Signal extraction for nonstationary multivariate time series with applications to trend inflation

Tucker McElroy
U.S. Census Bureau

United States™ Census Bureau

Preface

Acknowledgements: Co-authored with Thomas Trimbur (Federal Reserve Board)

United States™ Census Bureau

U.S. DEPARTMENT OF COMMERCE ★ BUREAU OF THE CENSUS ★

Abstract

We advance the theory of signal extraction by developing the optimal treatment of nonstationary vector time series that may have common trends. We present new formulas for exact signal estimation for both theoretical bi-infinite and finite samples. The formulas reveal the specific roles of inter-relationships among variables for sets of optimal filters, which makes fast and direct calculation feasible, and shows rigorously how the optimal asymmetric filters are constructed near the end points for a set of series. We develop a class of model-based low-pass filters for trend estimation and illustrate the methodology by studying statistical estimates of trend inflation.

Outline

- Trends of inflation: a data example to motivate results
- Multivariate filtering in frequency domain
- Time domain formulation of results
- Common trends models

Inflation

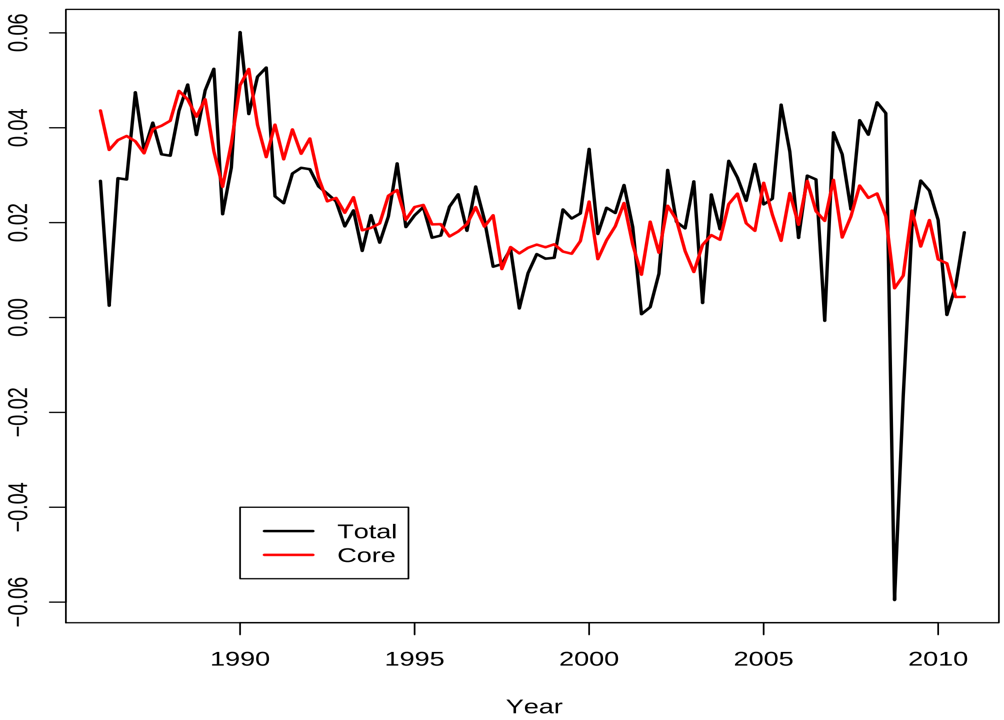

Figure 1: Core and Total PCE inflation, 1986 through 2010.

Inflation Data

- The underlying trend rate in inflation is monitored by central banks, and has a significant impact on monetary policy decisions.

- Core inflation excludes food and energy, and is less volatile. But Total inflation includes gasoline prices, and seems important too. We expect both series to be highly correlated in their movements

- Can we assess the trend via Total inflation, but with added information from Core inflation?

- We use the Personal Consumption Expenditures (PCE) Core price index and Total index, source is Bureau of Economic Analysis, quarterly from 1986 through end of 2010.

Common Trends in Inflation

- We might expect correlated trends: when correlation is full (between trend component innovations), we say there is a common trend; else we say there are similar trends.

- We want a trend estimation (or signal extraction) method that utilizes a bivariate time series model.

- Simple trend-plus-noise models are the Local Level Model (LLM) and the Smooth Trend Model (STM) of Harvey (1989). We can consider either similar or common trends for either, giving 4 models total.

- Given the specified model, we need to: (i) fit it, (ii) build the signal extraction filters, and (iii) compute the trends along with their standard errors. Our research describes (i), (ii), and (iii).

Unrelated Trends (LLM): Total PCE

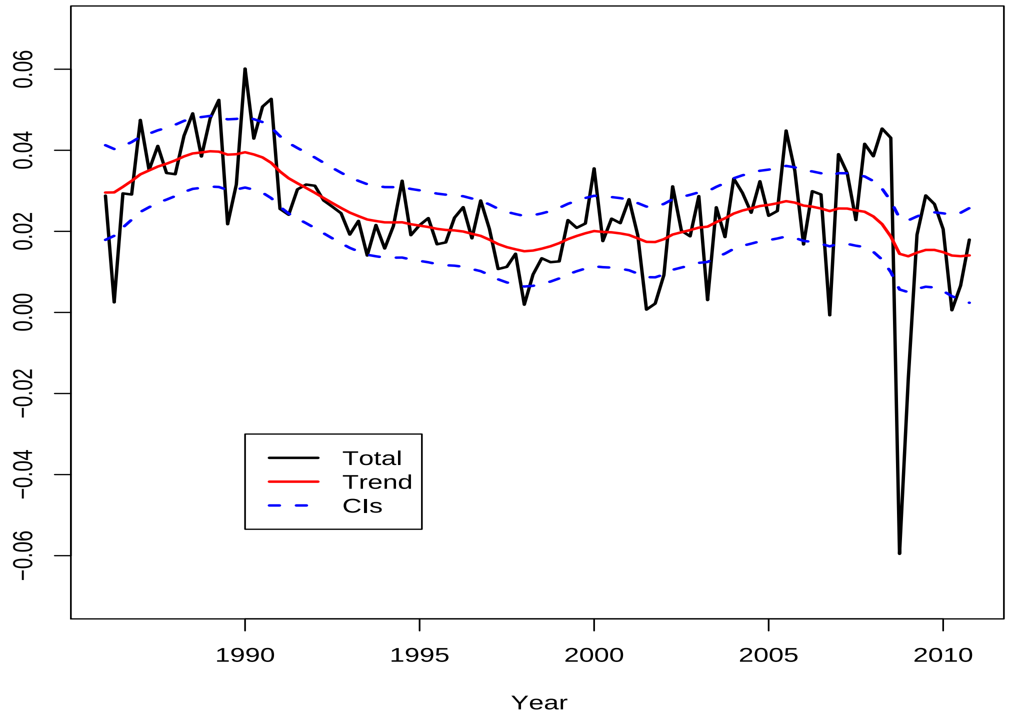

Figure 2: Trend estimates for the unrelated trends LLM model for Total inflation, with 2 SE confidence intervals.

Unrelated Trends (LLM): Core PCE

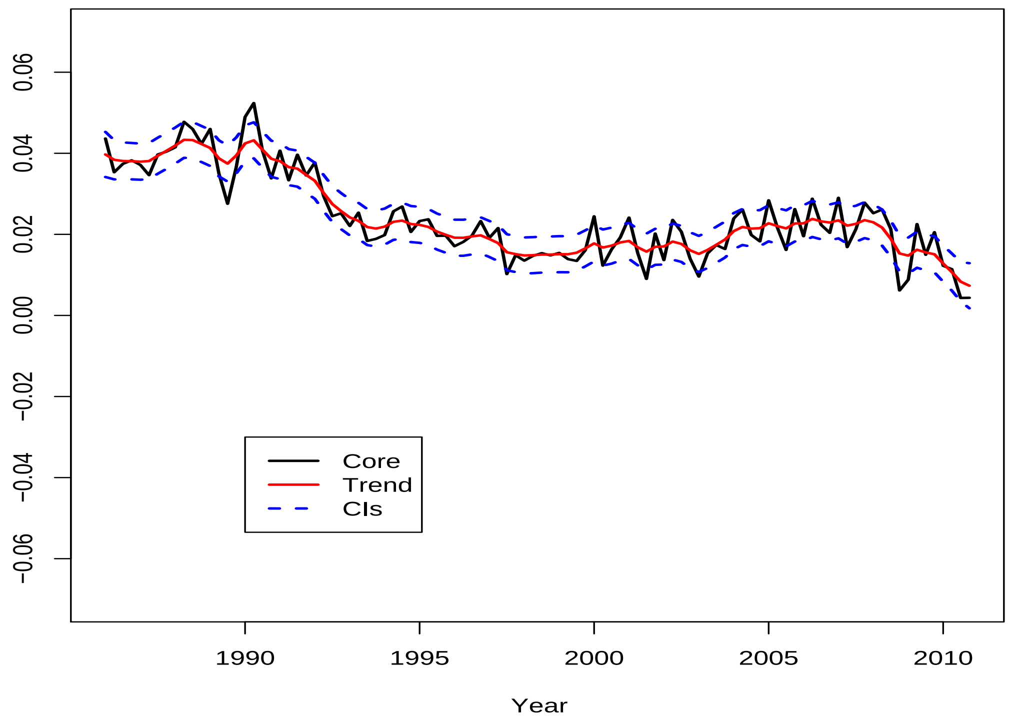

Figure 3: Trend estimates for the unrelated trends LLM model for Core inflation, with 2 SE confidence intervals.

Similar Trends (LLM): Total PCE

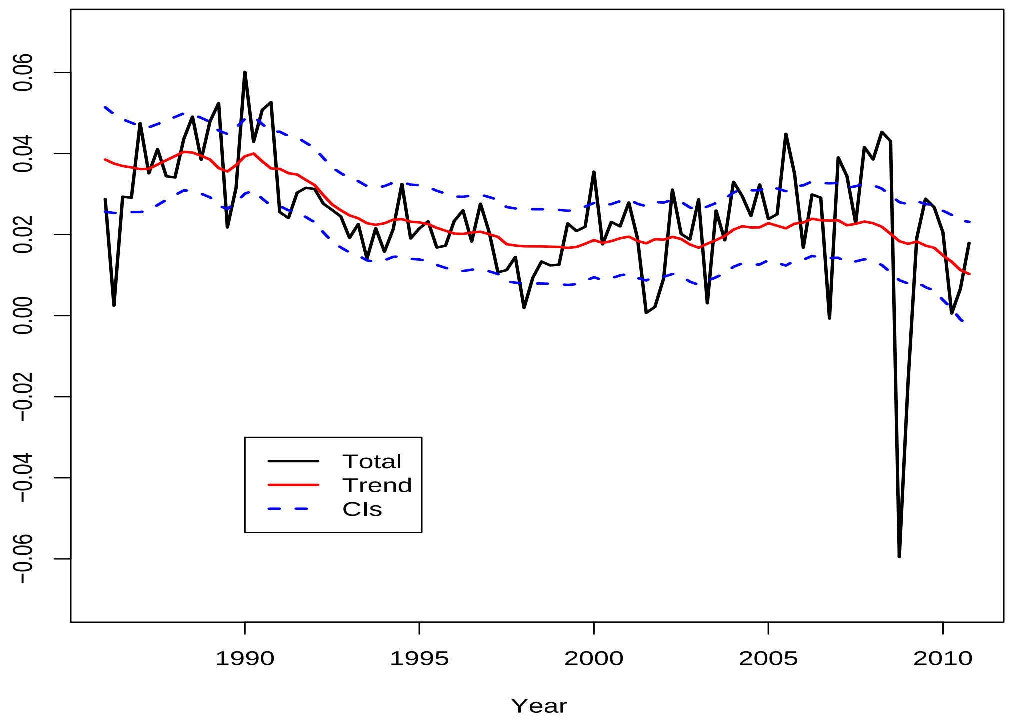

Figure 4: Trend estimates for the similar trends LLM model for Total inflation, with 2 SE confidence intervals.

Similar Trends (LLM): Core PCE

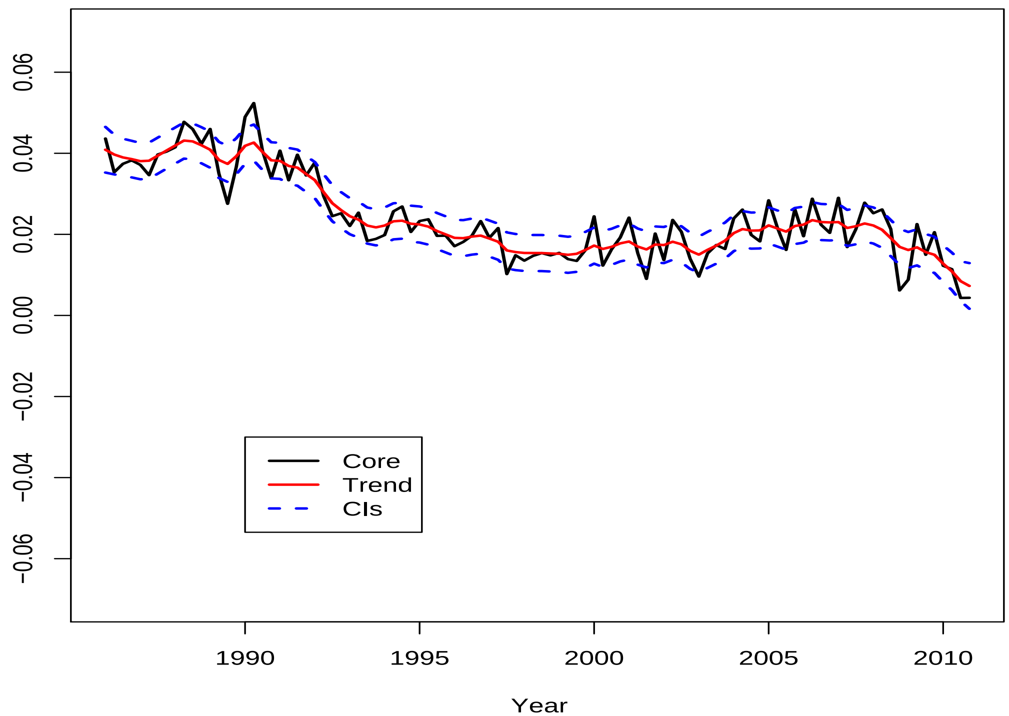

Figure 5: Trend estimates for the similar trends LLM model for Core inflation, with 2 SE confidence intervals.

Common Trends (LLM): Total PCE

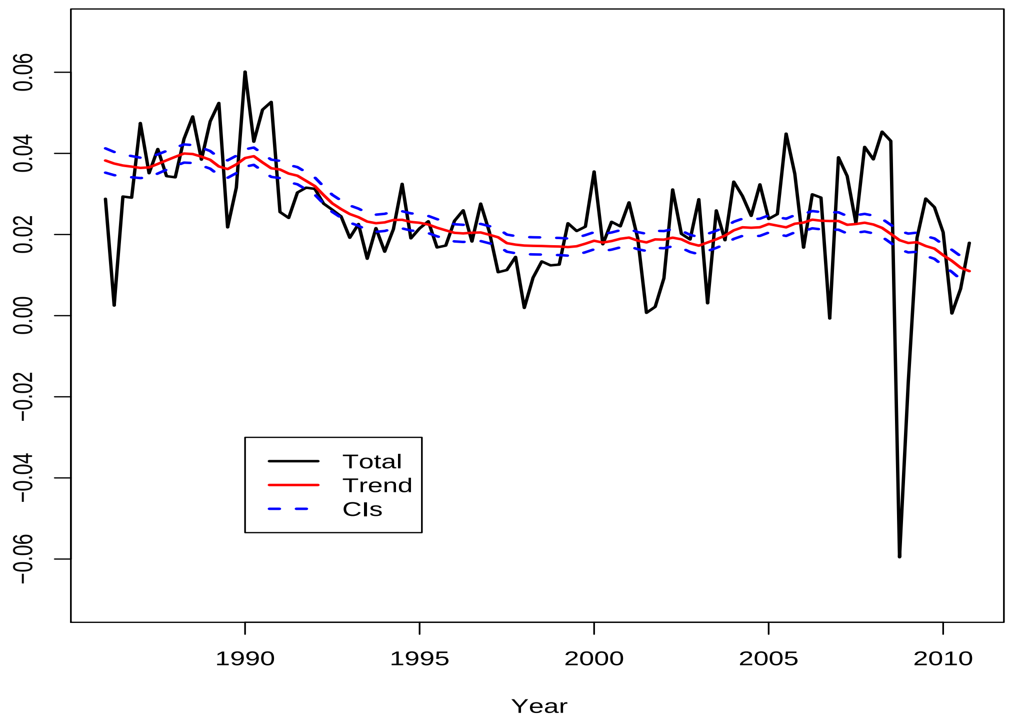

Figure 6: Trend estimates for the common trends LLM model for Total inflation, with 2 SE confidence intervals.

Common Trends (LLM): Core PCE

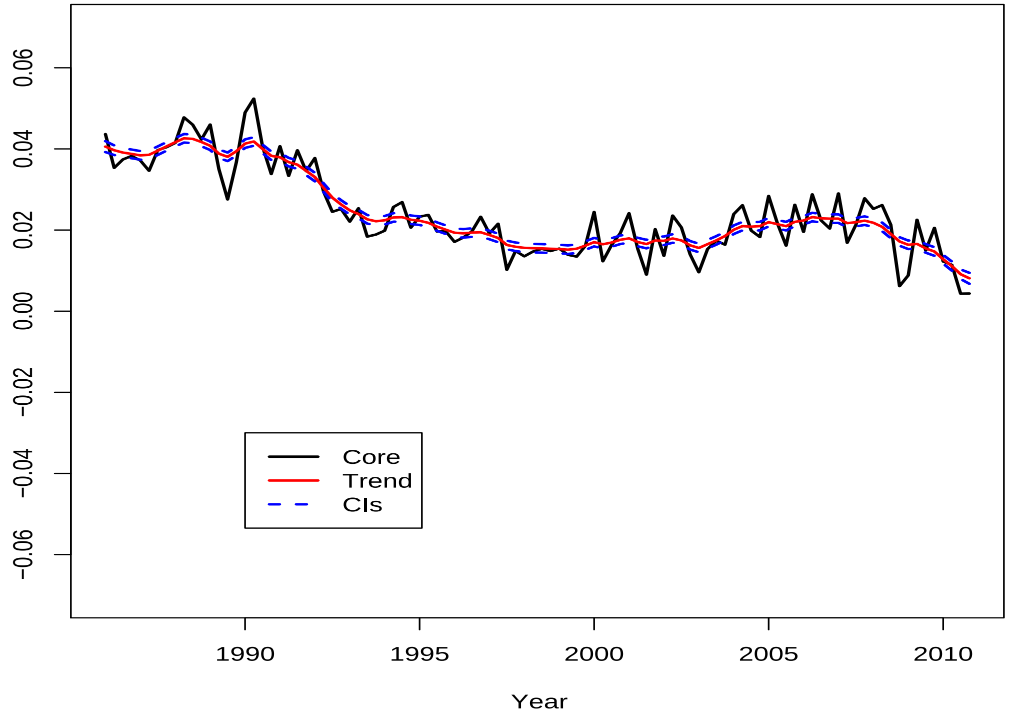

Figure 7: Trend estimates for the common trends LLM model for Core inflation, with 2 SE confidence intervals.

Unrelated Trends (LLM): Total and Core

Figure 8: Trend estimates for the unrelated trends LLM models for Total and Core inflation juxtaposed.

Similar Trends (LLM): Total and Core

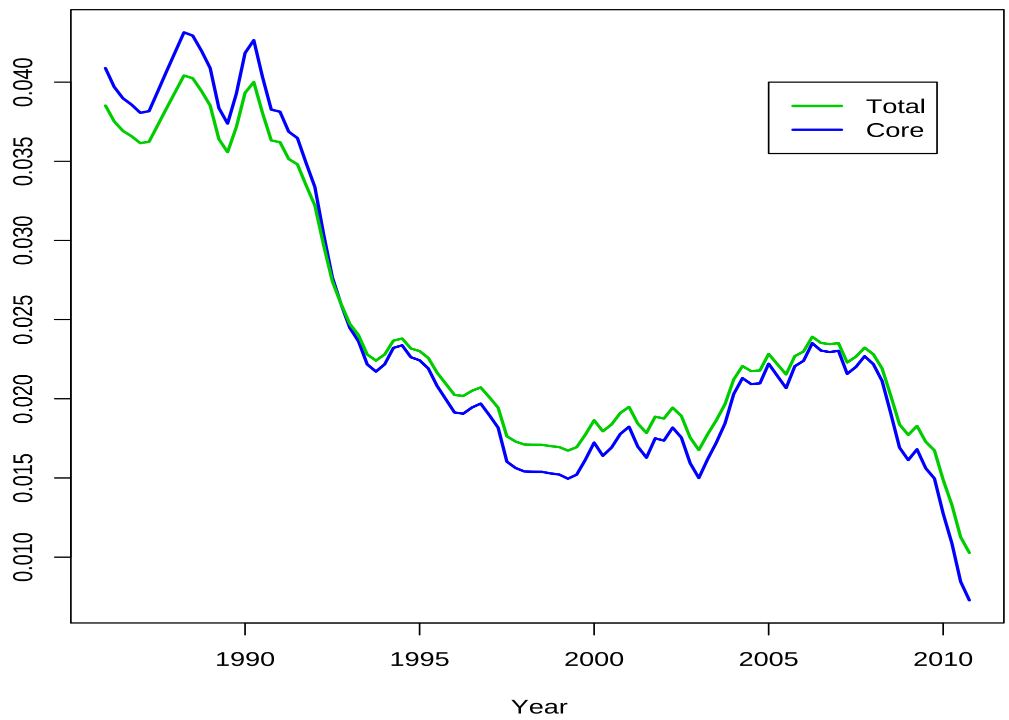

Figure 9: Trend estimates for the similar trends LLM models for Total and Core inflation juxtaposed.

Unrelated Trends (STM): Total PCE

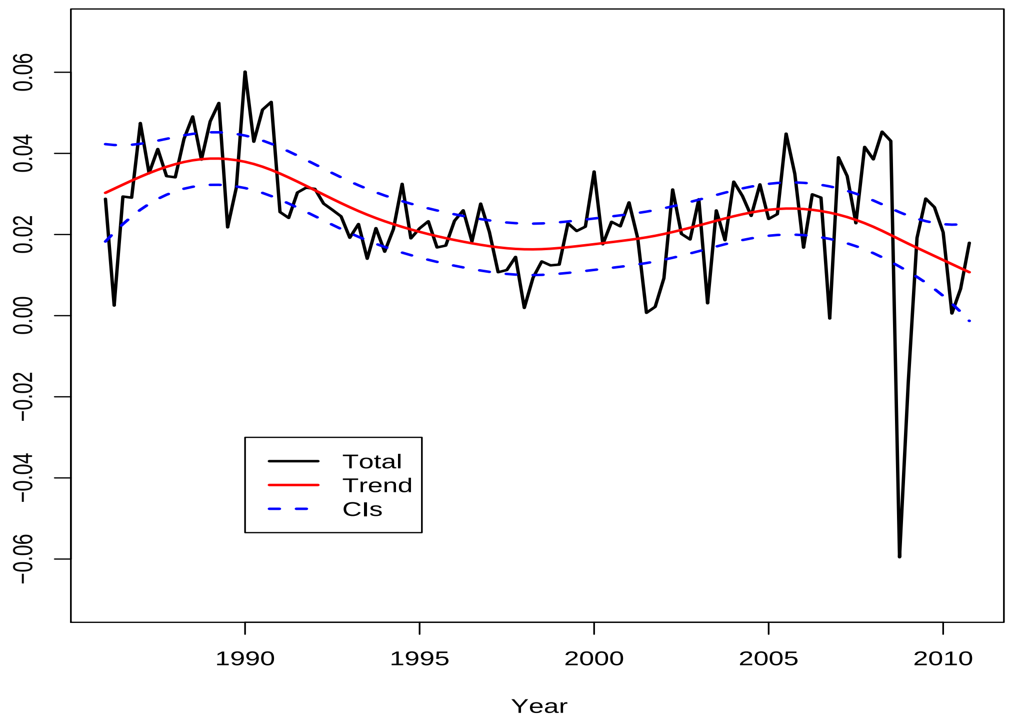

Figure 10: Trend estimates for the unrelated trends STM model for Total inflation, with 2 SE confidence intervals.

Unrelated Trends (STM): Core PCE

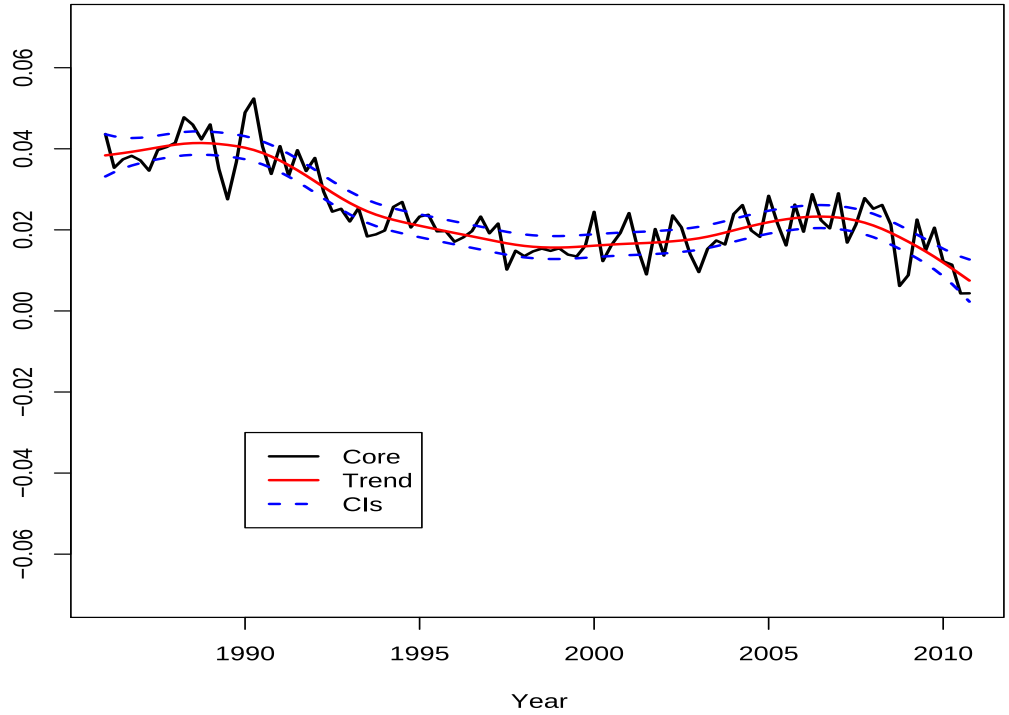

Figure 11: Trend estimates for the unrelated trends STM model for Core inflation, with 2 SE confidence intervals.

Similar Trends (STM): Total PCE

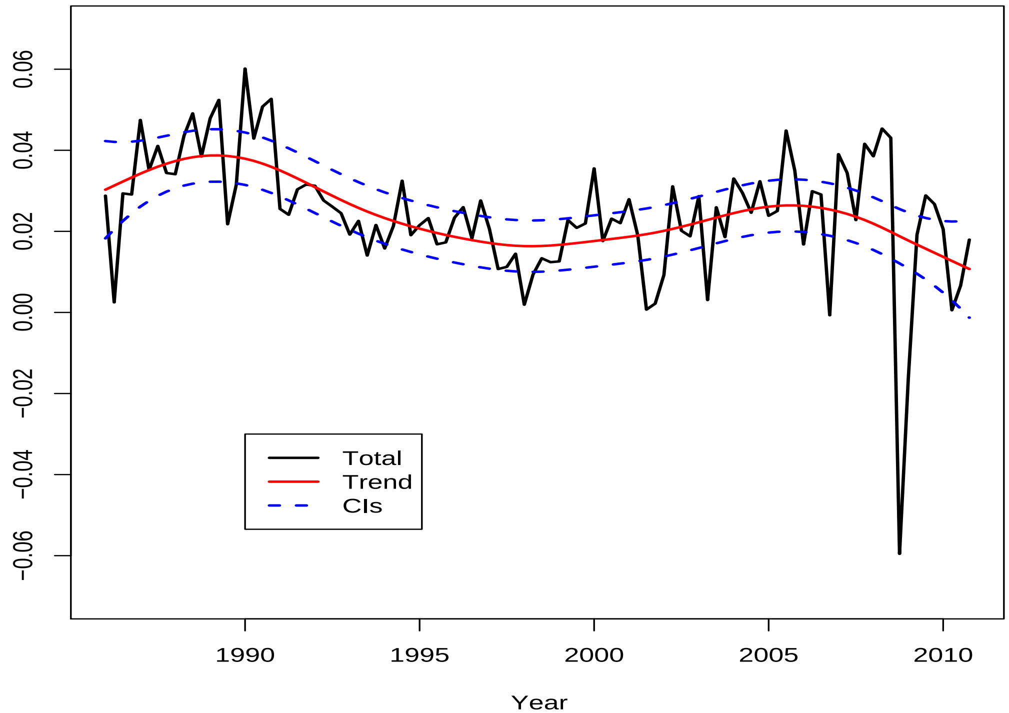

Figure 12: Trend estimates for the similar trends STM model for Total inflation, with 2 SE confidence intervals.

Similar Trends (STM): Core PCE

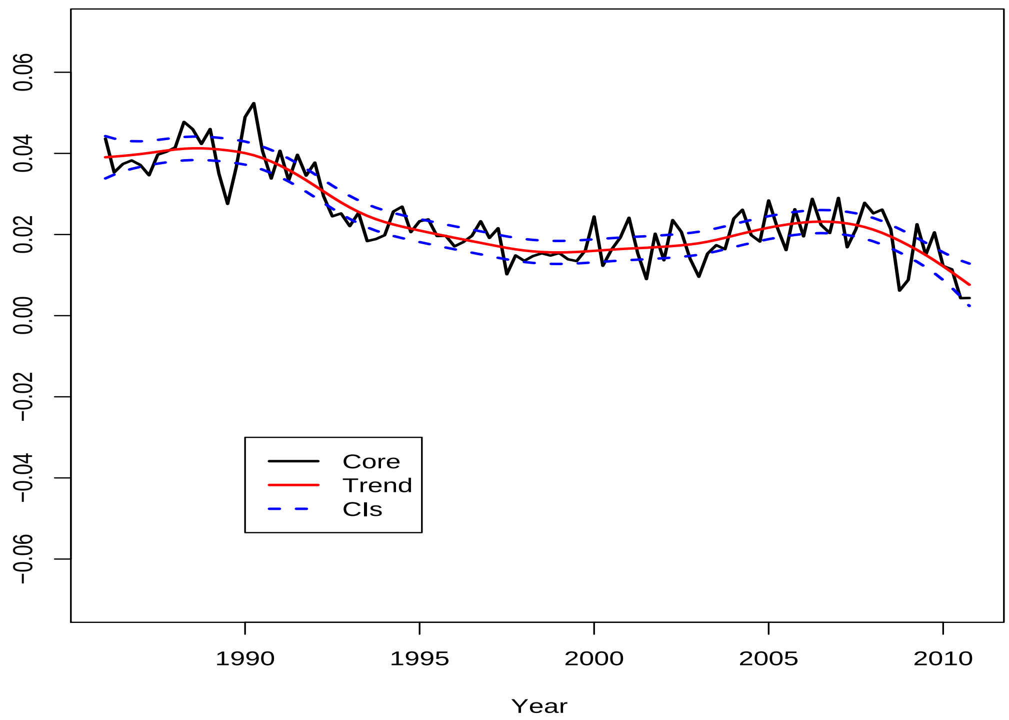

Figure 13: Trend estimates for the similar trends STM model for Core inflation, with 2 SE confidence intervals.

Common Trends (STM): Total PCE

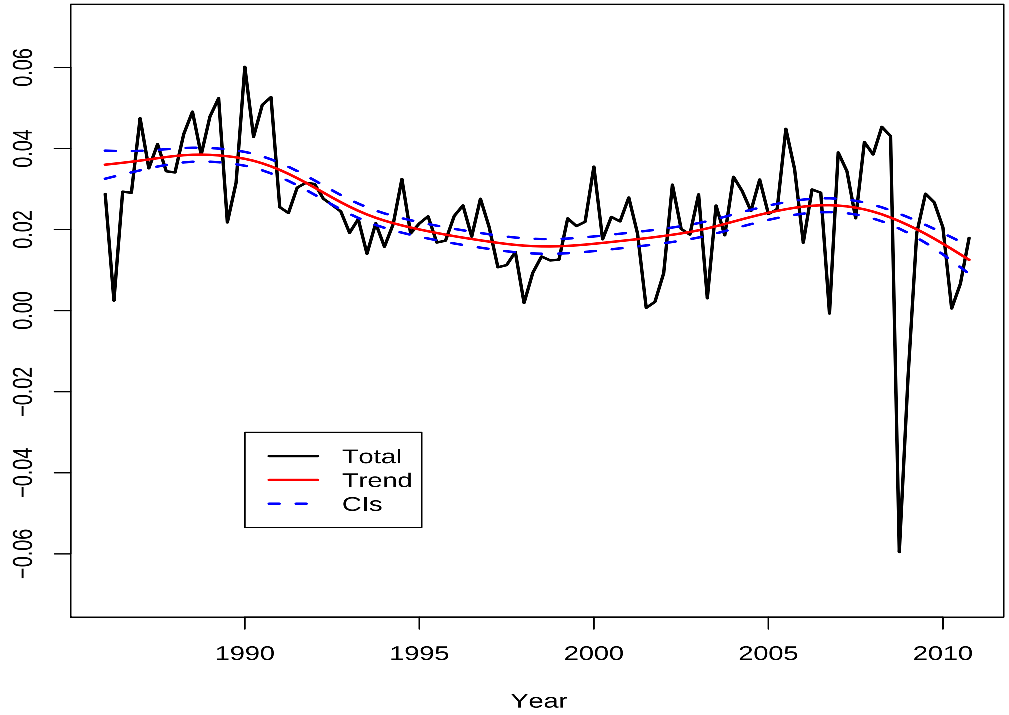

Figure 14: Trend estimates for the common trends STM model for Total inflation, with 2 SE confidence intervals.

Common Trends (STM): Core PCE

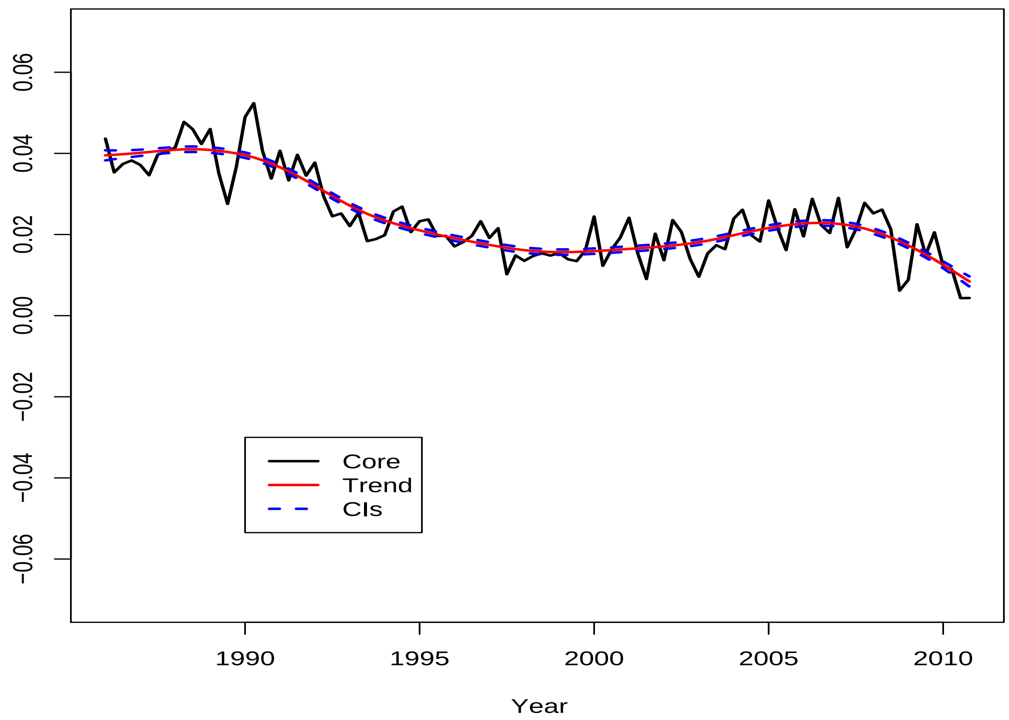

Figure 15: Trend estimates for the common trends STM model for Core inflation, with 2 SE confidence intervals.

Unrelated Trends (STM): Total and Core

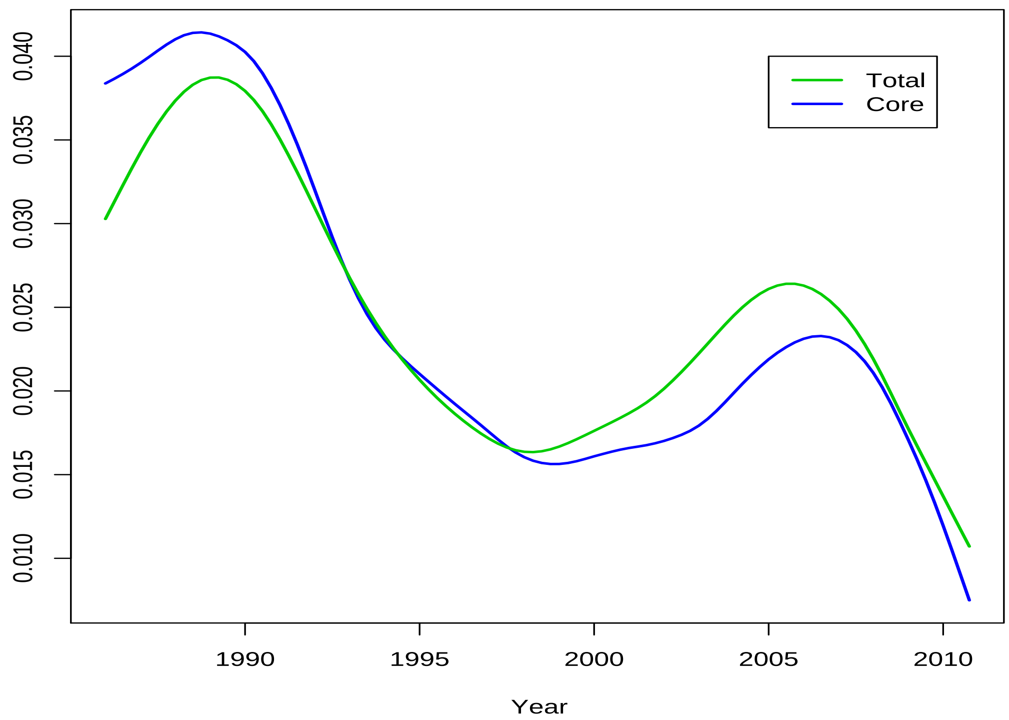

Figure 16: Trend estimates for the unrelated trends STM models for Total and Core inflation juxtaposed.

Similar Trends (STM): Total and Core

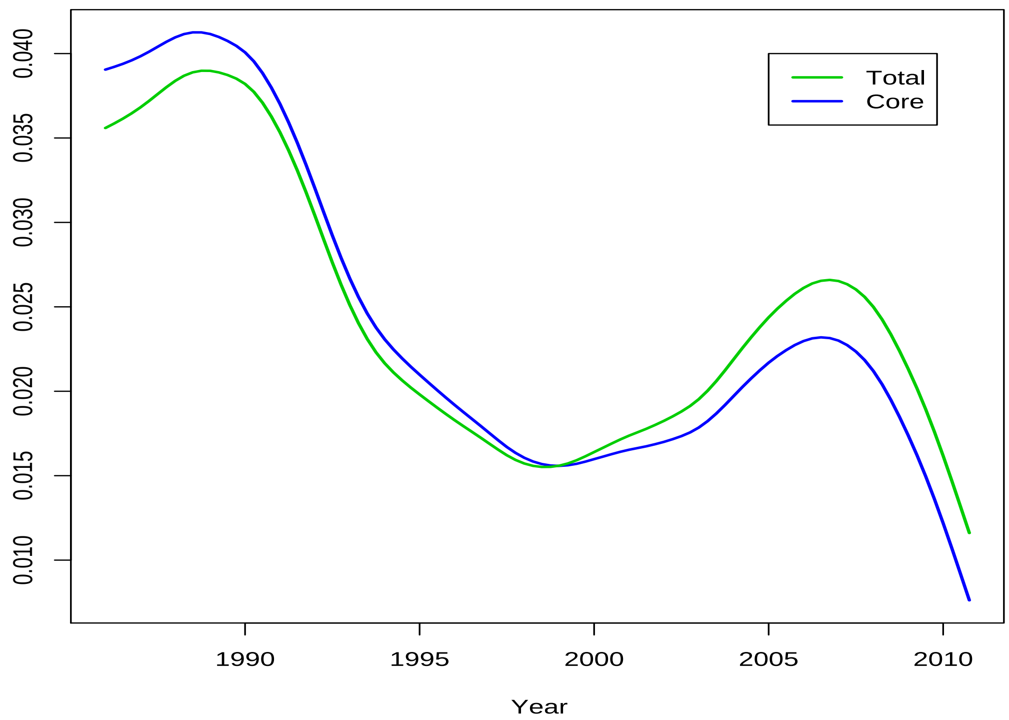

Figure 17: Trend estimates for the similar trends STM models for Total and Core inflation juxtaposed.

Common Trends (STM): Total and Core

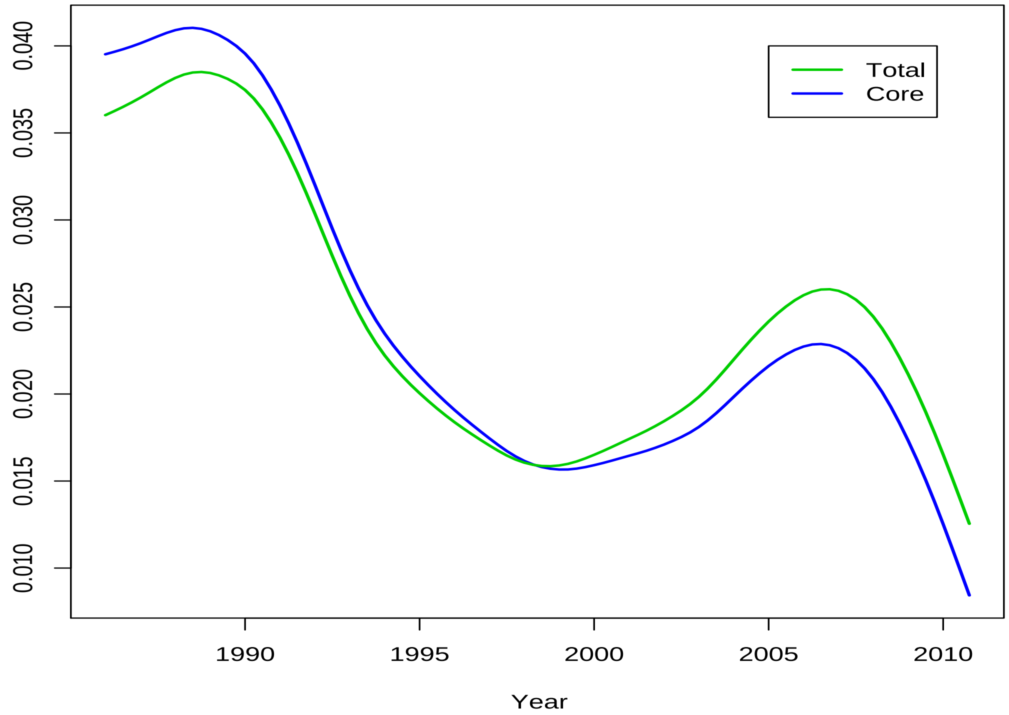

Figure 18: Trend estimates for the common trends STM models for Total and Core inflation juxtaposed.

Features of Inflation Trends

- Both trends and their confidence intervals (2 SE) are produced. Models allow us to quantify uncertainty.
- The SEs are much smaller for the common trends, vs. the similar and unrelated trends. This is because the correlations in similar trends models are almost full, so there is no loss in the simpler common trends model, and uncertainty is reduced.
- The trends are smoother in the STM vs. the LLM. The STM uses a twice-integrated random walk for the trend, while LLM uses a once-integrated random walk.
- The trend estimates for Total and Core behave similarly, depending on model.

Multivariate Signal Extraction

- We have signal S (trend in our case) and noise N for each of several time series. These signals may be cross-correlated among one another, but are independent of the noises.

- One way to think about this: say the vector signal process $\{S_t\}$ is driven by a vector innovation process, which is a multivariate white noise. Its covariance matrix Σ may be non-diagonal (implies similar signals) and may even have some zero eigenvalues (implies common signals, and non-full rank).

- Signal and noise processes are nonstationary, and they can be temporally differenced to a stationary process. E.g., a vector random walk.

Multivariate Filtering

Suppose we have N time series, or N components to our vector time series. A multivariate filter is written

$$\boldsymbol{F}(L) = \sum_{j=-\infty}^{\infty} \boldsymbol{W}_j L^j \tag{1}$$

where L is the standard lag operator, and $\boldsymbol{W}_j$ is the $N \times N$ matrix of coefficients for lag j. The cross-elements $\boldsymbol{W}_j^{IJ}$ and $\boldsymbol{W}_j^{JI}$ are generally unequal.

Multivariate Filtering

The bi-infinite assumption represents a useful hypothetical case for a theoretical analysis. Given $\{\mathbf{y}_t\} = \{\mathbf{y}_t, t = -\infty, \cdots, \infty\}$ an $N \times 1$ vector series, the filter produces output $\mathbf{z}_t$ as follows:

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

The filter output for each I equals a sum of N terms, each given by a weighting kernel applied to an element series. For $I = J$, we will call the profile of weights a self- or own-filter. We now have N input series for each output series, so there are N^2 filters to consider.

Spectral Representation

The spectral representation for a stationary multivariate time series involves a vector-valued orthogonal increments process $d\mathbf{Z}(\lambda)$ for frequencies $\lambda \in [-\pi, \pi]$:

$$\mathbf{y}_t = \int_{-\pi}^{\pi} e^{it\lambda} d\mathbf{Z}(\lambda) \qquad \mathbf{z}_t = \int_{-\pi}^{\pi} e^{it\lambda} \boldsymbol{F}(e^{-i\lambda}) d\mathbf{Z}(\lambda). \tag{3}$$

The quantity $\boldsymbol{F}(e^{-i\lambda})$ is the definition of the multivariate frequency response function (frf), equal to the Fourier Transform (FT) of the weights sequence $\boldsymbol{W}_j$.

Spectral Representation

A comparison of input and output in (3) indicates that the new orthogonal increments process for $\{\mathbf{z}_t\}$ is $\boldsymbol{F}(e^{-i\lambda})d\mathbf{Z}(\lambda)$; hence, the spectral density matrix of the output process is $\boldsymbol{F}(e^{-i\lambda})f(\lambda)\boldsymbol{F}'(e^{i\lambda})$. The diagonal entries of this matrix are the spectral densities of the component processes of $\{\mathbf{z}_t\}$, whereas the off-diagonal entries are cross-spectral densities, which are potentially complex-valued. If we examine the action on the Ith component output process, we have

$$\mathbf{z}_t^{(I)} = \sum_{J=1}^{N} \int_{-\pi}^{\pi} e^{it\lambda} \boldsymbol{F}_{IJ}(e^{-i\lambda}) d\mathbf{Z}_J(\lambda)$$

Autocovariance Generating Function

The spectral density f gives information about the second order structure of our vector time series. An equivalent tool is the multivariate autocovariance generating function (ACGF), which for any mean zero stationary series $\boldsymbol{x}_t$ is written as

$$\boldsymbol{G}_{\boldsymbol{x}}(L) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j L^j,$$

where $\boldsymbol{\Gamma}_j = E(\boldsymbol{x}_t \boldsymbol{x}_{t-j})$ is the covariance between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-j}$. Therefore, $\boldsymbol{G}_{\boldsymbol{x}}(L)$ contains information about the autocovariances of each component of the vector process, as well as the cross-covariances of the various elements at different lags. It is immediate that $\boldsymbol{G}_{\boldsymbol{x}}(e^{-i\lambda}) = f(\lambda)$.

VARMA Example

As an example, consider the stationary VARMA process (cf. Brockwell and Davis (1991)) written

$$\Phi(L)\boldsymbol{x}_t = \Theta(L)\kappa_t, \quad \kappa_t \sim WN(0, \Sigma_\kappa).$$

Then the ACGF is

$$\boldsymbol{G}_{\boldsymbol{x}}(L) = \Phi^{-1}(L)\Theta(L)\Sigma_\kappa\Theta'(L^{-1})\Phi^{\dagger}(L^{-1}), \tag{4}$$

where $\dagger$ stands for inverse transpose. The spectrum is given by substituting $L = e^{-i\lambda}$ in the above expression.

Signal Extraction Filters

Now suppose that the observed time series $\{\mathbf{y}_t\}$ can be decomposed in terms of unobserved signal and noise:

$$\mathbf{y}_t = \boldsymbol{s}_t + \boldsymbol{n}_t, \tag{5}$$

for $t = -\infty, ..., \infty$, where $\{\boldsymbol{s}_t\}$ and $\{\boldsymbol{n}_t\}$ are both stationary, have dimension $N \times 1$, and are uncorrelated with one another. The (classical) problem of multivariate signal extraction is to compute, for each I and at each time t, $\mathbb{E}[\mathbf{s}_t^{(I)} | \{\mathbf{y}_t\}]$, the estimate that minimizes the Mean Squared Error (MSE) criterion.

Signal Extraction Filters, Stationary Case

In the case that both the signal and noise processes are stationary, the optimal filter for extracting the signal vector (Gómez, 2006) is

$$\boldsymbol{F}_{WK}(L) = \boldsymbol{G}_{\boldsymbol{s}}(L)[\boldsymbol{G}_{\boldsymbol{s}}(L) + \boldsymbol{G}_{\boldsymbol{n}}(L)]^{-1}, \tag{6}$$

where WK stands for the Wiener-Kolmogorov filter (see Wiener (1949) for the classic univariate case). The filter for the noise is $\boldsymbol{G}_{\boldsymbol{n}}(L)[\boldsymbol{G}_{\boldsymbol{s}}(L) + \boldsymbol{G}_{\boldsymbol{n}}(L)]^{-1}$, which is $1_N - \boldsymbol{F}_{WK}(L)$ – here and throughout 1_N denotes an $N \times N$ identity matrix.

Signal Extraction Filters, Stationary Case

Now (6) gives the time-domain characterization. To convert to the frequency domain, substitute $e^{-i\lambda}$ for L, which then produces the WK frf:

$$\boldsymbol{F}_{WK}(e^{-i\lambda}) = \boldsymbol{G_s}(e^{-i\lambda})[\boldsymbol{G_s}(e^{-i\lambda}) + \boldsymbol{G_n}(e^{-i\lambda})]^{-1},$$

where the quantities $\boldsymbol{G_s}(e^{-i\lambda})$ and $\boldsymbol{G_n}(e^{-i\lambda})$ are the multivariate spectral densities of signal and noise, respectively. Clearly, the multivariate WK filter depends on the relationships between component series, expressed in the cross-correlations of components, as well as on the individual dynamic properties within each series.

Signal Extraction Filters, Non-stationary Case

Now suppose that signal and noise are difference-stationary processes: for the Jth observed process $\{\mathbf{y}_t^{(J)}\}$ suppose there exists an order d^J polynomial $\delta^{(J)}$ in the lag operator L such that $\{\mathbf{w}_t^{(J)}\} = \{\delta^{(J)}(L)\mathbf{y}_t^{(J)}\}$ is covariance stationary. We suppose similarly that there are signal and noise differencing polynomials $\delta_{\mathbf{s}}^{(J)}$ and $\delta_{\mathbf{n}}^{(J)}$ that render them stationary, so that $\{\mathbf{u}_t^{(J)}\} = \{\delta_{\mathbf{s}}^{(J)}(L)\mathbf{s}_t^{(J)}\}$ and $\{\mathbf{v}_t^{(I)} = \delta_{\mathbf{n}}^{(J)}(L)\mathbf{n}_t^{(J)}\}$ have an autocovariance function well-defined at all lags and a spectrum that exists at all frequencies. When the signal and noise differencing operators do not depend on J, so that they are the same for each series (though they still differ for signal versus noise), we refer to this situation as "uniform differencing operators."

Signal Extraction Filters, Non-stationary Case

Let $f_{\mathbf{u}}^{IJ}$, $f_{\mathbf{v}}^{IJ}$, and $f_{\mathbf{w}}^{IJ}$ denote the cross-spectral density functions for the Ith and Jth processes for the signal, noise, and observed processes, respectively. We suppose that the set Ω of frequencies where $f_{\mathbf{w}}$ is noninvertible has Lebesgue measure zero. Define the so-called "over-differenced" processes given by

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

These occur when the full-differencing operator $\delta^{(I)}(L)$ is applied to signal plus noise, resulting in a covariance stationary process with zeroes in its spectral density.

Signal Extraction Filters, Non-stationary Case

We require some technical assumptions. First, each nonstationary process $\{\mathbf{y}_t^{(I)}\}$ can be generated from d^I initial values $\mathbf{y}_*^{(I)}$ together with the difference-stationary process $\{\mathbf{w}_t^{(I)}\}$, for each I, in the manner elucidated in Bell (1984). The information contained in $\{\mathbf{y}_t^{(I)}\}$ is equivalent to that in $\{\mathbf{w}_t^{(I)}\} \cup \mathbf{y}_*^{(I)}$ for the purposes of linear projection, since the former is expressible as a linear transformation of the latter, for each I.

Signal Extraction Filters, Non-stationary Case

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

We also assume that the vector processes $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$ are uncorrelated with one another.

Signal Extraction Filters, Non-stationary Case

Set $z = e^{-i\lambda}$ and $\overline{z} = e^{i\lambda}$, and consider the filter $\Psi(L)$ defined as follows: it has frf defined via the formula

$$\Psi(z) = \mathbf{G}_{\partial \mathbf{u}}(z) \mathbf{G}_{\mathbf{w}}^{-1}(z) = f_{\partial \mathbf{u}}(\lambda) \, f_{\mathbf{w}}^{-1}(\lambda) \tag{7}$$

for $\lambda \in \Omega$, and by the limit of such for $\lambda \notin \Omega$. Then when differencing operators are uniform, the optimal estimate of the signal at time t is given by $\widehat{\mathbf{s}}_t = \Psi(L) \mathbf{y}_t$.

Signal Extraction Filters, Non-stationary Case

Since $\mathbf{G_w} = \mathbf{G}_{\partial \mathbf{u}} + \mathbf{G}_{\partial \mathbf{v}}$, (7) generalizes (6) to the nonstationary case. If some of the differencing polynomials are unity (i.e., no differencing is required to produce a stationary series), the formula collapses down to the classical case. In the extreme case that all the series are stationary, trivially $\partial \mathbf{u}_t = \mathbf{s}_t$ and $\partial \mathbf{v}_t = \mathbf{n}_t$ for all times t. The second expression for the frf in (7) shows how this is a direct multivariate generalization of the univariate frf in Bell (1984), which has the formula $|\delta_n(z)|^2 f_u(\lambda) / f_w(\lambda)$.

How Does it Apply?

This result is worded so as to include the important case of co-integrated vector time series (Engle and Granger, 1987). The main issue: for these processes, there exist a finite set of λ for which $f_{\mathbf{w}}(\lambda)$ is singular. Since the problem occurs on a set of measure zero, it has no impact on defining the optimal filters.

We next show that the key assumptions on the structure of $\Psi(z)$ are satisfied for a very wide class of co-integrated processes. We present our discussion in the context of uniform differencing operators.

Filters for Co-Integrated Processes

Assume the vector signal and noise processes satisfy $\delta_{\mathbf{s}}(L)\mathbf{s}_t = \mathbf{u}_t$ and $\delta_{\mathbf{n}}(L)\mathbf{n}_t = \mathbf{v}_t$, and that

$$\mathbf{u}_t = \Xi(L)\boldsymbol{\zeta}_t \qquad \mathbf{v}_t = \Omega(L)\boldsymbol{\kappa}_t, \tag{8}$$

where $\{\boldsymbol{\zeta}_t\}$ and $\{\boldsymbol{\kappa}_t\}$ are multivariate white noise processes. The MA filters $\Xi(z)$ and $\Omega(z)$ but are linear and invertible by assumption.

We assume that the white noise covariance matrices $\Sigma_{\boldsymbol{\zeta}}$ and $\Sigma_{\boldsymbol{\kappa}}$ are non-negative definite, but whereas $\Sigma_{\boldsymbol{\kappa}}$ is positive definite, some zero eigenvalues are present in $\Sigma_{\boldsymbol{\zeta}}$.

Filters for Co-Integrated Processes

Result: Suppose that the differencing operators are common and that the disturbances follow (8). Also suppose that $\Sigma_{\boldsymbol{\kappa}}$ is positive definite, but that some of the eigenvalues of $\Sigma_{\boldsymbol{\zeta}}$ are zero. Then $\Psi(z)$ defined in (7) can be continuously extended from its natural domain Ω to all of $[-\pi, \pi]$.

Filters for Co-Integrated Processes

Hence the formula for the WK frf is well-defined and (7) can be used to give a compact expression for the filter, formally substituting L for $z = e^{-i\lambda}$:

$$\Psi(L) = \mathbf{G}_{\partial\mathbf{u}}(L) \left[\mathbf{G}_{\partial\mathbf{u}}(L) + \mathbf{G}_{\partial\mathbf{v}}(L)\right]^{-1}. \tag{9}$$

This expresses the filter in terms of the ACGFs of the over-differenced signal and noise processes.

The components of the frf, when real-valued, are interpretable as "gain" functions, multiplying the input spectral density.

Gain Graphs

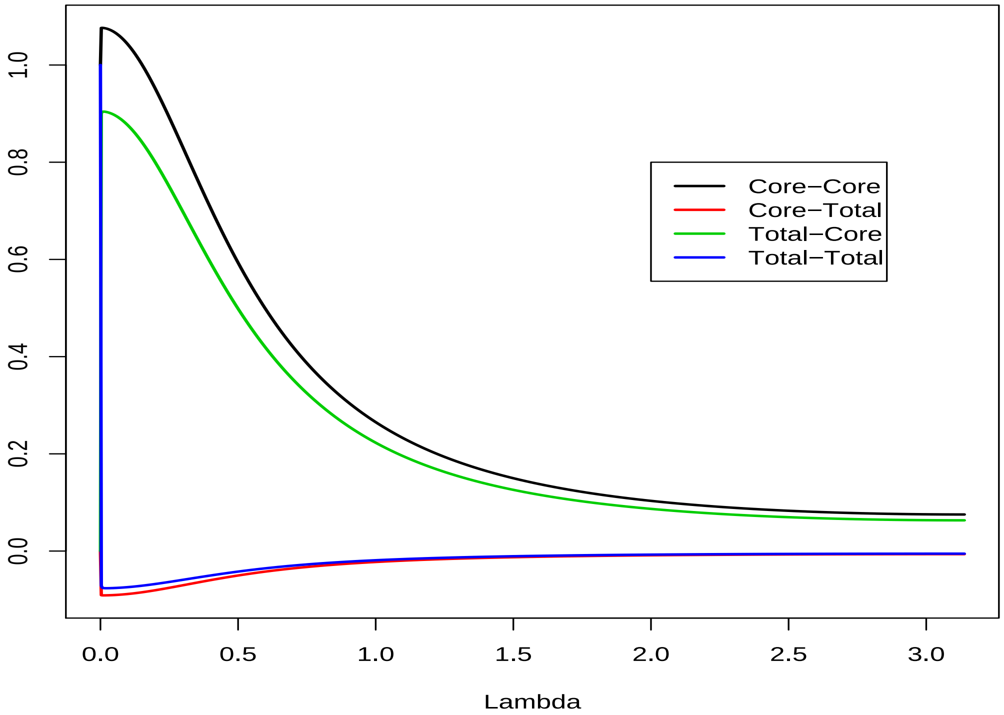

Figure 19: Gain plots for Core and Total. The legend reads with output series on the left, input series on the right. Model is LLM similar trends.

Gain Graphs

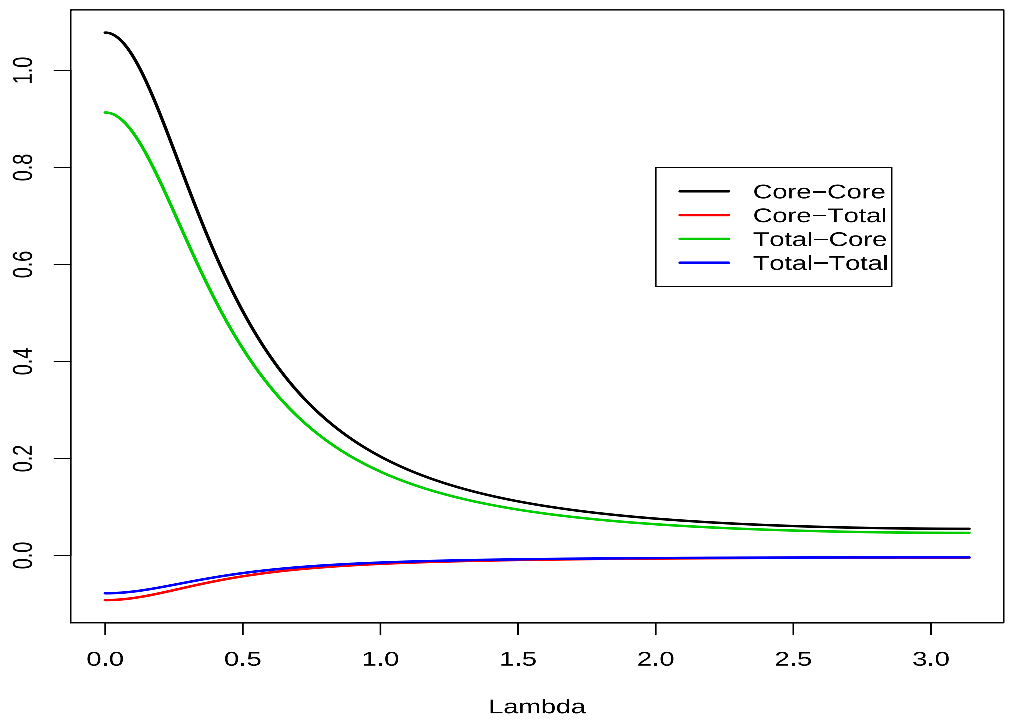

Figure 20: Gain plots for Core and Total. The legend reads with output series on the left, input series on the right. Model is LLM common trends.

Gain Graphs

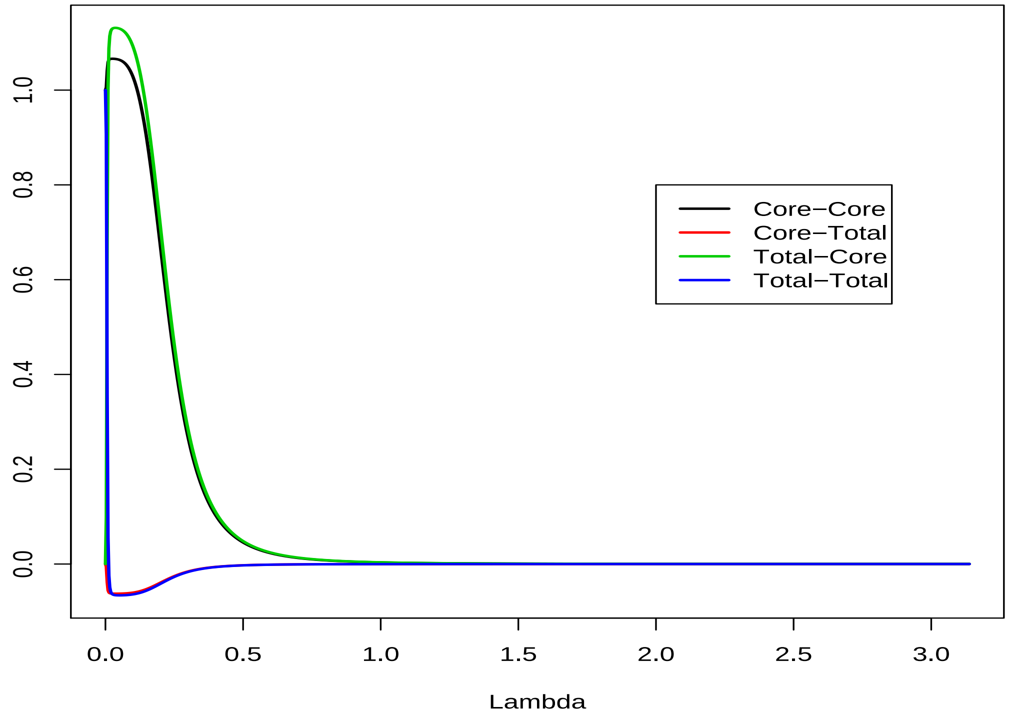

Figure 21: Gain plots for Core and Total. The legend reads with output series on the left, input series on the right. Model is STM similar trends.

Gain Graphs

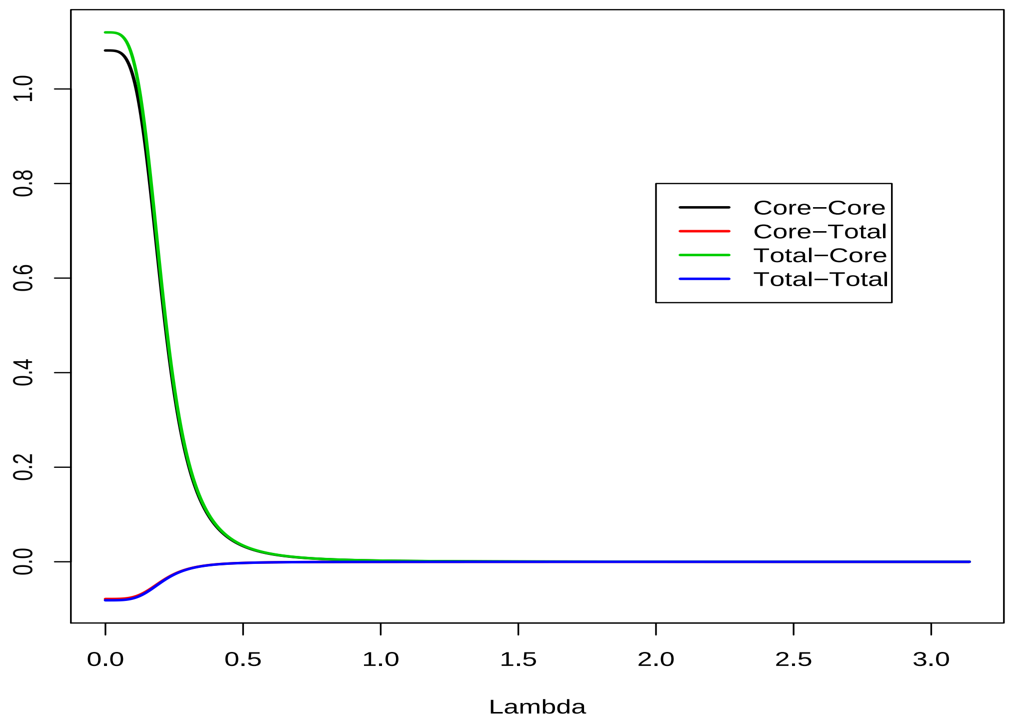

Figure 22: Gain plots for Core and Total. The legend reads with output series on the left, input series on the right. Model is STM common trends.

Summary of Gains

- All frfs are real-valued for these models, and hence equal to Gain.

- Weird behavior at frequency zero for similar trends models. This is due to ill-conditioning in the trend innovation variance matrix – the correlation is over $.999$ in both cases! Nicer behavior in common trends model, because collinearity in parameters has been accounted for in the model.

- Core is primarily determining the trends for both Core and Total. But Total is less relevant, and has a negative effect.

- All four filters are smoothers, since they die towards zero as frequency increases.

Finite versus Infinite Sample

- So far we have looked at the filtering of a bi-infinite sample. Then the filters can be examined in terms of frf, which is useful for understanding the long-term properties of a filter.

- In practice we want time-varying finite-sample filters, i.e., matrices acting on our data vector. Then we can get trends like at the start of the talk. The main issue is handling beginning and end of sample.

- So we now proceed to exposit this. All this can be done in a State Space Formulation (SSF), but we provide explicit formulas; plus there is no confusion about initializing the Kalman filter – matrix formulas are exact mathematically and fast in practice. Also more general than SSF, since they can handle long memory processes.

Signal Extraction, Finite Sample

Consider N time series $\{\mathbf{y}_t^{(I)}\}$ for $1 \leq I \leq N$. Furthermore, each series can be written as the sum of unobserved signal and noise components, denoted $\{\mathbf{s}_t^{(I)}\}$ and $\{\mathbf{n}_t^{(I)}\}$, such that (5) holds for all t. While previously we considered t unbounded in both directions, here we suppose the sample consists of $t = 1, 2, \cdots, T$. We will express the samples of each series as a length-T vector, namely $\mathbf{y}^{(I)} = [y_1^{(I)}, y_2^{(I)}, ..., y_T^{(I)}]'$, and similarly for signal, $\mathbf{s}^{(I)}$, and noise, $\mathbf{n}^{(I)}$. For each I, the estimate that minimizes the MSE criterion is $\mathbb{E}[\mathbf{s}^{(I)} | \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(N)}]$. If the samples are Gaussian then it suffices to consider linear estimators of the signal. For a general distribution, we can instead seek optimal $linear$ MSE estimators.

Signal Extraction, Finite Sample

So our estimate $\widehat{\mathbf{s}}^{(I)}$ can be expressed as a $T \times (NT)$ matrix acting on all the data vectors stacked up, or equivalently as

$$\widehat{\mathbf{s}}^{(I)} = \sum_{J=1}^{N} F^{IJ} \mathbf{y}^{(J)}.$$

Each matrix F^{IJ} is $T \times T$ dimensional.

Our task is to compute the entries of F^{IJ} such that the MSE, i.e., the covariance matrix of $\widehat{\mathbf{s}}^{(I)} - \mathbf{s}^{(I)}$, is minimized over the class of estimators that are linear in the data.

Signal Extraction, Finite Sample

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 $\mathbf{w}^{(J)} = [\mathbf{w}^{(J)}_{d^J+1}, \cdots, \mathbf{w}^{(J)}_T]'$). These vectors may be correlated with one another and among themselves, which is summarized in the notation $\mathbb{E}[\mathbf{w}^{(I)}\mathbf{w}^{(J)'}] = \Sigma^{IJ}_{\mathbf{w}}$. We further suppose that the differencing is taken such that all random vectors have mean zero. Note that this definition includes processes that are nonstationary only in second moments, i.e., heteroskedastic.

Signal Extraction, Finite Sample

Extending to signal and noise components, we form the matrices $\Delta_{\mathbf{s}}^{(J)}$ and $\Delta_{\mathbf{n}}^{(J)}$ corresponding to the signal and noise differencing polynomials $\delta_{\mathbf{s}}^{(J)}$ and $\delta_{\mathbf{n}}^{(J)}$. Let $\mathbf{u}^{(J)} = \Delta_{\mathbf{s}}^{(J)} \mathbf{s}^{(J)}$ and $\mathbf{v}^{(J)} = \Delta_{\mathbf{n}}^{(J)} \mathbf{n}^{(J)}$, with cross-covariance matrices denoted $\Sigma_{\mathbf{u}}^{IJ}$ and $\Sigma_{\mathbf{v}}^{IJ}$. Now assume there are no common roots among $\delta_{\mathbf{s}}^{(J)}$ and $\delta_{\mathbf{n}}^{(J)}$, so that $\delta^{(J)}(L) = \delta_{\mathbf{s}}^{(J)}(L) \delta_{\mathbf{n}}^{(J)}(L)$. Then

$$\Delta^{(J)} = \underline{\Delta}_{\mathbf{n}}^{(J)} \Delta_{\mathbf{s}}^{(J)} = \underline{\Delta}_{\mathbf{s}}^{(J)} \Delta_{\mathbf{n}}^{(J)}, \tag{10}$$

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

United States™ Census Bureau

U.S. DEPARTMENT OF COMMERCE BUREAU OF THE CENSUS

Signal Extraction, Finite Sample

It follows that

$$\mathbf{w}^{(J)} = \Delta^{(J)} \mathbf{y}^{(J)} = \underline{\Delta}_{\mathbf{n}}^{(J)} \mathbf{u}^{(J)} + \underline{\Delta}_{\mathbf{s}}^{(J)} \mathbf{v}^{(J)}, \tag{11}$$

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

$$\Sigma_{\mathbf{w}}^{IJ} = \underline{\Delta}_{\mathbf{n}}^{(I)} \Sigma_{\mathbf{u}}^{IJ} \underline{\Delta}_{\mathbf{n}}^{(J)'} + \underline{\Delta}_{\mathbf{s}}^{(I)} \Sigma_{\mathbf{v}}^{IJ} \underline{\Delta}_{\mathbf{s}}^{(J)'}.$$

Signal Extraction, Finite Sample

We can splice all these $\Sigma_{\mathbf{w}}^{IJ}$ 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 $\Sigma_{\mathbf{w}}$. Further, the Gaussian likelihood function for the differenced sample involves the quadratic form $\mathbf{w}'\Sigma_{\mathbf{w}}^{-1}\mathbf{w}$, so parameter estimation on this basis also requires a invertible covariance matrix. For signal-noise decompositions and homoskedastic disturbances, the invertibility of $\Sigma_{\mathbf{w}}$ is guaranteed (derivation in the paper).

Some Technical Assumptions

For the signal extraction formula below, we require a few additional assumptions: let $\Sigma_{\mathbf{u}}^{JJ}$ and $\Sigma_{\mathbf{v}}^{JJ}$ be invertible matrices for each J, assume that $\mathbf{u}^{(I)}$ and $\mathbf{v}^{(J)}$ are uncorrelated with one another for all I, J, and suppose that the initial values of $\mathbf{y}^{(I)}$ are uncorrelated with $\mathbf{u}^{(J)}$ and $\mathbf{v}^{(J)}$ for all J. These initial values consist of all the first d^I values of each series $\mathbf{y}^{(I)}$. This type of assumption is less stringent than M_∞ of the previous subsection, and will be called Assumption M_T instead.

Assumption $\boldsymbol{M_T}$. Suppose that, for each $I = 1, 2, \cdots, N$, the initial values of $\mathbf{y}^{(I)}$ are uncorrelated with $\mathbf{u}$ and $\mathbf{v}$.

Signal Extraction, Finite Sample

We use the notation $\widetilde{G}$ for a block-matrix of the same dimension as a given block-matrix G, where $\widetilde{G}$ consists of only the block-diagonal matrices. Let

$$M^{II} = \Delta_{\mathbf{n}}^{(I)\prime} \Sigma_{\mathbf{v}}^{II^{-1}} \Delta_{\mathbf{n}}^{(I)} + \Delta_{\mathbf{s}}^{(I)\prime} \Sigma_{\mathbf{u}}^{II^{-1}} \Delta_{\mathbf{s}}^{(I)}.$$

Fact: M^{II} is invertible.

Signal Extraction, Finite Sample

With $\widehat{\mathbf{s}} = F\mathbf{y}$, a compact matrix formula for F is given as follows. Define block-matrices A, B, C, D that have IJth block matrix entries given, respectively, by

$$
\begin{aligned}
A^{IJ} &= \Delta_{\mathbf{s}}^{(I)'} {\Sigma_{\mathbf{u}}^{II}}^{-1} \Sigma_{\mathbf{u}}^{IJ} {\Sigma_{\mathbf{u}}^{JJ}}^{-1} \Delta_{\mathbf{s}}^{(J)} \\
B^{IJ} &= \Delta_{\mathbf{n}}^{(I)'} {\Sigma_{\mathbf{v}}^{II}}^{-1} \Sigma_{\mathbf{v}}^{IJ} {\Sigma_{\mathbf{v}}^{JJ}}^{-1} \Delta_{\mathbf{n}}^{(J)} \\
C^{IJ} &= \Delta_{\mathbf{s}}^{(I)'} {\Sigma_{\mathbf{u}}^{II}}^{-1} \Sigma_{\mathbf{u}}^{IJ} \underline{\Delta}_{\mathbf{n}}^{(J)'} \\
D^{IJ} &= \Delta_{\mathbf{n}}^{(I)'} {\Sigma_{\mathbf{v}}^{II}}^{-1} \Sigma_{\mathbf{v}}^{IJ} \underline{\Delta}_{\mathbf{s}}^{(J)'}.
\end{aligned}
$$

Signal Extraction, Finite Sample

Also let $\widetilde{\Delta}$ denote a block diagonal matrix with the matrix $\Delta^{(I)}$ in the Ith diagonal. Then

$$\begin{aligned} M &= \widetilde{A} + \widetilde{B} \\ F &= M^{-1} \left[\widetilde{B} + (C - D)\, \Sigma_{\mathbf{w}}^{-1} \widetilde{\Delta} \right] \\ V &= A + B + (C - D) \Sigma_{\mathbf{w}}^{-1} (C - D)', \end{aligned}$$

and the covariance matrix of the error vector is $M^{-1} V M^{-1}$.

Commentary

These formulas tell us exactly how each series $\mathbf{y}^{(J)}$ contributes to the component estimate $\widehat{\mathbf{s}}^{(I)}$. When there is no cross-series information, i.e., $\Sigma_{\mathbf{u}}^{IJ}$ and $\Sigma_{\mathbf{v}}^{IJ}$ are zero for $I \neq J$, then clearly C and D are zero, and F reduces to an N-fold stacking of the univariate filter ($M^{-1}\widetilde{B}$ is just the stacking of the univariate matrix filters of McElroy (2008)).

When $C \neq D$, there is cross-series information entering into the filters, which also increases the uncertainty.

Commentary

Since $\Sigma_{\mathbf{w}}$ is invertible for processes consisting of co-integrated signal and non-co-integrated noise (or vice versa), this indicates that maximum likelihood estimation is viable; the Gaussian log likelihood is -2 times

$$\mathbf{w}' \Sigma_{\mathbf{w}}^{-1} \mathbf{w} + \log |\Sigma_{\mathbf{w}}|,$$

up to irrelevant constants, once we factor out the initial value vectors utilizing Assumption M_T. It is interesting that the Whittle likelihood is not well-defined since $f_{\mathbf{w}}$ is not invertible at all frequencies.

Implementation

- For the particular LLM and STM models (more details below), we have R code to fit via mle.

- Matrix formulas for signal extraction are very fast for moderate time series, and code is also written in R. All figures were produced this way.

- We also did all the same calculations in Ox, using SSF, to check our results. Initializing the Kalman filter was a tricky business, but no such tinkering is needed with our direct matrix approach.

Models for Multivariate Trends

Define the $N \times 1$ vector process $\boldsymbol{\mu}_t = (\mu_t^1, ..., \mu_t^N)'$ as the trend, $\boldsymbol{\varepsilon}_t = (\varepsilon_t^1, ..., \varepsilon_t^N)'$ as the $N \times 1$ irregular, and $\mathbf{y}_t = (y_t^1, ..., y_t^N)'$ as the $N \times 1$ observed series. Then the multivariate local level model specifies

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

where $WN(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ denotes that the vector is white noise, i.e., serially uncorrelated with zero mean vector and $N \times N$ positive semi-definite covariance matrix $\boldsymbol{\Sigma}_\varepsilon$.

United States Census Bureau

Models for Multivariate Trends

Likewise, $\boldsymbol{\Sigma}_\eta$ is multivariate white noise, but even though in the general case it is $N \times N$, $\boldsymbol{\Sigma}_\eta$ might be of reduced rank $K < N$. When this occurs, we can rewrite (12) as

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\Theta}\boldsymbol{\mu}_t^\dagger + \boldsymbol{\mu}_0^\dagger + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\mu}_t^\dagger &= \boldsymbol{\mu}_{t-1}^\dagger + \boldsymbol{\eta}_t^\dagger. \end{aligned} \tag{13}$$

For identification, the elements of the load matrix $\boldsymbol{\Theta}$ are constrained to satisfy $\boldsymbol{\Theta}^{IJ} = 0$ for $J > I$, and $\boldsymbol{\Theta}^{II} = 1$ for $I = 1, ..., K$.

Models for Multivariate Trends

Hence, the trend depends on a smaller set of processes, arranged in the $K \times 1$ vector $\boldsymbol{\mu}_t^\dagger$, that tie together the series and are called common trends. This is also called a Dynamic Factor Model. These common trends, or factors, are driven by the disturbance $\boldsymbol{\eta}_t^\dagger$ whose $K \times K$ covariance matrix, $\boldsymbol{\Sigma}_{\eta^\dagger}$, is diagonal and positive. The $N \times 1$ vector $\boldsymbol{\mu}_0^\dagger$ contains zeros in the first K positions, and constants elsewhere.

Models for Multivariate Trends

The common slopes model is

$$
\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu}_t^{\dagger} + \boldsymbol{\varepsilon}_t \\
\boldsymbol{\mu}_t^{\dagger} &= \boldsymbol{\mu}_t^{\dagger} + \boldsymbol{\Theta}_{\beta}\boldsymbol{\beta}_{t-1}^{\dagger} + \boldsymbol{\beta}_0^{\dagger}, \\
\boldsymbol{\beta}_t^{\dagger} &= \boldsymbol{\beta}_{t-1}^{\dagger} + \boldsymbol{\zeta}_t^{\dagger}, \qquad \boldsymbol{\zeta}_t^{\dagger} \sim WN(0, \boldsymbol{\Sigma}_{\zeta^{\dagger}})
\end{aligned}
\tag{14}
$$

where $\boldsymbol{\beta}_t^{\dagger}$ is $K_{\beta} \times 1$ $(K_{\beta} < K), \boldsymbol{\Sigma}_{\zeta^{\dagger}}$ is diagonal, and $\boldsymbol{\Theta}_{\beta}$ has elements $(\boldsymbol{\Theta}_{\beta})^{IJ} = 0$ for $J > I$, and $(\boldsymbol{\Theta}_{\beta})^{II} = 1$ for $I = 1, ..., K_{\beta}$.

A Few Derivations for these Models

Consider the signal of interest is $\mathbf{s}_t = \boldsymbol{\mu}_t$ for all t, with common differencing operator $\delta_{\mathbf{s}}(L) = (1 - L)^m$. To derive the frf for the trend extraction problem, start with

$$f_{\mathbf{u}}^{IJ}(\lambda) = \Sigma_{\boldsymbol{\zeta}}^{IJ}$$

$$f_{\mathbf{v}}^{IJ}(\lambda) = \Sigma_{\epsilon}^{IJ}$$

$$f_{\mathbf{w}}^{IJ}(\lambda) = \Sigma_{\boldsymbol{\zeta}}^{IJ} + |1 - z|^{2m} \Sigma_{\epsilon}^{IJ}.$$

A Few Derivations for these Models

So the quantities in the matrix formulation of the bi-infinite filter are $\mathbf{G}_{\partial \mathbf{u}}(\lambda) = \Sigma_{\boldsymbol{\zeta}}$ and $\mathbf{G}_{\partial \mathbf{v}}(\lambda) = |1-z|^{2m} \Sigma_{\epsilon}$, so that the frf is

$$\Psi(z) = \Sigma_{\boldsymbol{\zeta}} \left(\Sigma_{\boldsymbol{\zeta}} + |1-z|^{2m} \Sigma_{\epsilon} \right)^{-1}. \tag{15}$$

A Few Derivations for these Models

Consider $\Psi(1)$, or the value of the gain function at $\lambda = 0$; this is of special interest, since it relates to how the very lowest frequency is passed by the filter. In the case that Σ_ζ is invertible (i.e., the similar trends case – where Θ can be taken as an identity matrix), we easily see that $\Psi(1) = 1_N$; in other words, related series have *no impact* on the frequency parts at the lowest region of the spectrum (the quintessential long-term), and the filter behaves as a collection of univariate filters.

A Few Derivations for these Models

But it is more interesting if Σ_ζ is non-invertible; supposing that $\Sigma_\zeta = \Theta \Sigma_\zeta^\dagger \Theta'$, we obtain

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

This formula reveals how the filter treats the lowest-frequency components. When $\Theta = \iota = [1, 1, \cdots, 1]'$ and Σ_ϵ is a multiple of the identity matrix, we get $\Psi(1) = \iota \iota' / N$, which equally weights the contribution of each input series. But for a more general Θ matrix, there may well be unequal weighting of each input series – for inflation we have $\Theta = (1, .847)'$.

Summary

- We give novel results on optimal signal extraction for multivariate time series, for bi-infinite and finite samples, including co-integration case.

- Matrix formulas are easy to implement (compared to SSF), fast to compute, and intuitive.

- Multivariate frf and Gain give insight into filtering.

- Co-movements in Core and Total PCE are nicely handled by the method. Method is quite flexible, allowing for different models.

Contact

tucker.s.mcelroy@census.gov