# Марковский подход к анализу тренда и сезонности в модели нестационарной регрессии

Д. С. Кононенко

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель д.т.н., профессор, в.н.с. ВЦ РАН В. В. Моттль

Москва,
2011 г.

## План доклада

- Линейная нормальная модель сигнала.
- Процедуры фильтрации и интерполяции.
- Тренд и сезонность.
- Задача анализа ВВП.
- Вычислительный эксперимент.

# Общая постановка задачи

- Дан $Y = (y_t,\ t = 1 \dots N)$ — наблюдаемый сигнал, $y_t \in \mathbb{Y} \subseteq \mathbb{R}$.
- Скрытый процесс $X = (\mathbf{x}_t,\ t = 1 \dots N)$ — модель наблюдаемого сигнала, $\mathbf{x}_t \in \mathbb{X} \subseteq \mathbb{R}^n$.
- Нужно для сигнала $Y$ подобрать наиболее подходящую модель $X$.

# Байесовский подход

## Марковская модель

$$\psi_t(x_t|x_1,\dots,x_{t-1}) = \psi_t(x_t|x_{t-1}).$$

$$\varphi_t(y_t|y_1,\dots,y_{t-1},X) = \varphi_t(y_t|y_1,\dots,y_{t-1},x_t).$$

## Оптимизационная задача

$$\eta_0(x_0) = \log\psi_0(x_0),\ \ \eta_t(x_t|Y^t) = \log\varphi_t(y_t|Y^{t-1};x_t),$$
$$\gamma_t(x_{t-1},x_t) = \log\psi_t(x_t|x_{t-1}),\ \ t = 1,\dots,N.$$

$$\hat{X}(Y) = arg\max_{X\in\mathcal{X}}\sum_{t=0}^{N}\eta_t(x_t|Y^t) + \sum_{t=1}^{N}\gamma_t(x_{t-1},x_t).$$

# Линейная нормальная модель сигнала со скрытой компонентой

Предполагаем следующую модель порождения данных:

$$\begin{cases} \mathbf{x}_t = \mathbf{V}_t \mathbf{x}_{t-1} + \varepsilon_t; \\ y_t = \mathbf{c}_t \mathbf{x}_t + \xi_t. \end{cases}$$

$\mathbf{V}_t (n \times n)$, $M(\varepsilon_t) = \mathbf{0}$, $M(\varepsilon_t \varepsilon_s^T) = \mathbf{0}$ при $t \neq s$,
$M(\varepsilon_t \varepsilon_t^T) = \sigma^2 \mathbf{U}_t^{-1}$.
$\mathbf{c}_t \in \mathbb{R}^n$, $M(\xi_t) = 0$, $M(\xi_t \xi_s^T) = 0$ при $t \neq s$, $M(\xi_t \xi_t^T) = \sigma^2$,
$M(\varepsilon_t \xi_s) = \mathbf{0}$.

# Предположения о нормальности

## Предположения о шуме

$\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{U}_t^{-1})$, $\xi_t \sim \mathcal{N}(0, \sigma^2)$.

## Начальное значение скрытого процесса — априорная модель

$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_{0|0}, \mathbf{B}_{0|0}^{-1})$.

Отсутствие априорных знаний о скрытом процессе может быть выражено следующим образом:

$$\mathbf{B}_{0|0}^{-1} = \begin{pmatrix} \rho & 0 & \dots & 0 \\ 0 & \rho & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho \end{pmatrix} \Rightarrow \mathbf{B}_{0|0} = \mathbf{0},$$

$\rho \to \infty$.

# Процедура фильтрации

## Фильтрационные плотности распределения

$Y^t = (y_1, \dots, y_t).$

$p_{t|t}(\mathbf{x}_t) = p(\mathbf{x}_t | Y^t) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|t}, \mathbf{B}_{t|t}^{-1}).$

Процедура фильтрации заключается в последовательном нахождении фильтрационных плотностей распределения. Начальные значения берутся из априорной модели.

## Шаг процедуры фильтрации

$$\begin{cases} \hat{\mathbf{x}}_{t|t} = \mathbf{V}_t \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_{t|t}^{-1} \mathbf{c}_t \left( y_t - \mathbf{c}_t^T \mathbf{V}_t \hat{\mathbf{x}}_{t-1|t-1} \right), \\ \mathbf{B}_{t|t} = \mathbf{c}_t \mathbf{c}_t^T + \left( \mathbf{V}_t \mathbf{B}_{t-1|t-1}^{-1} \mathbf{V}_t^T + \mathbf{U}_t^{-1} \right)^{-1}. \end{cases}$$

# Процедура интерполяции

## Интерполяционные плотности распределения

$$p_{t|N}(\mathbf{x}_t) = p(\mathbf{x}_t|Y) \sim \mathcal{N}(\hat{\mathbf{x}}_{t|N}, \mathbf{B}_{t|N}^{-1}).$$

Процедура интерполяции заключается в последовательном нахождении интерполяционных плотностей распределения. Апостериорное матожидание $\hat{\mathbf{x}}_{N|N}$ и ковариационная матрица $\mathbf{B}_{N|N}^{-1}$ вычислены на последнем шаге процедуры фильтрации.

## Шаг процедуры интерполяции

$$\begin{cases} \hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_{t|t} + \mathbf{H}_t \left(\hat{\mathbf{x}}_{t+1|N} - \mathbf{V}_{t+1}\hat{\mathbf{x}}_{t|t}\right), \\ \mathbf{B}_{t|N}^{-1} = \mathbf{H}_t\mathbf{B}_{t+1|N}^{-1}\mathbf{H}_t^T + \left(\mathbf{V}_{t+1}^T\mathbf{U}_{t+1}\mathbf{V}_{t+1} + \mathbf{B}_{t|t}\right)^{-1}, \\ \mathbf{H}_t = \mathbf{B}_{t|t}^{-1}\mathbf{V}_{t+1}^T \left(\mathbf{V}_{t+1}\mathbf{B}_{t|t}^{-1}\mathbf{V}_{t+1}^T + \mathbf{U}_{t+1}^{-1}\right)^{-1}. \end{cases}$$

# Модель тренда

Пусть тренд выражается полиномом степени не выше $k$: $g(t) = a_k t^k + \cdots + a_1 t + a_0$. Тогда, зная значения $g(t_0), \ldots,$ $g(t_k)$, можно восстановить все коэффициенты полинома и значение $g(t)$ в любой момент времени. Это может быть выражено следующим образом:

$$\begin{pmatrix} g_\tau \\ \vdots \\ g_{\tau+k} \end{pmatrix} = V^g \begin{pmatrix} g_{\tau-k-1} \\ \vdots \\ g_{\tau-1} \end{pmatrix}.$$

Здесь матрица $V^g$ размера $(k+1) \times (k+1)$ зависит только от степени полинома $k$.

# Модель сезонности

Пусть сезонность выражается $l$ гармониками:

$h(t) = a_0 + \sum_{i=1}^{l} a_i \cos(w_i t) + b_i \sin(w_i t),\ w_i = 2\pi i$. Зная $2l+1$ значений $h(t_0), \dots, h(t_{2l})$, можно восстановить коэффициенты и значение $h(t)$ в любой момент времени. Это может быть выражено следующим образом:

$$\begin{pmatrix} h_\tau \\ \vdots \\ h_{\tau+2l} \end{pmatrix} = V^h \begin{pmatrix} h_{\tau-2l-1} \\ \vdots \\ h_{\tau-1} \end{pmatrix}.$$

Здесь матрица $V^h$ размера $(2l+1) \times (2l+1)$ зависит только от $l$.

# Скрытые переменные тренда и сезонности

$$\mathbf{z}_t = (\mathbf{x}_t^T, \mathbf{g}_t^T, \mathbf{h}_t^T)^T$$

## Линейная нормальная модель с трендом и сезонностью

$$\begin{cases} \mathbf{z}_t = \mathbf{V}_t \mathbf{z}_{t-1} + \varepsilon_t, \quad t = 2, \dots, N; \\ y_t = \mathbf{c}_t^T \mathbf{z}_t + \xi_t = (\mathbf{c}_t^x)^T \mathbf{x}_t + g_t^1 + \dots + g_t^{k+1} + h_t^1 + \dots + h_t^{2l+1} + \xi_t. \end{cases}$$

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{V}_t^x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & V^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & V^h \end{pmatrix}, \quad \mathbf{U}_t^{-1} = \begin{pmatrix} (\mathbf{U}_t^x)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

# Задача анализа ВВП

$y_t$ — ВВП за квартал, $t = 1 \dots N$

Скрытый процесс $\mathbf{z}_t = (\mathbf{x}_t^T, \mathbf{g}_t^T, \mathbf{h}_t^T)^T$, где

$$\mathbf{x}_t = \begin{pmatrix} x_{t,1} \\ x_{t,2} \\ x_{t,3} \end{pmatrix}$$ — сигнал без учета тренда и сезонности по месяцам,

$$\mathbf{g}_t = \begin{pmatrix} g_{t,1} \\ g_{t,2} \\ g_{t,3} \end{pmatrix}$$ — тренд за квартал по месяцам,

$$\mathbf{h}_t = \begin{pmatrix} h_{t,1} \\ h_{t,2} \\ h_{t,3} \end{pmatrix}$$ — сезонность за квартал по месяцам.

$y_t = \mathbf{c}^T \mathbf{z}_t + \xi_t =$

$x_{t,1} + x_{t,2} + x_{t,3} + g_{t,1} + g_{t,2} + g_{t,3} + h_{t,1} + h_{t,2} + h_{t,3} + \xi_t$,

t=1,...,N.

## Компонента с шумом

$x_\tau = x_{\tau-1} + \eta_\tau,\ \tau = 2, \dots, 3N,$
$M(\eta_\tau) = 0,\ D(\eta_\tau) = \sigma_x^2,\ M(\eta_{\tau_1}\eta_{\tau_2}) = 0$ при $\tau_1 \neq \tau_2$.

$\mathbf{x}_t = \begin{pmatrix} x_\tau \\ x_{\tau+1} \\ x_{\tau+2} \end{pmatrix}$ — компонента скрытого процесса.

$$V^x = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

$$(U^x)^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$$

$\tau$ — номер месяца.

# Компоненты тренда и сезонности

$$g_\tau = 3g_{\tau-1} - 3g_{\tau-2} + g_{\tau-3}, \ \tau = 4, \dots, 3N.$$

$$\mathbf{g}_t = \begin{pmatrix} g_\tau \\ g_{\tau+1} \\ g_{\tau+2} \end{pmatrix}$$ — компонента скрытого процесса.

$$V^g = \begin{pmatrix} 1 & -3 & 3 \\ 3 & -8 & 6 \\ 6 & -15 & 10 \end{pmatrix}.$$

$$h_\tau = 2\left[\cos\left(\frac{2\pi}{T}\right)\right] h_{\tau-1} - h_{\tau-2}, \ \tau = 3, \dots, 3N.$$

$$\mathbf{h}_t = \begin{pmatrix} h_\tau \\ h_{\tau+1} \\ h_{\tau+2} \end{pmatrix}$$ — компонента скрытого процесса.

$$V^h = \begin{pmatrix} 0 & -1 & 2\cos(w) \\ 0 & -2\cos(w) & 4\cos^2(w) - 1 \\ 0 & -(4\cos^2(w) - 1) & 8\cos^3(w) - 4\cos(w) \end{pmatrix}.$$

## Линейная нормальная модель

Модель сводится к виду, рассмотренному выше:

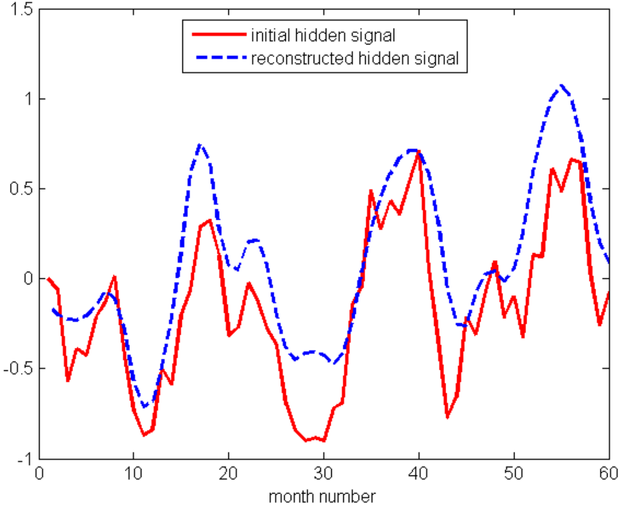
$$\begin{cases} \mathbf{z}_t = V\mathbf{z}_{t-1} + \varepsilon_t, & t = 2, \dots, N; \\ y_t = \mathbf{c}_t\mathbf{z}_t + \xi_t, & t = 1, \dots, N; \end{cases}$$

где скрытый процесс $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{g}_t, \mathbf{h}_t)^T$,

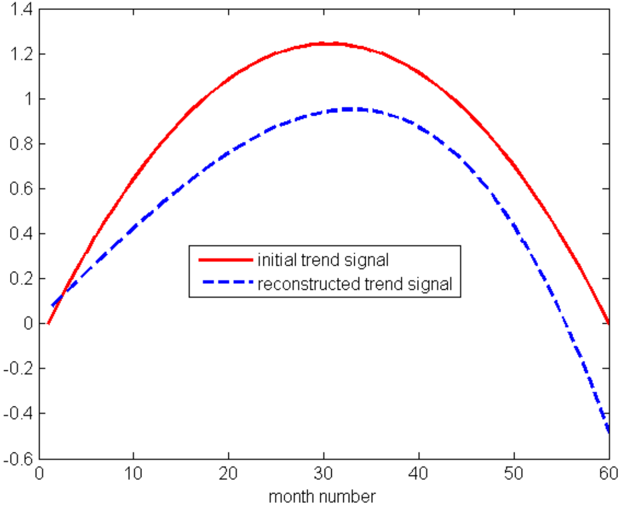
$\mathbf{c}_t = (1 \dots 1)^T \quad [9 \times 1]$,

$$V = \begin{pmatrix} V^x & 0 & 0 \\ 0 & V^g & 0 \\ 0 & 0 & V^h \end{pmatrix} \quad [9 \times 9].$$

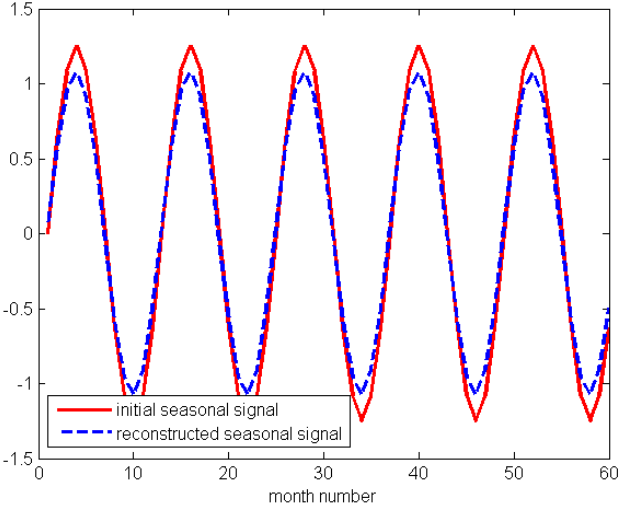
# Зашумленная скрытая компонента



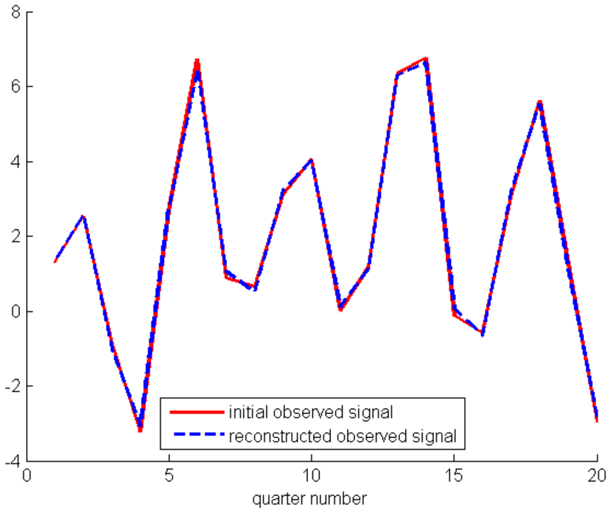
# Скрытая компонента тренда



# Скрытая компонента сезонности



# Наблюдаемый сигнал



# Результаты работы

- Предложен способ моделирования тренда и сезонности в рамках линейной нормальной модели нестационарной регрессии.
- Предложенный метод реализован в среде MATLAB.
- Предложенные модели проиллюстрированы на задаче анализа ВВП.
- Проведен вычислительный эксперимент, который показал работоспособность предложенного метода, но также и выявил некоторые его недостатки.