Frequency decompositions in autoregression models

Wouter Cames van Batenburg [*] Aleksander Czechowski[†]
Joey van der Leer Duran [‡] Bert Lindenhovius [§] Eric Siero [¶]

Abstract

Autoregression models are used by Ortec Finance to forecast the evolution of economic variables, such as interest rates. To distinguish the impact of short, medium and long term fluctuations, the company decomposes their models into three components: month, business and trend, respectively. We answer the question of how to design a model, so that predictions generated for a given frequency band do not overlap with other frequencies. We also discuss several other related matters, i.e. how to address the frequency leaking problem, how to choose the number of frequencies in each band and how our method generalizes to time-dependent models.

KEYWORDS: Fourier filter, autoregression, time series forecast

1 Introduction

This paper contains results on the problem of designing a good filtering method for autoregression models posed by Ortec Finance for the 114th European Study Group Mathematics with Industry. The general setting of the problem is as follows. Suppose we have a time series $\mathbf{r} = \{r_t\}_t$, where r is a quantity of interest (such as interest rate, oil price etc.), or a collection thereof, and $t \in \mathbb{Z}$ is a time parameter which takes discrete steps (representing months, years etc.). We want to make future predictions of r_t, given a historical set of values. A natural approach for forecasting based on data of such a time series is to describe it as a function of its predecessors

$$r_t = f(r_{t-1}, r_{t-2}, \ldots) + \epsilon_t, \tag{1}$$

where f is some function and ϵ_t is a sequence of independently, identically distributed random variables representing the probabilistic nature of future predictions. Vaguely

[*]Radboud University Nijmegen
[†]Jagiellonian University, Kraków
[‡]Utrecht University
[§]Radboud University Nijmegen
[¶]Leiden University

put, the aim is to choose f as simple as possible, while minimizing the deviation of the error terms ϵ_t. Often f is taken to be linear and dependent on only finitely many predecessors, in which case the model is called the AR(k) model:

$$r_t = c + \sum_{p=1}^{k} a_p r_{t-p} + \epsilon_t,$$

with $a_p \in \mathbb{R}$. We will assume that the ϵ_t are independent and identically distributed with mean 0 and the same standard deviation σ. Such a set $\{\epsilon_t\}_t$ is called *white noise*.

The AR(k) model is a special case of the *vector regression model*, where r_t and ϵ_t are both vector valued and the recursive structure is given by

$$r_t = c + Ar_{t-1} + \epsilon_t \tag{2}$$

where A is a matrix and c a vector. We will impose the restriction $||A|| < 1$, where $||A||$ denotes the operator norm of A, which will allow us to discard high powers of A. Intuitively, this condition amounts to stability of the model, but we will not make this statement precise.

By demeaning the data we can take $c = 0$. Let us also assume that the sequence starts at $t = 0$ with value r_0. Then, equation (2) has the following solution

$$r_t = \sum_{l=0}^{t-1} A^l \epsilon_{t-l} + A^t r_0.$$

Note that the expectation $\mathbb{E}(r_t)$ decays to 0 as time goes to infinity, because the ϵ_t have zero mean and $||A|| < 1$. Furthermore, as time becomes large the effect of the initial value diminishes.

Of course one cannot expect a single forecast based on such a rough model to be accurate. The value of the method is that it can be used to quickly generate a large number of scenarios and evaluate probabilities of future states via Monte Carlo experiments.

Ortec's approach to forecasting economic variables via autoregression models is to decompose the time series into a sum

$$\mathbf{r} = \mathbf{r}^T + \mathbf{r}^B + \mathbf{r}^M, \tag{3}$$

where $\mathbf{r}^T$ represents the long term (*trend*) fluctuations, $\mathbf{r}^B$ the medium term (*business cycle*) oscillations, and $\mathbf{r}^M$ the short term (*month*) movements. This is performed via so-called *filters*. We will elaborate more on them in Section 2, but to give some intuition, let us mention that a basic example is a filter is based on a discrete Fourier transform (*the Fourier filter*). For this filter the terms $\mathbf{r}^T$ correspond to low Fourier modes, the terms $\mathbf{r}^B$ to medium ones and $\mathbf{r}^M$ to the high frequencies.

The motivation for the decomposition (3) is that short, medium and long term terms are (to a certain degree) independent from each other, and, as such, their evolution should also be forecasted independently. However, a naive approach of applying an AR model separately for the trend, business cycle and month components can result in an undesirable effect, where a forecast for shorter fluctuations starts developing long term movements e.g. a forecast for the month term evolves a trend on its own.

For more background information about the intuition and practical applications of the frequency decomposition approach, we refer to Van der Schans and Steenhouwer (2012) and references therein.

In this paper we propose a *filtered AR model*, where long-term predictions are made for each term separately, in such a way that they stay in their own frequency band (which is chosen on the basis of historical data). Given a frequency decomposition, i.e. the choice of the filter, the recipe for such a prediction for a given term is as follows:

1. Firstly, we choose a forecasting period, which we specify by an integer $N \in \mathbb{Z}$, so that the outcome of the prediction will be a set $\{r_1, \ldots, r_N\}$, with initial condition r_0.

2. Secondly, we generate a time series of noise $\{\epsilon_t\}_{1 \leq t \leq N}$ of length equal to the forecasting period N, using the white noise probability distribution.

3. Thirdly, we apply the filter to the sequence $\{\epsilon_t\}_t$, in order to obtain a *filtered noise sequence* $\{\epsilon_t^*\}_t$.

4. Finally, we use the sequence $\{\epsilon_t^*\}_t$ to generate a prediction by the formula (2).

The filter F is typically chosen on the basis of the last N consecutive historical data points, so that both the employed historical data and the prediction are represented by a $N-$dimensional time series, implying that both are in the domain of F. As a consequence, it is possible to apply the *same* filter to both the employed historical data and the prediction, thus facilitating a meaningful comparison between the filtered historical data and the filtered prediction.

In order to apply the recipe, the forecasting period N can be arbitrary. In practice however, it is limited by the amount of historical data we have available.

In Section 2 we show, for the class of filters that are linear, weakly translation invariant and commuting with the autoregression parameter matrix, that such predictions will indeed remain in their own frequency band. Next, we give an example of two linear filters. The Fourier filter, presented in Subsection 2.1 is translation invariant, hence it can be employed in the filtered AR model. Another example is the Christiano-Fitzgerald Band Pass Filter, treated in Subsection 2.2. It is not clear whether this filter is weakly translation invariant. However, its advantage is that it deals with the *frequency leaking* problem, discussed later. In Section 3 we extend this method to regression models with time-dependent parameters.

Another problem we deal with is how to choose a partition into frequency bands. Ortec Finance chooses its decompositions based on heuristic reasoning backed by

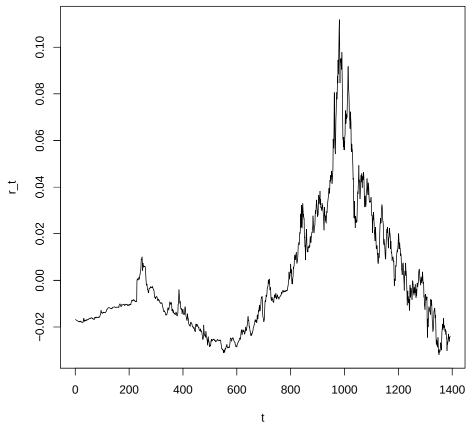

Figure 1: Demeaned interest rate of US bonds with a 10 year term over the past 116 years (in months; x-axis).

economic theories. We propose a different approach, where the allocation of the frequencies to each of the three terms is chosen to minimize the total variance of the given historical time series $\{r_t\}_t$ with respect to the filtered regression model. The rationale behind this is that the variance gives a measure of how well the regression model fits the given data, which is also the reason why least squares methods are often used to estimate the parameters of such models.

Due to analytical difficulties, we only performed a numerical study, and implemented our idea on historical data of the (univariate) interest rate of US bonds with a term of 10 years (see Figure 1). The details are presented in Section 4.

2 Filters

As discussed previously, a time series of interest can sometimes be regarded as a superposition of other time series with different kinds of evolution behaviors. To take that into account, we introduce a filtering process below, whose purpose is to decompose time series into its different constituents.

Let $l^\infty(\mathbb{R}) := \{(x_n)_{n\in\mathbb{Z}} | \sup|x_n| < \infty\}$ be the vector space of bounded sequences, and denote by $L : l^\infty(\mathbb{R}) \to l^\infty(\mathbb{R})$ the shift operator $(L(x))_n = x_{n-1}$. We are mainly interested in finite sets of data, which we regard as a subset of $l^\infty(\mathbb{R})$ by periodic extension. More precisely, we fix an $N \in \mathbb{Z}_{>0}$ and define $V \subset l^\infty(\mathbb{R})$ as the subspace of N-periodic sequences. Obviously, V is invariant under L.

Definition 2.1. A *linear filter* is a linear map $F : V \to V$. We call F *weakly translation invariant* if $L(im(F)) \subset im(F)$.

We think of F as a device that takes a time series and picks out a component with a specific evolution behavior. In particular, time series in the image of F are to be thought of as evolving in this specific way. The question at hand is how to produce a regression model that has as output a time series in the image of F. To this end, we modify the vector regression model (2) as follows. Let $F^1, \dots, F^d$ be linear filters and define $F := diag(F^1, \dots, F^d)$, considered as a linear map from $V^{\oplus d}$ to itself. Let ϵ_t^i be a set of random variables, with $1 \le i \le d$ and $t \in \{0, \dots, N-1\}$, put together into a sequence of vectors $\epsilon_t := (\epsilon_t^1, \dots, \epsilon_t^d)$. We take all ϵ_t^i independent of each other and, for each fixed i, we take the ϵ_t^i identically distributed with zero mean and variance σ_i. We can regard $(\epsilon_t^i)_{1 \le t \le N}$ as an element of V, and we will denote it by ϵ^i. Then, we define

$$\epsilon_t^* := (F\epsilon)_t = ((F^1\epsilon^1)_t, \dots, (F^d\epsilon^d)_t).$$

Simply put, we have d sequences of random variables and d filters, and we apply the filters component-wise. The *filtered vector regression model* is then defined by an initial value r_0, with time evolution given by

$$r_t = Ar_{t-1} + \epsilon_t^*, \tag{4}$$

where, as before, A is a matrix satisfying the stability condition $||A|| < 1$. Note that, in contrast to the non-filtered regression models, we need to specify the prediction period N in advance, in order for (4) to make sense. Indeed, we first need all the ϵ_t in order to apply the filter, after which the regression model can be initiated. The answer to the above question is given by the following proposition.

Proposition 2.1. *If all the F^i are linear and weakly translation invariant and if A commutes*[1] *with $F = \mathrm{diag}(F^1, \dots, F^d)$, then*

$$\Big(r_t - A^t r_0\Big)_{1 \le t \le N} \in Im(F).$$

So, except for the initial value terms $A^t r_0$ that converge to 0, the output of the filtered regression model is contained in the image of F.

Proof. By writing out the definitions and using that $[F, A] = 0$ we get

$$\begin{aligned} r_t &= \sum_{l=0}^{t-1} A^l \epsilon_{t-l}^* + A^t r_0 = \sum_{l=0}^{t-1} A^l (L^l F\epsilon)_t + A^t r_0 = \sum_{l=0}^{t-1} A^l (Fy^l)_t + A^t r_0 \\ &= \Big(F\big(\sum_{l=0}^{t-1} A^l (y^l)\big)\Big)_t + A^t r_0. \end{aligned}$$

[1] To be precise, A is a $d \times d$ matrix which acts naturally on $V^{\oplus d}$, while F acts on $V^{\oplus d}$ in a diagonal way by applying F^i to each component.

In the third equality we used that L preserves the image of F, so that we can find sequences y^l with the property that $L^l F\epsilon = Fy^l$. □

This proposition tells us that if we think of the filter as forcing the noise $(\epsilon_t)_t$ to have a certain time evolution, then the prediction for r will have this time evolution as well (at least in the long run, if we ignore the contribution from the initial value), provided that we use the same filters for those components of r that are interacting with each other (i.e. we need $[A, F] = 0$). We will discuss interactions between evolutions lying in different filters in Section 3.

2.1 The Fourier filter

An example of a linear filter F is the *Fourier filter* defined below. The *discrete Fourier transform* (DFT) of a sequence $x \in V$ is given by

$$\tilde{x}_k := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{\frac{-2\pi i k n}{N}}.$$

The sequence $\tilde{x}_k$ is obviously also N–periodic and the inverse DFT is given by

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \tilde{x}_k e^{\frac{2\pi i k n}{N}}.$$

Given a subset $K \subset \{0, \ldots, N-1\}$ with the property that $k \in K \Leftrightarrow N-k \in K$, we define the Fourier filter with respect to K as the map $x \mapsto Fx =: x^*$, where

$$x_n^* = \frac{1}{\sqrt{N}} \sum_{k \in K} \tilde{x}_k e^{\frac{2\pi i k n}{N}}.$$

Basically, the Fourier filter is given by first applying DFT, then applying a linear projection by forgetting some of the frequencies and then applying the inverse DFT. This example is prototypical for the concept of filter, designed with the purpose of making a separation between different time scales or frequencies, which are expected to have different driving mechanisms. Note that the Fourier filter F is translation invariant, satisfies $F^2 = F$ and $F^* = F$, with respect to the inner product on V given by

$$\langle x, y \rangle := \sum_{n=0}^{N-1} x_n y_n.$$

If $\{0, \ldots, N-1\} = K_1 \cup \ldots \cup K_q$ is a disjoint decomposition, the associated Fourier filters F_{K_i} additionally satisfy: $1 = \sum_i F_{K_i}$ and $F_{K_i} F_{K_j} = 0$ for all $i \neq j$.

2.2 The Christiano-Fitzgerald Band Pass Filter

In this section, we investigate the *Christiano-Fitzgerald filter*, for the following reasons. Firstly, the Fourier filter has the disadvantage of frequency leaking (see also below). For this reason, Ortec Finance is using another filter which is, however, non-linear. The Christiano-Fitzgerald filter might be the most prominent choice of a linear filter that prevents frequency leaking.

The discrete Fourier transform only filters a discrete set of frequencies. However, it may be possible (even plausible) that the frequencies of the input signal do not (perfectly) match the frequencies that are chosen to be filtered by the discrete Fourier transform. The discrete Fourier transform assumes that the input signal is periodic with a certain period, but it could happen that the input signal has a slightly different period. For instance, we want to filter the business cycle component of the interest rate and we assume a period of 8 years. However, the actual period of the rate turns out to be 7 years. It follows that if we use the discrete Fourier filter in order to filter certain frequencies out, we might damp certain eigenfrequencies of the input signal nearby the frequencies we actually want to keep, which is not desirable. This effect is called *frequency leaking*.

Therefore, it is desirable to filter an interval of frequencies. This leads to the *Ideal Band Pass Filter*. Unfortunately, this filter has the disadvantage that it requires the use of an infinite number of input values, whereas data sets are usually finite sets. Hence, an approximation is required, leading to the *Christiano-Fitzgerald Band Pass Filter*. This filter assumes that the historical data follows a random walk pattern (even though in most cases, this is a false assumption).

We start by defining the Ideal Band Pass Filter.

Definition 2.2. Let $(x_n)_{n\in\mathbb{Z}}$ be a time series. Choose $0 < a < b \leq \pi$ and let L be the shift operator sending x_n to x_{n-1}. Then, the Ideal Band Pass Filter is given by

$$B = \sum_{n\in\mathbb{Z}} B_n L^n$$

with

$$B_n = \begin{cases} \frac{b-a}{\pi}, & n = 0, \\ \frac{\sin(nb)-\sin(na)}{n\pi}, & n \neq 0. \end{cases}$$

The sum of all B_n is zero and $B_{-n} = B_n$. Moreover, we have

$$\sum_{n\in\mathbb{Z}} B_n e^{-in\omega} = \begin{cases} 1, & \omega \in (a,b) \cup (-b,-a), \\ 0, & \text{otherwise.} \end{cases}$$

Hence B is a filter that 'accepts' frequencies between a and b. Usually, the data set x_n is split into a 'trend' component t_n and a 'cyclic' component y_n, such that $x_n = y_n + t_n$, where $y_n = Bx_n$ for each $n \in \mathbb{Z}$. By definition, we have

$$y_n = \sum_{k\in\mathbb{Z}} B_k x_{n-k}.$$

Hence we require all x_k in order to calculate y_n. However, usually we only have a finite data set $(x_n)_{n=1}^N$, so the output y_k might not be accurate. We now define the Christiano-Fitzgerald Band Pass Filter (abbreviated by *CF Filter*) C as follows. Let z_n be the solution of minimizing the mean square error

$$E\big((y_n - z_n)^2 | x_1, \ldots, x_N\big).$$

Then, we define $Cx_n = z_n$. For $k = 1, \ldots, N-1$, define

$$\tilde{B}_{N-k} = -\frac{1}{2}B_0 - \sum_{j=1}^{N-k-1} B_j.$$

It is stated in Christiano and Fitzgerald (2003) that for $k \in \{2, \ldots, N-1\}$, we have

$$z_k = B_0 x_k + \sum_{j=1}^{N-k-1} B_j x_{k+j} + \sum_{j=1}^{k-2} B_j x_{k-j} + \tilde{B}_{k-1} x_1.$$

The values of z_1 and z_N are given by

$$z_1 = \frac{1}{2}B_0 x_1 + \sum_{j=1}^{N-2} B_j x_{j+1} + \tilde{B}_{T-1} x_N$$

and

$$z_N = \frac{1}{2}B_0 x_N + \sum_{j=1}^{N-2} B_j x_{N-j} + \tilde{B}_{T-1} x_1.$$

More generally, the CF filter is of the following form (cf. Schleicher (2003)):

$$z_k = \sum_{j=-n_{1,k}}^{n_{2,k}} C_{k,j} x_{k+j}$$

for some coefficients $C_{k,j}$. Clearly, this formula is only translation invariant if $C_{k,j} = C_j$ for each k, which is not the case for the CF filter. It is also not clear, whether the CF filter is weakly translation invariant. Hence, the CF filter might not be a useful filter if one wants to implement the methods that are developed in this contribution. Another filter which deals with frequency leakage is the Hodrick-Prescott filter. It is both linear and translation invariant, therefore it may be better suited to our purposes. For more information about the Hodrick-Prescott filter we refer to (Schleicher, 2003, §2.5.2).

3 Coupling models with different frequencies

In Proposition 2.1 we saw that the filtered regression model produces sequences lying in the image of the filter, provided that the filter is linear and weakly translation invariant in all components *and* commutes with the regression matrix A. Another way to describe the final condition is: different components that interact with each other need to be filtered in the same way. In practice however, there are situations where components with different time evolution still influence each other in some way. We now give one strategy to incorporate such interactions whilst preserving the conclusion of Proposition 2.1.

We can generalize the regression model (2) a bit if we allow A and the distribution of the ϵ_t's to depend on time as well. For instance, one can consider the AR(1)-model with constant $a = a(t)$ and standard deviations $\sigma = \sigma(t)$ that depend on time. In this way one can incorporate interactions between different models by letting them act via the parameters. Suppose that ϵ_t is a sequence of vector-valued random variables with zero mean and standard deviations σ_t and that A_t is a sequence of matrices with operator norms $||A_t|| < 1$ where $t \in \{0, \dots, N-1\}$. If $F = (F^1, \dots, F^d)$ is a filter, we can define, as before, the filtered regression by starting with an initial value r_0 and applying the recursive formula

$$r_t = A_t r_{t-1} + \epsilon_t^* \tag{5}$$

with $\epsilon^* = F\epsilon$. As before, we have

Proposition 3.1. *If F is linear and weakly translation invariant and commutes with A_t for all t, then the solution of the filtered regression (5) lies in the image of F, modulo an initial value term that converges to zero.*

Proof. In this case, the solution of (5) is given by

$$r_t = \sum_{l=0}^{t-1} A_t A_{t-1} \cdots A_{t-l+1} \epsilon_{t-l}^* + A_t \cdots A_1 r_0.$$

From here on the proof is identical to the proof of Proposition 2.1. □

4 A minimal variance approach to band decomposition

In this section we numerically investigate the (optimal) decomposition into frequency bands. For this, we focus on the (univariate) time series of the monthly interest rate of US bonds with a term of 10 years. We restrict our attention to the AR(1) model and a Fourier filter (Section 2.1).

As explained in the introduction, the current decomposition used by Ortec Finance consists of three bands – trend, business cycle and month. For the Fourier filter, the table below explicilty shows which of the frequencies are part of which band.

	period	frequencies
Month	2 months-2 years	$K_M = [N/24, N - N/24]$
Business	2 years-16 years	$K_B = [N/192, N/24) \cup (N - N/24, N - N/192]$
Trend	longer than 16 years	$K_T = [1, N/192) \cup (N - N/192, N - 1]$

Given the equidistant frequency distribution of the Fourier filter, about 92% of the frequency components is in the month component, 7% is in the business cycle component and only 1% is in the trend component. We note that $0 \in K$ corresponds to the mean interest rate which we have (without loss of generality) disregarded. We also note that Ortec Finance currently uses a nonlinear filter which may lead to a different distribution of frequencies.

How does the decomposition compare to other possible partitions in the simple univariate setting? We will compare the different decompositions based on the linear Fourier filter by comparing the total variances from the AR(1) model. As mentioned before, the total variance is a certain measure of fit of the model to the time series, so our idea for the choice of decomposition is to select the one that has the minimal total variance.

Since the interest rate time series is real and we want the filtered time series to be real as well, we impose that $j \in K$ implies $N - j \in K$. Furthermore, to make the computation feasible, we make the reasonable assumption that each of the three parts of the partition is 'connected' in the sense that there exist integers $2 \leq a \leq b \leq \lfloor \frac{N}{2} \rfloor$ such that $K_T = \{1, \ldots, a-1\} \cup \{N-a+1, \ldots, N-1\}, K_B = \{a, \ldots, b-1\} \cup \{N-b+1, \ldots, N-a\}$ and $K_M = \{b, \ldots, N-b\}$.

Write $\mathbf{r} = (F_{K_M}(\mathbf{r}), F_{K_B}(\mathbf{r}), F_{K_T}(\mathbf{r})) = (\mathbf{r}^M, \mathbf{r}^B, \mathbf{r}^T)$ for the decomposition of the interest rate in a month, business and trend component. We initialize the AR(1)-model for each frequency band by using the ordinary least squares method.

To have the best fit with the historical data, we should find (a^M, a^B, a^T) such that the total variance

$$\mathrm{Var_{tot}} := \frac{1}{N-1} \sum_{t=1}^{N-1} \left|\left|r_t^M + r_t^B + r_t^T - a^M r_{t-1}^M - a^B r_{t-1}^B - a^T r_{t-1}^T\right|\right|^2 \tag{6}$$

is minimal. Instead, in the filtered AR(1) framework based on the least squares method, the parameters a^M, a^B, a^T are chosen separately to minimize the separate variances:

$$\begin{aligned} \mathrm{Var_M} &:= \frac{1}{N-1} \sum_{t=1}^{N-1} \left|\left|r_t^T - a^T r_{t-1}^T\right|\right|^2, \\ \mathrm{Var_B} &:= \frac{1}{N-1} \sum_{t=1}^{N-1} \left|\left|r_t^B - a^B r_{t-1}^B\right|\right|^2, \\ \mathrm{Var_T} &:= \frac{1}{N-1} \sum_{t=1}^{N-1} \left|\left|r_t^M - a^M r_{t-1}^M\right|\right|^2. \end{aligned} \tag{7}$$

The separate minimizers a^M, a^B and a^T of (7) together yield an (almost) minimal value of (6). To see this, we show that $\mathrm{Var_M} + \mathrm{Var_B} + \mathrm{Var_T} \approx \mathrm{Var_{tot}}$. We make use of the inner product $\langle \cdot, \cdot \rangle$ on V and orthogonality of the Fourier basis. It holds that

$$
\begin{aligned}
(N-1)\mathrm{Var}_{\mathrm{tot}} = \sum_{t=0}^{N-2} \left|\left| r_t^M + r_t^B + r_t^T - a^M r_{t-1} - a^B r_{t-1} - a^T r_{t-1} \right|\right|^2 \\
=& \langle \mathbf{r}^M + \mathbf{r}^B + \mathbf{r}^T - a^M L\mathbf{r} - a^B L\mathbf{r} - a^T L\mathbf{r}, \mathbf{r}^M + \mathbf{r}^B + \mathbf{r}^T - a^M L\mathbf{r} - a^B L\mathbf{r} - a^T L\mathbf{r} \rangle \\
& - \left|\left| r_{N-1}^M + r_{N-1}^B + r_{N-1}^T - a^M r_0^M - a^B r_0^B - a^T r_0^T \right|\right|^2 \\
=& \langle F_{K_M}\mathbf{r} - a^M F_{K_M} L\mathbf{r}, F_{K_M}\mathbf{r} - a^M F_{K_M} L\mathbf{r} \rangle + \langle F_{K_B}\mathbf{r} - a^B F_{K_B} L\mathbf{r}, F_{K_B}\mathbf{r} - a^B F_{K_B} L\mathbf{r} \rangle \\
& + \langle F_{K_T}\mathbf{r} - a^T F_{K_T} L\mathbf{r}, F_{K_T}\mathbf{r} - a^T F_{K_T} L\mathbf{r} \rangle \\
& - \left|\left| r_{N-1}^M + r_{N-1}^B + r_{N-1}^T - a^M r_0^M - a^B r_0^B - a^T r_0^T \right|\right|^2 \\
=& (N-1)\mathrm{Var}_M + (N-1)\mathrm{Var}_B + (N-1)\mathrm{Var}_T \\
& - \left|\left| r_{N-1}^M + r_{N-1}^B + r_{N-1}^T - a^M r_0^M - a^B r_0^B - a^T r_0^T \right|\right|^2 \\
& + \left|\left| r_{N-1}^M - a^M r_0^M \right|\right|^2 + \left|\left| r_{N-1}^B - a^B r_0^B \right|\right|^2 + \left|\left| r_{N-1}^T - a^T r_0^T \right|\right|^2 .
\end{aligned}
$$

After dividing the equality by $N-1$, we see that the error that is made by minimizing the separate variances (7) instead of (6), represented by the terms on the last two lines of the equation, is small if the amount of data N is large.

Since it is computationally much more efficient to optimize three times over a one dimensional set than once over a three dimensional data set, our script minimizes the separate variances (7).

We performed numerical tests on the monthly time series of interest rates from the past 116 years, consisting of 1392 data points (Figure 1). We computed the resulting total variance of all possible frequency decompositions (Figure 2). We found that (in this case) $\mathrm{Var}_{\mathrm{tot}}$ is minimal for a decomposition given by 624 frequencies in the month component, 410 in the business cycle component and 358 in the trend component. This results in the decomposition of the interest rate as shown in Figure 3. The corresponding frequency and period decomposition is shown in the table below.

	frequencies	period
Month	$K_M = [384, 1008]$	2 months-3.6 months
Business	$K_B = [179, 384) \cup (1008, 1087]$	3.6 months-7.8 months
Trend	$K_T = [1, 179) \cup (1087, 1391]$	longer than 7.8 months

so that about 45% of the frequencies are in the month component, 29% in the business component and 26% in the trend component.

5 Concluding remarks

We have proposed two ways to possibly improve the filtered regression models used by Ortec Finance. The first one is a method for generating predictions, which ensures that predictions via regression stay in the same frequency band as the one corresponding to the filtered historical time series. To put it shortly, a sequence of samples from

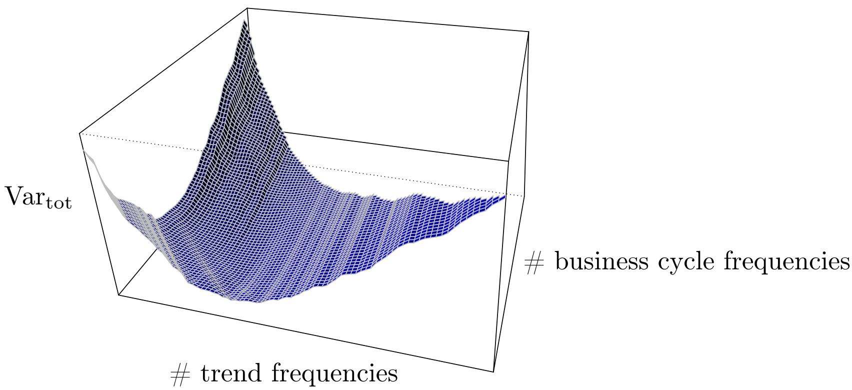

Figure 2: The total variance for all frequency decompositions attains its minimum $(N-1) \times \mathrm{Var}_{\mathrm{tot}} \approx 0.004640976$ for 358 frequencies in the trend component and 410 frequencies in the business cycle component. The total number of frequencies is $N = 1392$, so the remaining ones (624) belong to the month component.

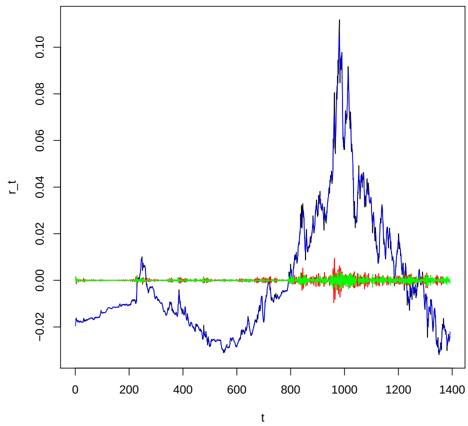

Figure 3: The optimal (minimizing the total variance) decomposition of the interest rate. The month component in green, the business cycle component in red and the trend component in blue.

white noise needs to be generated *a priori* for the whole prediction period, and then filtered, as opposed to sampling the white noise at each time step of the prediction. We identify a group of filters for which the method is applicable – the class of linear, weakly translation invariant filters that commute with the parameter matrix. In particular, the method can be readily applied for scalar, Fourier filtered AR(1) models, and can incorporate time-dependent parameters. However, currently Ortec Finance is using nonlinear filters to address the frequency leaking problem, and further investigation has to be performed to find a weakly translation invariant, linear filter that prevents frequency leaking. In particular, even though the Christiano-Fitzgerald band pass filter is linear and prevents frequency leaking, it is not applicable as it does not possess good translation invariance properties.

Our second contribution is the idea that by optimizing the number of frequencies in each band, one can further reduce the total variance of the model with respect to the given time series. This way, the frequency decomposition can be adapted to the time series, rather than arbitrarily fixed beforehand. Our numerical calculations based on the data set of demeaned interest rates of US bonds and a Fourier filtered AR(1) model indeed shows that there seems to be a clear global minimum for the total variance; see Figure 2. The method is, in principle, independent of the filter and applicable to any filtered autoregression model. We note that in the application to this particular dataset, only 45% of the frequencies entered the month component, as opposed to 92% in the decomposition used by Ortec Finance and consequently the size of the business cycle component, and particularly of the trend component was much bigger. Perhaps increasing the number of frequency bands (e.g. to four or five) would make a clear narrow trend similar to the one from Ortec's decomposition reveal itself, and what has been captured as trend in the three frequency band setting is in fact a new, intermediate pattern.

References

L. Christiano and T. Fitzgerald. The band pass filter. *International Economic Review*, 44:435–465, 2003.

C. Schleicher. Essays on the decomposition of economic variables. *PhD Thesis, University of British Columbia*, 2003.

M. Van der Schans and H. Steenhouwer. Imposing views on frequency domain factor models, methodological working paper no. 2012-01. *Ortec Finance Research Center*, 2012.