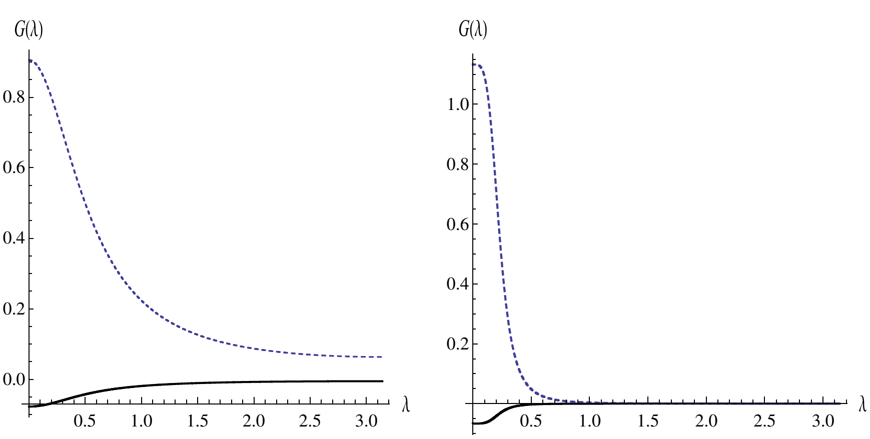

Figure 3. Total trend inflation gain functions, as transfer functions from core (dashed) and total (solid) inflation respectively, based on the local-level model (left panel) and smooth trend model (right panel)

The gain functions (away from the end points) for estimating trend inflation are shown in the left panel of Figure 3. The filter gain for the core-to-total pattern essentially has the usual shape of a low-pass filter with its focus on low frequencies. The cross-gain for the total-to-total pattern is essentially an inverted low pass. Analogous to the weight functions, there are small shifts in the gains compared with a univariate model's signal gain, which would range between zero and unity. Reflecting the filter weights, the gains for estimating the signal in core inflation have a similar profile to those in Figure 3 and are omitted for brevity.

DOI: 10.1111/jtsa.12102

Table III. Smooth trend model (STM) parameter estimates for core–total inflation (1986Q1–2010Q4), based on separate univariate STMs and the related trend bivariate STM

Series	σ_η^2	σ_ϵ^2	q	ν_η	ν_ϵ	R_D^2	LB
			Parameter estimates for STM univariate				
Core	6.36e−008	2.33e−005	0.00273	NA	NA	0.133	16.37
Total	6.17e−008	0.000185	0.000334	NA	NA	0.22	12.64
			Related trend bivariate				
Core	6.11564e−008	2.34043e−005	0.00261	1.000	0.513972	0.146	15.30
Total	6.89953e−008	0.000179544	0.000384	1.000	0.513972	0.346	14.07

See (12) for definitions. LB is the Ljung–Box statistic based on the first 10 residual autocorrelations, and R_D^2 is the coefficient of determination with respect to first differences. For the common trend STM, the estimate of θ was 1.06219.
NA, not applicable.

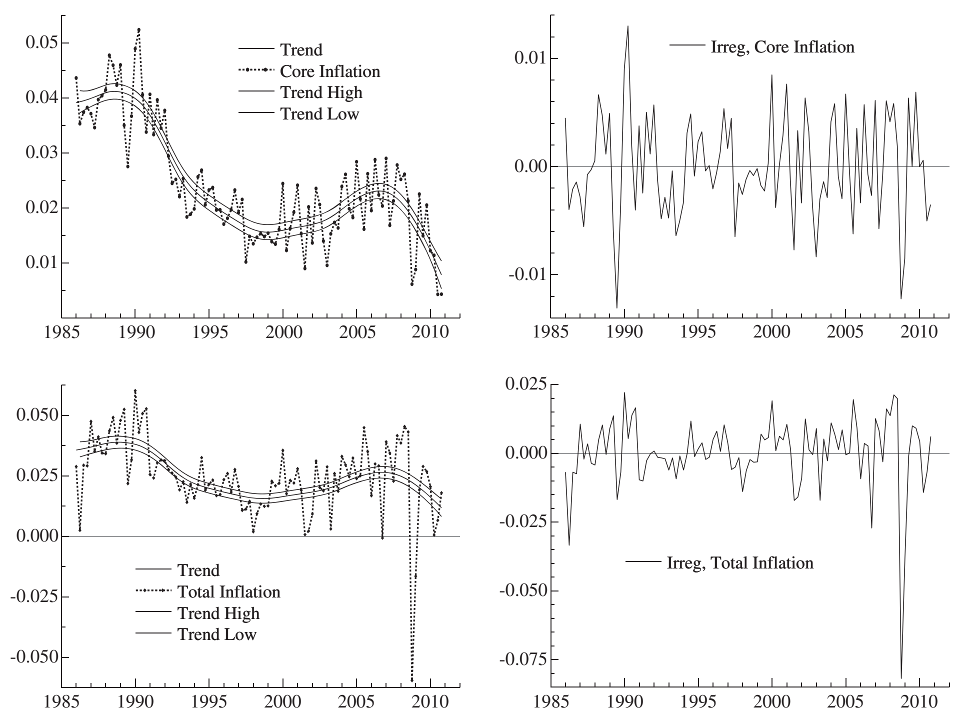

Figure 4. Trend estimates from the bivariate smooth trend model for core–total inflation (1986Q1–2010Q4)

The results for the STM are shown in Table III. The fit measures and diagnostics are similar to those for the LLM, and there is again a clear superiority of the bivariate model over the univariate specification. The cross-correlation for the irregulars is $\nu_\epsilon = 0.514$, while the correlation between the trend disturbances is essentially unity. The common trend loading factor is $\theta = 1.062$, and the respective AICs are -1332.06 for related trends and -1334.06 for common trends. The estimated trends, displayed in Figure 4, show more stable variation and clearer turning points than the LLM. The weighting kernels for the STM case are shown in Figure 5, and the gain functions for the resulting pair of low-pass filters for estimating trend inflation are shown in the right panel of Figure 3. The gain applied to core inflation cuts off more sharply than the LLM gains (left panel), giving greater

DOI: 10.1111/jtsa.12102

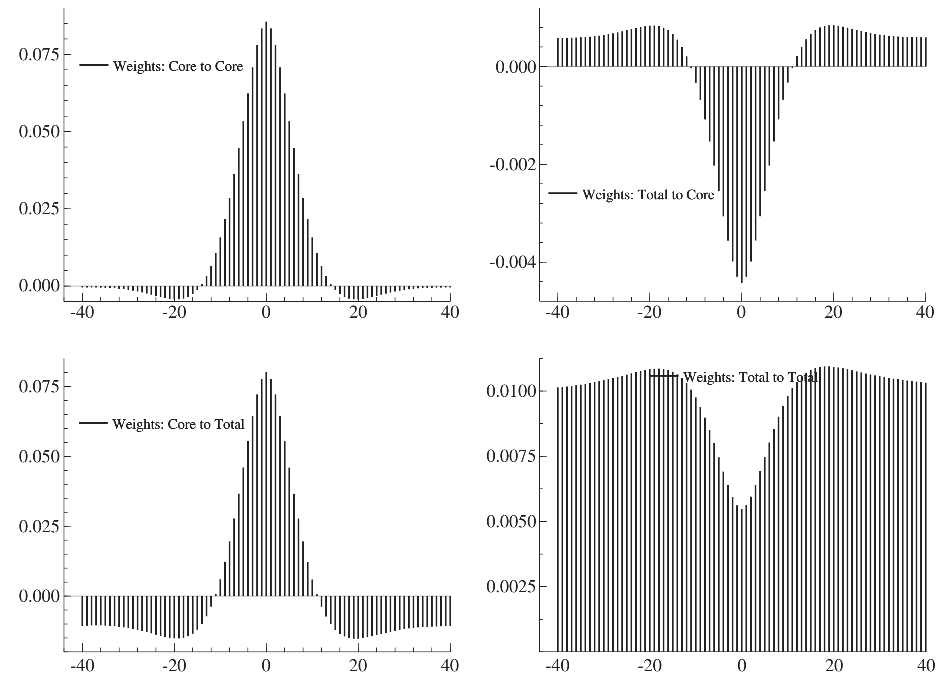

Figure 5. Observation weights for estimating trends in core and total inflation for the smooth trend model. The horizontal axis is the separation between the signal estimate and observation time points. Note that the y-axes have differing scales

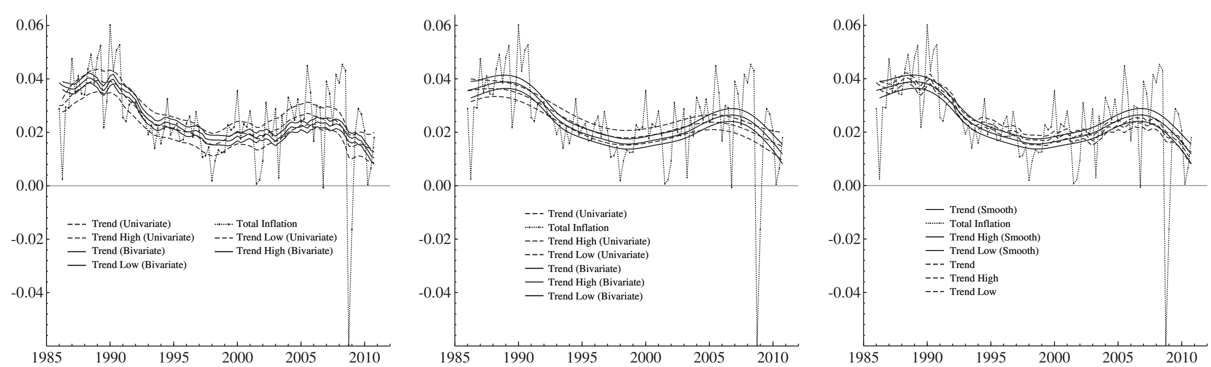

Figure 6. Trend estimates for total inflation (1986Q1–2010Q4) from various models: univariate and bivariate local-level model (LLM; left panel), univariate and bivariate smooth trend model (STM; middle panel), and bivariate LLM and STM (right panel)

persistence in the direction of the trend as well as increased smoothness. Further comparisons of the estimated trends of total inflation can be found in the middle and right panels of Figure 6.

While our illustrations have used simple stochastic trend plus noise models, they already show that the exact optimal weighting of just two series has non-obvious aspects, such as a slight over-emphasis on the signalling series compensated by subtraction of a weighted average of the noisier target.

5. CONCLUSION

We have provided explicit matrix formulae for multivariate signal extraction computed from either finite or bi-infinite samples drawn from difference stationary processes, continuing a thread of literature (e.g. Wiener, 1949;

DOI: 10.1111/jtsa.12102

Whittle, 1963; Bell, 1984) on signal extraction theory. These results are the most general in the literature to date, encompassing processes embeddable in a linear SS framework, while also being more broadly applicable to vector long-memory processes. A precise discussion and treatment of co-integration is provided. We highlight, in particular, how an invariance property arises: trend estimators (from a bi-infinite sample) have the same collinearity relations as the unobserved stochastic trends. This is intuitively attractive, but not obviously true beforehand. These new tools are illustrated on a bivariate inflation application.

In the statistical interpretation and analysis of co-movements across related time series, a diverse set of multivariate signal extraction systems arise in economics and other fields. The bi-infinite analysis shows, in compact form, precisely how to best use the rich set of information in a set of variables to construct signal estimates. Correspondingly, we can now examine the analytical expressions for the optimal set of weight polynomials, applied to the observation vectors, based on the matrix ACGFs or pseudo-ACGFs in the time domain; in the frequency domain, the generalized gain function is now based on matrix (pseudo-)spectra, illustrated by the introduction of multivariate low-pass filters for I(1) and I(2) series.

The finite-sample time-domain formulae, which we have used for the actual application of the filters, expand the available methodology and give a strategy distinct from the SS approach. Our compact method gives exact results at all time points in a single matrix formula, and the entire error covariance matrix is explicitly available, which is beneficial when studying the uncertainty in signal growth rate estimates, or cross-sectional aggregate signal estimates. However, linear SS methods have already been implemented in several excellent software packages (e.g. STAMP) and tend to be faster than the matrix methods introduced here. A practitioner may lean towards using SS methods when speed is of utmost importance but may use the matrix formulae when it is necessary to gain insight into filter weights, or for models not embeddable in the SS framework.

ACKNOWLEDGEMENTS

This article was presented at the Society for Nonlinear Dynamics and Econometrics Conference (in Washington DC in March 2011), at the National Bureau of Economic Research/National Science Foundation time-series conference (in Lansing, Michigan, in September 2011), at a University of Maryland–College Park seminar in October 2011, and at an Federal Reserve Bank seminar in November 2011. We would like to thank the participants at these presentations, as well as two referees, for helpful feedback and discussion.

DISCLAIMER

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the US Census Bureau or the Federal Reserve Board.

REFERENCES

Ansley C, Kohn R. 1985. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics* **13**: 1286–1316.

Basistha A, Startz D. 2008. Measuring the NAIRU with reduced uncertainty: a multiple-indicator common-cycle approach. *Review of Economics and Statistics* **90**: 805–811.

Bell W. 1984. Signal extraction for nonstationary time series. *The Annals of Statistics* **12**: 646–664.

Bell W, Hillmer S. 1988. A matrix approach to likelihood evaluation and signal extraction for ARIMA component time series models. *SRD research report RR-88/22*, Bureau of the Census.

Bell W, Hillmer S. 1991. Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis* **12**: 283–300.

Brockwell P, Davis R. 1991. *Time Series: Theory and Methods*, 2nd Ed. New York: Springer.

Engle R, Granger C. 1987. Cointegration and error correction: representation, estimation, and testing. *Econometrica* **55**: 251–276.

Golub G, Van Loan C. 1996. *Matrix Computations*. Baltimore and London: The Johns Hopkins University Press.
Gómez V. 2006. Wiener–Kolmogorov filtering and smoothing for multivariate series with state-space structure. *Journal of Time Series Analysis* **28**: 361–385.
Harvey A. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
Harvey AC, Trimbur T. 2003. General model-based filters for extracting cycles and trends in economic time series. *Review of Economics and Statistics* **85**: 244–255.
Johansen S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.
Kiley M. 2008. Estimating the common trend rate of inflation for consumer prices and consumer prices excluding food and energy prices. *Finance and Economics Discussion Series, Federal Reserve Board*, 2008-38. http://www.federalreserve.gov/pubs/feds/2008/200838/200838pap.pdf [Accessed on 22 August 2008].
McElroy T. 2008. Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* **24**: 1–22.
McElroy T, Trimbur T. 2012. Signal extraction for nonstationary multivariate time series with illustrations for trend inflation. *Finance and Economics Discussion Series, Federal Reserve Board*, 2012-45. http://www.federalreserve.gov/pubs/feds/2012/201245/201245abs.html [Accessed on 9 November 2012].
Pinheiro J, Bates D. 1996. Unconstrained parametrizations for variance–covariance matrices. *Statistics and Computing* **6**: 289–296.
Pollock S. 2007. Wiener–Kolmogorov filtering, frequency-selective filtering, and polynomial regression. *Econometric Theory* **23**: 71–88.
Stock J, Watson M. 1988. Testing for common trends. *Journal of the American Statistical Association* **83**: 1097–1107.
Taniguchi M, Kakizawa Y. 2000. *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer-Verlag.
Whittle P. 1963. *Prediction and Regulation*. London: English Universities Press.
Wiener N. 1949. *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley.

APPENDIX A:

Proof of Theorem 1.
The action of each filter on a time series is obtained by replacing z by L and $\overline{z}$ by L^{-1}. The error process is defined via $\boldsymbol{\epsilon}_t = \boldsymbol{W}(L)\mathbf{y}_t - \mathbf{s}_t$. [It is also easy to derive an expression for $1_N - \boldsymbol{W}(z)$, which has the same formula as (5) but with signal and noise swapped.] Then applying (5) to this error decomposition yields

$$\boldsymbol{\epsilon}_t = \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(L)\,\mathbf{B}(L)\,\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(L)\,\mathbf{w}_t + \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(L)\,\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(L)\mathbf{n}_t - \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(L)\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{v}}(L)\mathbf{s}_t. \tag{A1}$$

Now even in the case that differencing operators are not uniform, it is true that

$$\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z)\,\mathbf{n}_t = \tilde{\boldsymbol{\Gamma}}_{\mathbf{u}}(z)\,\tilde{\boldsymbol{\delta}}_{\mathbf{n}}(\overline{z})\,\mathbf{v}_t,$$
$$\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{v}}(z)\,\mathbf{s}_t = \tilde{\boldsymbol{\Gamma}}_{\mathbf{v}}(z)\,\tilde{\boldsymbol{\delta}}_{\mathbf{s}}(\overline{z})\,\mathbf{u}_t.$$

Substituting these expressions into (A1) shows that $\{\boldsymbol{\epsilon}_t\}$ is stationary. Moreover, by the definition of $\mathbf{B}(z)$, we see that $\boldsymbol{\epsilon}_t$ is orthogonal to $\mathbf{w}_{t+h}$ for all h and any t. In the expectation calculations, we can use the Cramer representation for $\{\mathbf{w}_t\}$ to obtain expressions involving integrals – then the fact that the FRF is not defined on a set of measure zero becomes irrelevant. Moreover, since the error process only depends on $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$, it is orthogonal to the initial values of the data process as well, by Assumption M_∞. Hence, the error process is uncorrelated with $\{\mathbf{y}_t\}$, proving the optimality of the signal extraction filter. The formula for the error spectral density also follows from the earlier description of the error process, using the fact that the spectra are Hermitian [i.e. $\boldsymbol{\Gamma}(z) = \boldsymbol{\Gamma}'(\overline{z})$].

Under the assumption that the differencing operators are uniform, we have (since such diagonal matrices commute with any other matrix)

$$\mathbf{B}(z)\,\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z)\,\tilde{\boldsymbol{\delta}}(z) = \left[\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{v}}(z)\,\boldsymbol{\Gamma}_{\partial\mathbf{u}}(z) - \tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z)\,\boldsymbol{\Gamma}_{\partial\mathbf{v}}(z)\right]\,\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z).$$

DOI: 10.1111/jtsa.12102

Using this relation produces

$$\begin{aligned}\boldsymbol{W}(z) &= \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(z)\left[\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z) + \left(\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{v}}(z)\boldsymbol{\Gamma}_{\partial\mathbf{u}}(z) - \tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z)\boldsymbol{\Gamma}_{\partial\mathbf{v}}(z)\right)\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z)\right] \\ &= \tilde{\boldsymbol{\Gamma}}_{\mathbf{w}}^{-1}(z)\left[\tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{u}}(z)\boldsymbol{\Gamma}_{\partial\mathbf{u}}(z) + \tilde{\boldsymbol{\Gamma}}_{\partial\mathbf{v}}(z)\boldsymbol{\Gamma}_{\partial\mathbf{u}}(z)\right]\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z) = \boldsymbol{\Gamma}_{\partial\mathbf{u}}(z)\boldsymbol{\Gamma}_{\mathbf{w}}^{-1}(z).\end{aligned}$$

Similar substitutions in the error spectral density produce the stated formula. □

Proof of Proposition 1.
The spectral density of the differenced data process is

$$\boldsymbol{F}_{\mathrm{w}}(\lambda) = |\delta_{\mathbf{n}}(z)|^2\,\Xi(z)\Sigma_{\boldsymbol{\zeta}}\,\Xi'(\overline{z}) + |\delta_{\mathbf{s}}(z)|^2\,\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z}).$$

For any fixed λ, this expression is the linear combination of a positive definite matrix and a non-negative definite matrix. So long as $|\delta_{\mathbf{s}}(z)|^2$ is non-zero, $\boldsymbol{F}_{\mathrm{w}}(\lambda)$ will be invertible, but otherwise, the matrix is singular. Let $p(\lambda) = |\delta_{\mathbf{n}}(z)|^2$ and $q(\lambda) = |\delta_{\mathbf{s}}(z)|^2$ for short. If λ_s is a root of q, then

$$\mathbf{F}_{\mathrm{w}}(\lambda_s) = p(\lambda_s)\,\Xi(z_s)\Sigma_{\boldsymbol{\zeta}}\,\Xi'(\overline{z_s}),$$

where $z_s = \mathrm{e}^{-i\lambda_s}$. If this matrix has determinant zero, then either $|\Sigma_{\boldsymbol{\zeta}}| = 0$, or there exists some matrix a (depending on λ_s) such that $a'\Xi(z_s) = 0$. The former case corresponds to the collinearity of the differenced signal process, whereas the latter case is discussed further in the following. If the differenced signal process is collinear, then from the singular value decomposition of $\Sigma_{\boldsymbol{\zeta}}$, we can write this matrix as $\Theta D_{\boldsymbol{\zeta}}\Theta'$. Here, $D_{\boldsymbol{\zeta}}$ has a reduced dimension corresponding to the number of non-zero eigenvalues of $\Sigma_{\boldsymbol{\zeta}}$, and Θ is rectangular. Then the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996) yields

$$\begin{aligned}\boldsymbol{F}_{\mathrm{w}}^{-1}(\lambda) &= q^{-1}(\lambda)\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1} - \frac{p(\lambda)}{q(\lambda)}\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1}\Xi(z)\Theta D_{\boldsymbol{\zeta}} \\ &\quad \cdot \left\{q(\lambda)1_N + p(\lambda)\Theta'\Xi'(\overline{z})\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1}\Xi(z)\Theta D_{\boldsymbol{\zeta}}\right\}^{-1} \cdot \Theta'\Xi'(\overline{z})\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1}.\end{aligned}$$

This formula is quite informative: the preceding derivation shows that $q\boldsymbol{F}_{\mathrm{w}}^{-1}$ is bounded for all λ, even in a neighbourhood of λ_s. Unfortunately, the matrix $\boldsymbol{F}_{\mathrm{w}}$ is not invertible at λ_s, and the rate of explosion of $q^{-1}(\lambda)$ is generally such that the Whittle likelihood (3.1.5 of Taniguchi and Kakizawa, 2000) is not well defined. However, the FRF (6) for any λ such that $q(\lambda) \neq 0$ is given by

$$\begin{aligned}\boldsymbol{W}(z) &= 1_N - q(\lambda)\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\boldsymbol{F}_{\mathrm{w}}^{-1}(\lambda) \\ &= p(\lambda)\Xi(z)\Theta D_{\boldsymbol{\zeta}} \cdot \left\{q(\lambda)1_N + p(\lambda)\Theta'\Xi'(\overline{z})\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1}\Xi(z)\Theta D_{\boldsymbol{\zeta}}\right\}^{-1} \\ &\quad \cdot \Theta'\Xi'(\overline{z})\left[\Omega(z)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z})\right]^{-1},\end{aligned}$$

and the limit as $\lambda \to \lambda_s$ of this is well defined; we set $q(\lambda) = 0$ and $p(\lambda) = p(\lambda_s)$ (which is non-zero, since the signal and noise differencing polynomials are relatively prime by assumption). That is,

$$\begin{aligned}\boldsymbol{W}(z_s) &= p(\lambda_s)\Xi(z_s)\Theta\Sigma_{\boldsymbol{\eta}} \cdot \left\{p(\lambda_s)\Theta'\Xi'(\overline{z_s})\left[\Omega(z_s)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z_s})\right]^{-1}\Xi(z_s)\Theta D_{\boldsymbol{\zeta}}\right\}^{-1} \\ &\quad \cdot \Theta'\Xi'(\overline{z_s})\left[\Omega(z_s)\Sigma_{\boldsymbol{\kappa}}\Omega'(\overline{z_s})\right]^{-1}.\end{aligned} \qquad \text{(A2)}$$

DOI: 10.1111/jtsa.12102

Now let us treat the case that the innovations are not collinear, and yet $a'\Xi(z) = 0$ at some measure zero set of frequencies (with a different a for each frequency λ). Let Λ denote the set of frequencies where $\mathbf{F_w}$ is invertible, so that

$$\mathbf{\Gamma}_{\partial \mathbf{u}}(z)\mathbf{\Gamma}_{\mathbf{w}}^{-1}(z) = 1_N - q(\lambda)\big(1_N + p(\lambda)\mathbf{F_u}(\lambda)\mathbf{F_v}^{-1}(\lambda)\big)^{-1},$$

which is well defined for all frequencies λ; for $\lambda \notin \Lambda$, we can take the limit as $\lambda \to \lambda_s$, since the matrix inverse in the preceding expression always exists. The limit is just 1_N, so $\mathbf{W}(z)$ can be continuously extended to all frequencies. This completes the proof. □

Proof of Theorem 2.
First note that $M^{jj} = \Delta_{\mathbf{n}}^{(j)\prime}\Sigma_{\mathbf{v}}^{jj^{-1}}\Delta_{\mathbf{n}}^{(j)} + \Delta_{\mathbf{s}}^{(j)\prime}\Sigma_{\mathbf{u}}^{jj^{-1}}\Delta_{\mathbf{s}}^{(j)}$, and invertibility follows from the univariate argument of McElroy (2008), using the fact that $\delta_{\mathbf{s}}^{(j)}$ and $\delta_{\mathbf{n}}^{(j)}$ share no common factors. Next, the jkth block matrix of the putative formula for F is

$$\sum_{\ell=1}^{N} M^{j\ell^{-1}}\left(\tilde{B}^{\ell k} + \sum_{p,q=1}^{N}\left(C^{\ell p} - D^{\ell p}\right)\Gamma_{\mathbf{w}}^{pq}\Delta^{qk}\right) = M^{jj^{-1}}\left(B^{jk} + \sum_{p=1}^{N}\left(C^{jp} - D^{jp}\right)\Gamma_{\mathbf{w}}^{pk}\Delta^{(k)}\right),$$

using the block-diagonal structure of M, $\tilde{B}$, and Δ. Letting $\Gamma_{\mathbf{w}} = \Sigma_{\mathbf{w}}^{-1}$, the formula for F^{jk} is then

$$\begin{aligned} F^{jj} &= M^{jj^{-1}}\left[\Delta_{\mathbf{n}}^{(j)\prime}\Sigma_{\mathbf{v}}^{jj^{-1}}\Delta_{\mathbf{n}}^{(j)} + \Sigma_{\ell=1}^{N}\left(\Delta_{\mathbf{s}}^{(j)\prime}\Sigma_{\mathbf{u}}^{jj^{-1}}\Sigma_{\mathbf{u}}^{j\ell}\,\underline{\Delta}_{\mathbf{n}}^{(\ell)\prime} - \Delta_{\mathbf{n}}^{(j)\prime}\Sigma_{\mathbf{v}}^{jj^{-1}}\Sigma_{\mathbf{v}}^{j\ell}\underline{\Delta}_{\mathbf{s}}^{(\ell)\prime}\right)\Gamma_{\mathbf{w}}^{\ell j}\Delta^{(j)}\right] \\ F^{jk} &= M^{jj^{-1}}\Sigma_{\ell=1}^{N}\left(\Delta_{\mathbf{s}}^{(j)\prime}\Sigma_{\mathbf{u}}^{jj^{-1}}\Sigma_{\mathbf{u}}^{j\ell}\underline{\Delta}_{\mathbf{n}}^{(\ell)\prime} - \Delta_{\mathbf{n}}^{(j)\prime}\Sigma_{\mathbf{v}}^{jj^{-1}}\Sigma_{\mathbf{v}}^{j\ell}\underline{\Delta}_{\mathbf{s}}^{(\ell)\prime}\right)\Gamma_{\mathbf{w}}^{\ell k}\Delta^{(k)} \end{aligned}$$

for $j \neq k$. Note that in these formulae, for $\ell = j$, we obtain

$$\Delta_{\mathbf{s}}^{(j)\prime}\Sigma_{\mathbf{u}}^{jj^{-1}}\Sigma_{\mathbf{u}}^{jj}\,\underline{\Delta}_{\mathbf{n}}^{(j)\prime} - \Delta_{\mathbf{n}}^{(j)}\Sigma_{\mathbf{v}}^{jj^{-1}}\Sigma_{\mathbf{v}}^{jj}\,\underline{\Delta}_{\mathbf{s}}^{(j)\prime} = \Delta^{(j)} - \Delta^{(j)} = 0$$

by (8), which causes some simplification. The signal error for the jth signal is

$$\epsilon^{(j)} = \widehat{\mathbf{s}}^{(j)} - \mathbf{s}^{(j)} = \sum_{k\neq j} F^{jk}\mathbf{y}^{(k)} + F^{jj}\mathbf{n}^{(j)} - (1 - F^{jj})\mathbf{s}^{(j)}.$$

Optimality of the matrix formulae follows from demonstrating that $\epsilon^{(j)}$ is orthogonal to $\mathbf{y}$, the stacking of the individual data vectors $\mathbf{y}^{(k)}$. As in Bell (1984), we can write $\mathbf{y}$ as a linear combination of $\mathbf{w}$ and various initial values for all the component series. By utilizing Assumption M_T, it is then sufficient to demonstrate that $\epsilon^{(j)}$ is uncorrelated with $\mathbf{w}$. Noting the formula for $1 - F^{jj}$, we obtain

$$\begin{aligned} \epsilon^{(j)} = & \sum_{k\neq j} M^{jj^{-1}}\sum_{\ell=1}^{N}\left(C^{j\ell} - D^{j\ell}\right)\Gamma_{\mathbf{w}}^{\ell k}\mathbf{w}^{(k)} \\ & + M^{jj^{-1}}\left[\Delta_{\mathbf{n}}^{(j)\prime}\Sigma_{\mathbf{v}}^{jj^{-1}}\mathbf{v}^{(j)} + \Sigma_{\ell=1}^{N}\left(C^{j\ell} - D^{j\ell}\right)\Gamma_{\mathbf{w}}^{\ell j}\,\underline{\Delta}_{\mathbf{s}}^{(j)}\mathbf{v}^{(j)}\right] \\ & - M^{jj^{-1}}\left[\Delta_{\mathbf{s}}^{(j)\prime}\Sigma_{\mathbf{u}}^{jj^{-1}}\mathbf{u}^{(j)} - \Sigma_{\ell=1}^{N}\left(C^{j\ell} - D^{j\ell}\right)\Gamma_{\mathbf{w}}^{\ell j}\,\underline{\Delta}_{\mathbf{n}}^{(j)}\mathbf{u}^{(j)}\right]. \end{aligned}$$

DOI: 10.1111/jtsa.12102

Next, we show that the covariance between this error vector and each $\mathbf{w}^{(p)}$ is zero for each p. Noting (9), we obtain

$$\begin{aligned}
\mathbb{E}\left[\epsilon^{(j)}\mathbf{w}^{(p)\prime}\right] &= \sum_{k \neq j} M^{jj^{-1}} \sum_{\ell=1}^{N} \left(C^{j\ell} - D^{j\ell}\right) \Gamma_{\mathbf{w}}^{\ell k} \Sigma_{\mathbf{w}}^{kp} \\
&\quad + M^{jj^{-1}} \left[\Delta_{\mathbf{n}}^{(j)\prime} \Sigma_{\mathbf{v}}^{jj^{-1}} \Sigma_{\mathbf{v}}^{jp} \underline{\Delta}_{\mathbf{s}}^{(p)\prime} + \Sigma_{\ell=1}^{N} \left(C^{j\ell} - D^{j\ell}\right) \Gamma_{\mathbf{w}}^{\ell j} \underline{\Delta}_{\mathbf{s}}^{(j)} \Sigma_{\mathbf{v}}^{jp} \underline{\Delta}_{\mathbf{s}}^{(p)\prime}\right] \\
&\quad - M^{jj^{-1}} \left[\Delta_{\mathbf{s}}^{(j)\prime} \Sigma_{\mathbf{u}}^{jj^{-1}} \Sigma_{\mathbf{u}}^{jp} \underline{\Delta}_{\mathbf{n}}^{(p)\prime} - \Sigma_{\ell=1}^{N} \left(C^{j\ell} - D^{j\ell}\right) \Gamma_{\mathbf{w}}^{\ell j} \underline{\Delta}_{\mathbf{n}}^{(j)} \Sigma_{\mathbf{u}}^{jp} \underline{\Delta}_{\mathbf{n}}^{(p)\prime}\right] \\
&= M^{jj^{-1}} \sum_{\ell=1}^{N} \left(C^{j\ell} - D^{j\ell}\right) \sum_{k \neq j} \Gamma_{\mathbf{w}}^{\ell k} \Sigma_{\mathbf{w}}^{kp} \\
&\quad + M^{jj^{-1}} \Sigma_{\ell=1}^{N} \left(C^{j\ell} - D^{j\ell}\right) \Gamma_{\mathbf{w}}^{\ell j} \Sigma_{\mathbf{w}}^{jp} \\
&\quad - M^{jj^{-1}} \left(C^{jp} - D^{jp}\right).
\end{aligned}$$

By the definition of $\Gamma_{\mathbf{w}}^{\ell k}$, we know that $\sum_k \Gamma_{\mathbf{w}}^{\ell k} \Sigma_{\mathbf{w}}^{kp}$ is a matrix of zeroes unless $\ell = p$, in which case it is an identity matrix. As a result, the preceding calculation simplifies to zero. This completes the proof of optimality. The signal extraction covariance matrix is produced by using the preceding expression for $\epsilon^{(j)}$ in $\mathbb{E}[\epsilon^{(j)}\epsilon^{(k)\prime}]$, after much algebra. □

DOI: 10.1111/jtsa.12102