

Marinozzi, Tomás; Nallar, Leandro; Pernice, Sergio A.

**Working Paper**

## Intuitive mathematical economics series: Linear algebra techniques to measure business cycles

Serie Documentos de Trabajo, No. 802

**Provided in Cooperation with:**

University of CEMA, Buenos Aires

*Suggested Citation:* Marinozzi, Tomás; Nallar, Leandro; Pernice, Sergio A. (2021) : Intuitive mathematical economics series: Linear algebra techniques to measure business cycles, Serie Documentos de Trabajo, No. 802, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA), Buenos Aires

This Version is available at:

<https://hdl.handle.net/10419/248616>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**UNIVERSIDAD DEL CEMA**  
**Buenos Aires**  
**Argentina**

Serie  
**DOCUMENTOS DE TRABAJO**

**Área: Matemática y Economía**

**Intuitive Mathematical Economics Series.**  
**Linear Algebra Techniques to Measure Business Cycles**

**Tomás Marinozzi, Leandro Nallar, Sergio A. Pernice**

**Julio 2021**  
**Nro. 802**

**[www.cema.edu.ar/publicaciones/doc\\_trabajo.html](http://www.cema.edu.ar/publicaciones/doc_trabajo.html)**  
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina  
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)  
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding [jae@cema.edu.ar](mailto:jae@cema.edu.ar)



# Intuitive Mathematical Economics Series

## Linear Algebra Techniques to Measure Business Cycles

Tomás Marinozzi<sup>†</sup>

Leandro Nallar<sup>\*</sup>

Sergio Pernice<sup>§</sup>

*University of CEMA*

August 2021

### Abstract

Linear algebra is without a doubt a fundamental tool to deal with empirical economic problems. The goal of this paper is to use some of these techniques to treat business cycles. To do that, we present the classic ordinary least square approach to estimate the coefficients of a detrended time series in addition to the matrix form of the Hodrick–Prescott (HP) filter. This is a paper is part of “*Intuitive Mathematical Economic Series*”.

KEY WORDS: Linear Algebra, business cycles, trend.

---

<sup>†</sup>Email: tomasmarinozzi1996@gmail.com

<sup>\*</sup>Email: lean.nallar@hotmail.com

<sup>§</sup>Email: sp@ucema.edu.ar

The authors’ point of view do not reflect that of the University of CEMA.

# 1 Introduction

In macroeconomic, business cycles are generally described as fluctuations found in country's overall economic activity. Cycles need to be defined in terms of a secular component<sup>1</sup>. For instance, take the following 1960-2019 yearly GDP data for of Argentina and USA<sup>2</sup>.

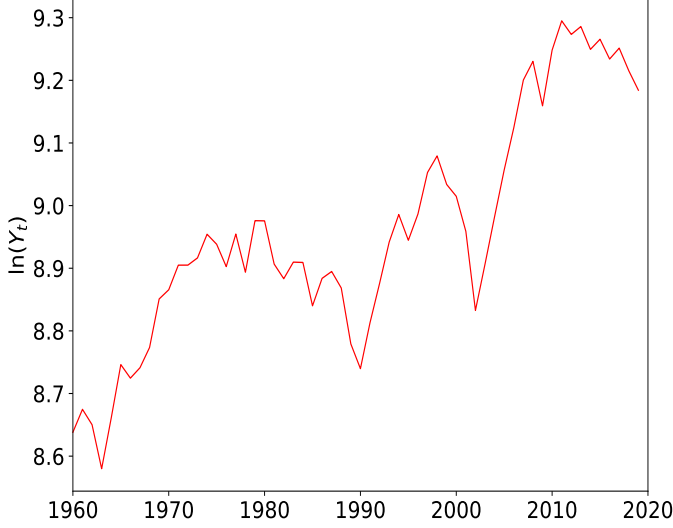


Figure 1: Argentina. GDP Data

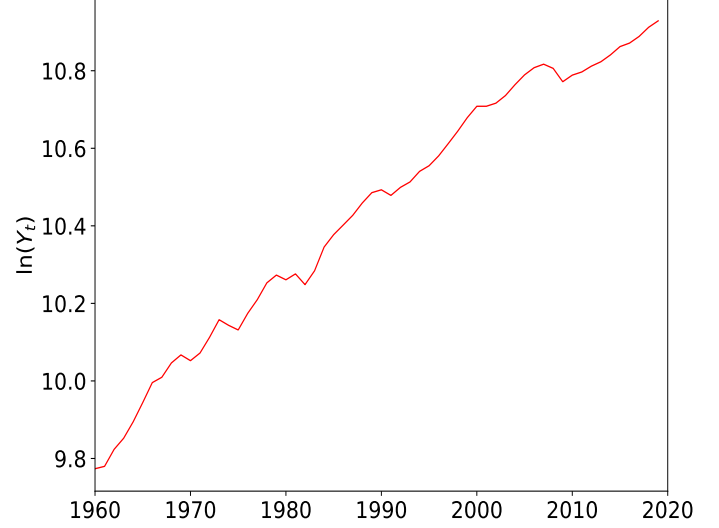


Figure 2: United States. GDP Data

For a regular time series pattern, the value of the series “ $y_t$ ” is usually a function of previous values and time explicitly. The the goal is to build a model of the type:

$$y_t = f(t, y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-n}) + e_t \quad (1)$$

but also with the ability to decompose the secular and cyclical components as

$$y_t \equiv y_t^c + y_t^s \quad (2)$$

There is an important trade-off between respecting the data generating process and providing meaningful information regarding business cycles. There are several methods to do this, in this paper we use *Ordinary Least Square*, the objective is to make the underlying liner algebra as transparent and intuitive as possible.

## 2 Ordinary Least Square (OLS)

Before dialing straight to our economic problem, in the spirit of this series, we briefly review some linear algebra techniques from the perspective presented in [Pernice \(2019\)](#). Ordinary least squares is one of

---

<sup>1</sup>See [Burns \(1946\)](#).

<sup>2</sup>Source: World Bank Data. We could have used a different serie such as unemployment, consumption, investment, or prices. For non stationary series is common practice to apply logarithms before dealing with cycle analysis. We will follow that approach and work with  $y_t \equiv \ln(Y_t)$ .

many linear algebra techniques to observe some relationship between a predicted variable  $y \in \mathbb{R}$  and some covariates  $\mathbf{x} = \langle x_0, \dots, x_p \rangle^\top \in \mathbb{R}^{p+1}$ . For instance, a linear regression could be described as

$$y = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3)$$

where  $\beta = \langle \beta_0, \dots, \beta_p \rangle$  are the parameters of the linear regression.

In practice, we typically have a data set containing  $n$  observations of  $(\mathbf{x}_i; y_i)$  that are supposed to come from the same data generating process. We structure the data as<sup>3</sup>

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n \quad X = \begin{bmatrix} -\mathbf{x}_1^\top - \\ \vdots \\ -\mathbf{x}_n^\top - \end{bmatrix} = \begin{bmatrix} x_{10} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n0} & \cdots & x_{np} \end{bmatrix} \in \mathbb{R}^{n \times (p+1)}$$

In economic applications we normally have significantly more data points than the amount of variables we are using, meaning  $n \gg p$ , so in general  $\mathbf{y}$  is not in the span of the columns of  $X$ . Therefore, we need to specify the estimation error as<sup>4</sup>

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times k} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix}_{k \times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

We need to decide the criteria for how to choose the most adequate  $\beta$ 's to fit the data<sup>5</sup>. The most conventional approach is to find the vector  $\beta \in \mathbb{R}^k$  such that the linear combination  $X\beta \in \mathbb{R}^n$  of columns of  $X$ , minimizes the sum of squared residuals (RSS). Formally,

$$\begin{aligned} \beta &= \operatorname{argmin}_{\beta \in \mathbb{R}^k} \|\mathbf{y} - X\beta\|_2^2 \\ &= \operatorname{argmin}_{\beta \in \mathbb{R}^k} \|\varepsilon\|_2^2 \end{aligned} \quad (4)$$

From the first expression it is clear that the minimum corresponds to the linear combination of the columns of  $X$  closest to  $\mathbf{y}$ . By Pythagoras,  $\mathbf{y} - X\beta$  should be orthogonal to  $X\beta$ . Using vector space notation, the

---

<sup>3</sup>We use small Roman or Greek letters such as “ $a$ ” or “ $\alpha$ ” for scalars, boldsymbol small Roman or Greek letters such as “ $\mathbf{y}$ ” or “ $\beta$ ” for vectors, and big letters such as “ $X$ ” for matrices.

<sup>4</sup>where  $k \equiv p + 1$  and  $x_{n0} \equiv 1 \forall n$

<sup>5</sup>Note that we normalize the first column, this is a conventional way of controlling by fixed base effects that are unrelated to the set of variables  $\{x_1, \dots, x_p\}$ . Therefore, there are  $p$  variables and  $k \equiv p + 1$  parameters to estimate.

square of the Euclidean norm<sup>6</sup>  $\|\boldsymbol{\varepsilon}\|_2^2$ . may be written as  $\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}$  or  $\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$

$$\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix}_{1 \times n} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} = [\varepsilon_1 \times \varepsilon_1 + \varepsilon_2 \times \varepsilon_2 + \dots + \varepsilon_n \times \varepsilon_n]_{1 \times 1}$$

We can write this expression as:

$$\begin{aligned} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} &= (\mathbf{y} - X\boldsymbol{\beta})^\top (\mathbf{y} - X\boldsymbol{\beta}) \\ &= \mathbf{y}^\top \mathbf{y} - \boldsymbol{\beta}^\top X^\top \mathbf{y} - \mathbf{y}^\top X\boldsymbol{\beta} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \\ &= \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top X^\top \mathbf{y} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \end{aligned}$$

where we use the fact that the transpose of a scalar is the same scalar, i.e.  $\mathbf{y}^\top X\boldsymbol{\beta} = (\mathbf{y}^\top X\boldsymbol{\beta})^\top = \boldsymbol{\beta}^\top X^\top \mathbf{y}$ . To find the  $\hat{\boldsymbol{\beta}}$  that minimizes the sum of squared residuals, the gradient with respect to  $\boldsymbol{\beta}$  should vanish:

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} (\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}) &= \nabla_{\boldsymbol{\beta}} [\mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top X^\top \mathbf{y} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta}] = 0 \\ -2X^\top \mathbf{y} + 2X^\top X\hat{\boldsymbol{\beta}} &= 0 \end{aligned}$$

or

$$(X^\top X) \hat{\boldsymbol{\beta}} = X^\top \mathbf{y}$$

$X^\top X$  is a symmetric matrix, and since  $n \gg p$ , it is invertible with high probability, which means that the  $\boldsymbol{\beta}$  that minimizes the squared errors ( $\hat{\boldsymbol{\beta}}$ ) is

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} \quad (5)$$

The components of the vector  $\hat{\boldsymbol{\beta}}$  correspond to the coordinates of the vector in the subspace of  $\mathbb{R}^n$  spanned by the columns of  $X$ , closest to the vector  $\mathbf{y}$ , and written as a linear combination of these columns.

This OLS estimator produces a  $\hat{\mathbf{y}}$  estimation which may be written as

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} = X(X^\top X)^{-1} X^\top \mathbf{y} = \hat{H}\mathbf{y} \quad (6)$$

Where  $\hat{H} \in \mathbb{R}^{n \times n}$  is the orthogonal projector on the  $\text{span}(X)$ <sup>7</sup>. This technique will be very useful when determining business cycles. Equations (5) and (6) are important since they are going to determine the coefficients and trend, respectively, for the approach discussed in the next section.

---

<sup>6</sup>There are several other types of norms used in computer science. The general approach of an  $L^p$ -norm is given by  $\|x\|_p = \left(\sum_{i=0}^N \|x_i\|^p\right)^{1/p}$  while the particular case of the Euclidean norm is when  $p=2$ . See [Lange, Zühlke, Holz, Villmann, and Mittweida \(2014\)](#) for some applications on  $L^p$ -norms.

<sup>7</sup>Note that  $H^2 = [X(X^\top X)^{-1}X^\top][X(X^\top X)^{-1}X^\top] = X(X^\top X)^{-1}(X^\top X)(X^\top X)^{-1}X^\top = H$ .

### 3 Power series expansion approach

Let us assume that a time series can be described by the following linear model

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (7)$$

The evolution of the series is composed by a secular component identified by a linear trend

$$y_t^s = \beta_0 + \beta_1 t$$

while the residuals represent the cyclical component

$$y_t^c = \varepsilon_t$$

The problem is straight forward, there is a parametric approach to the time series where  $\beta_0$  and  $\beta_1$  have to be estimated using a specific criteria (OLS for instance). For our particular case we define the relevant vectors and matrices as

$$\mathbf{y} = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & t \\ 1 & t-1 \\ 1 & t-2 \\ \vdots & \vdots \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

The product  $\mathbf{X}\boldsymbol{\beta}$  results in a linear combination of the columns of  $\mathbf{X}$ , where the first column of  $\mathbf{X}$  (the constant component) is multiplied by the parameter  $\beta_0$  while the second column is multiplied by  $\beta_1$ . Applying (5) we find the coefficients<sup>8</sup>.

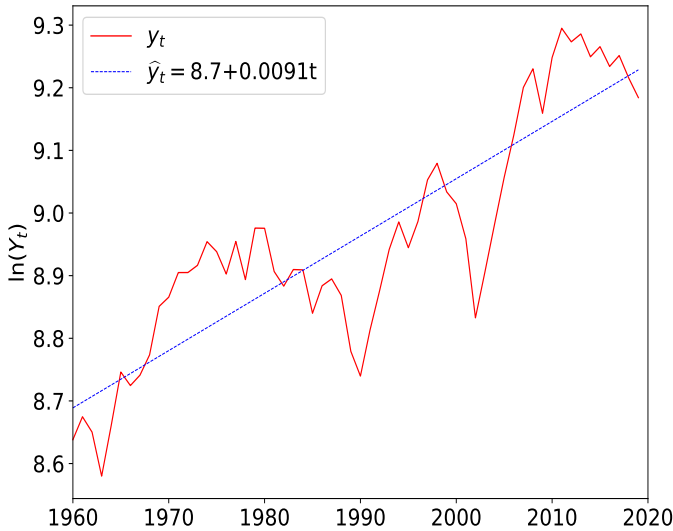


Figure 3: Argentina. Log-Linear Trend

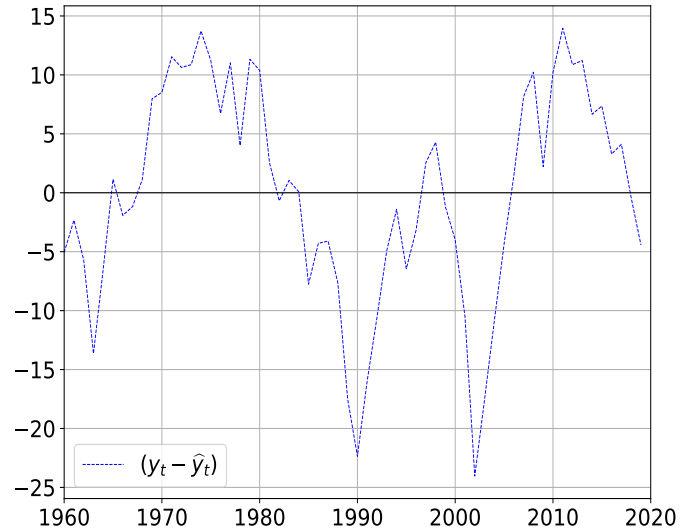


Figure 4: Argentina. Log-Linear Cycle

<sup>8</sup>The Appendix contains the python codes for the linear problem (7).



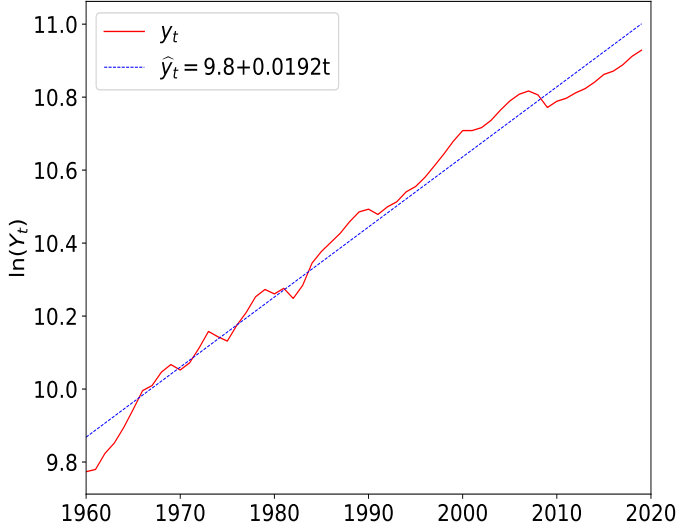


Figure 5: United States. Log-Linear Trend

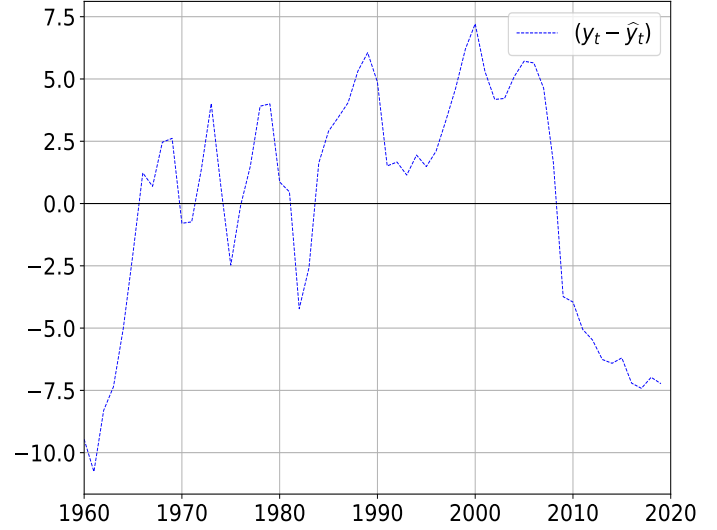


Figure 6: United States. Log-Linear Cycle

What if the proposed model is not the best approach? For instance, what if the addition of a quadratic term fits the data much better than using only a linear term? The OLS methodology easily handles extensions to the model. Moreover, it makes very intuitive to see to what extent we can complexify the model while still remaining meaningful. Consider a quadratic specification of the form,

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t \quad (8)$$

where the secular component is described by

$$y_t^s = \beta_0 + \beta_1 t + \beta_2 t^2$$

while the cyclical component is described by

$$y_t^c = \varepsilon_t$$

For this particular case we define the relevant vectors and matrices as

$$\mathbf{y} = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & t & t^2 \\ 1 & t-1 & (t-1)^2 \\ 1 & t-2 & (t-2)^2 \\ \vdots & \vdots & \vdots \\ 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

The crucial observation is that in  $\mathbb{R}^t$ , the three columns of the new matrix  $\mathbf{X}$  are linearly independent, so for arbitrary  $\beta_i$ ,  $i = 0, 1, 2$ ,  $\mathbf{X}\boldsymbol{\beta}$  span a 3-dimensional subspace in  $\mathbb{R}^t$ . Therefore, once again, using the OLS methodology we estimate  $\beta_0$ ,  $\beta_1$  and  $\beta_2$ . These values determine the vector in the subspace spanned by the columns of  $\mathbf{X}$  closest to  $\mathbf{y}$ <sup>9</sup>.

<sup>9</sup>The Appendix contains the python codes for the quadratic problem (8).

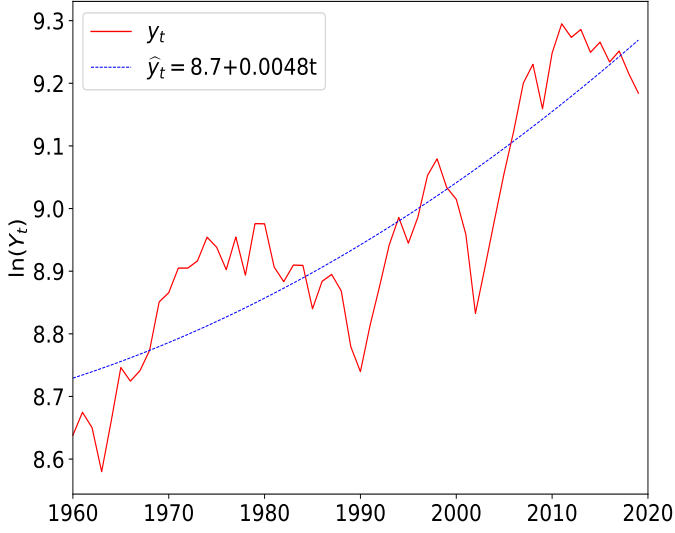


Figure 7: Argentina. Log-Quadratic Trend

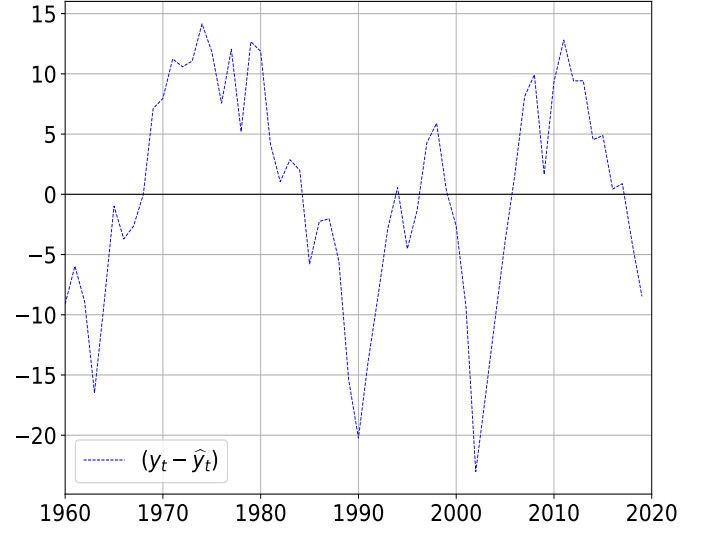


Figure 8: Argentina. Log-Quadratic Cycle

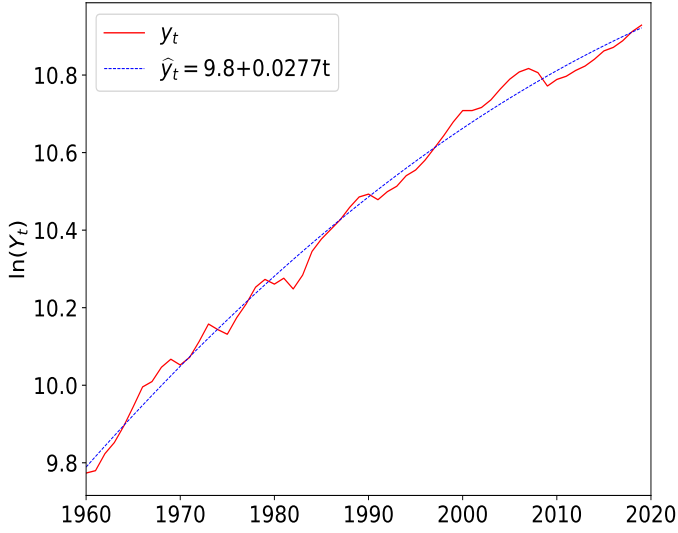


Figure 9: United States. Log-Quadratic Trend

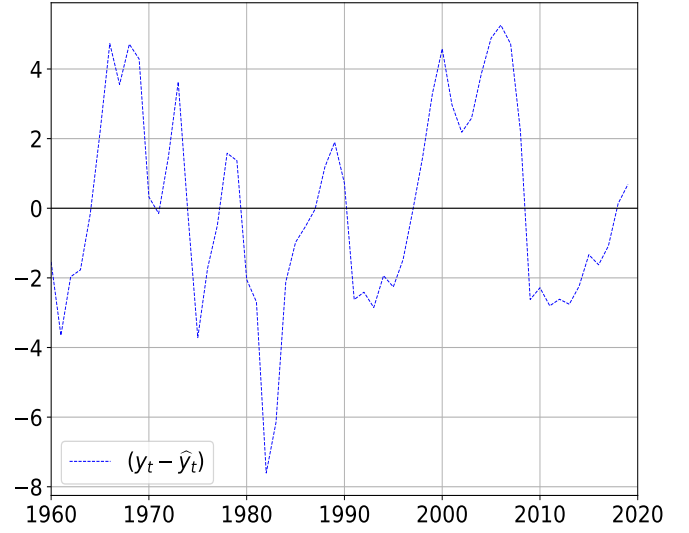


Figure 10: United States. Log-Quadratic Cycle

By now it should be evident that this methodology extends to a  $p$ -degree power series

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t \quad (9)$$

with the relevant vectors and matrices now given by<sup>10</sup>

$$\mathbf{y} = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_0 \end{bmatrix} \quad X = \begin{bmatrix} 1 & t & t^2 & \dots & t^{p-1} & t^p \\ 1 & t-1 & (t-1)^2 & \dots & (t-1)^{p-1} & (t-1)^p \\ 1 & t-2 & (t-2)^2 & \dots & (t-2)^{p-1} & (t-2)^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & 4 & \dots & 2^{p-1} & 2^p \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \\ \beta_p \end{bmatrix}$$

As long as  $p+1 \leq t$ , the  $p+1$  columns of  $X$  are linearly independent and span a  $p+1$  dimensional subspace in  $\mathbb{R}^t$ . However, if  $p \sim t$  we would be over-fitting the model. This linear algebra perspective makes it very intuitive to understand including higher powers of  $t$  makes sense only to the extent that  $k \ll t$ .

## 4 Hodrick–Prescott Filter

The Hodrick–Prescott filter<sup>11</sup> (usually referred to as the HP filter) is a non-linear filter used to remove the cyclical component of a time series from raw data. It is used to obtain a smoothed-curve representation of a time series, one that is more sensitive to long-term than to short-term fluctuations.

$$\begin{aligned} & \underset{\{y_t^s\}_{t=1}^T}{\operatorname{argmin}} \left\{ \sum_{t=1}^T (y_t - y_t^s)^2 + \lambda \sum_{t=2}^{T-1} [(y_{t+1}^s - y_t^s) - (y_t^s - y_{t-1}^s)]^2 \right\} \\ & \text{or} \\ & \underset{\mathbf{y}^s}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{y}^s)^\top (\mathbf{y} - \mathbf{y}^s) + \lambda (A\mathbf{y}^s)^\top (A\mathbf{y}^s) \right\} \\ & \text{or} \\ & \underset{\mathbf{y}^s}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{y}^s\|^2 + \lambda \|A\mathbf{y}^s\|^2 \right\} \end{aligned} \tag{10}$$

where  $\lambda \in \mathbb{R}$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_t \end{bmatrix} \quad \mathbf{y}^s = \begin{bmatrix} y_0^s \\ y_1^s \\ \vdots \\ y_t^s \end{bmatrix} \quad A_{T-2 \times T} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

where  $A$  is the second order difference matrix. Although this is a non-linear filter, when reduced to matrix form, simple linear algebra techniques should do the trick.

<sup>10</sup>In practice it is convenient to form the matrix  $X$  by dividing the elements of the  $j^{\text{th}}$  column by its norm, so that every column of  $X$  has norm 1. This simply re-scales the constants  $\beta_j$ .

<sup>11</sup>Popularized in the 1990s and named after economists Robert J. Hodrick and Edward C. Prescott.

Let's stop for a minute and try to grasp the concept of what this filter actually does. The left term is simply the square of the Euclidian norm (meaning the dot product) of the cyclical component,

$$\sum_{t=1}^T (y_t - y_t^s)^2 = \sum_{t=1}^T y_t^{c2} \Rightarrow (\mathbf{y} - \mathbf{y}^s)^\top (\mathbf{y} - \mathbf{y}^s) = \|\mathbf{y}^c\|_2^2$$

If we were to minimize solely this expression then the optimal solution would be  $y_t - y_t^s = 0 \forall t$ , meaning that the trend would fit the data to perfection providing no information regarding a cycle component.

On the other hand, the right term of the expression has to do with the second order differentiation of  $y_{t+1}^2$ ,

$$\begin{aligned} y_{t+1}^s - 2y_t^s + y_{t-1}^s &= (y_{t+1}^s - y_t^s) - (y_t^s - y_{t-1}^s) \\ &= \Delta y_{t+1}^s - \Delta y_t^s \\ &= \Delta (y_{t+1}^s - y_t^s) \\ &= \Delta (\Delta y_{t+1}^s) \\ &= \Delta^2 y_{t+1}^s \end{aligned}$$

This is strictly related to the curvature of the trend component<sup>12</sup>. For instance,  $\Delta^2 y_{t+1}^s$  is positive if  $\Delta y_{t+1}^s > \Delta y_t^s$ , meaning the trend will exhibit a convex behavior (positive curvature) with respect to the instantly previous path.  $\Delta^2 y_{t+1}^s$  is negative if  $\Delta y_{t+1}^s < \Delta y_t^s$ , meaning the trend will exhibit a concave behavior (negative curvature) with respect to the instantly previous path. The only way that curvature is not affected is if  $\Delta y_{t+1}^s = \Delta y_t^s$ .

This last statement requires that

$$\begin{aligned} \Delta y_{t+1}^s - \Delta y_t^s &= 0 \Rightarrow y_{t+1}^s - 2y_t^s + y_{t-1}^s = 0 \\ y_t^s &= \frac{y_{t+1}^s - y_{t-1}^s}{2} \end{aligned}$$

meaning that  $y_t^s$  has to stay equidistant between the strictly past and consecutive data point along a straight line. But notice something, if we were to minimize the right term, we should apply this equidistant linear concept throughout our whole data set, meaning that three points  $y_{t+1}^s, y_t^s, y_{t-1}^s$  are on the same line throughout all  $T - 1$ . Essentially, the second term in the objective function is zero if and only if  $y_t^s$  is affine, i.e., has the form  $y_t^s = \beta_0 + \beta_1 t$  for some constants  $\beta_0$  and  $\beta_1$ .

We are now able to discuss the role that  $\lambda$  plays. When  $\lambda \rightarrow 0$  there is convergence to the original data. On the other hand,  $\lambda \rightarrow \infty$  the HP estimated trend converges to the best affine (straight-line) fit to the time series data which is identical to the linear case in section 3.

Going back to the problem of minimizing (10) as a whole, using

$$(\mathbf{y} - \mathbf{y}^s)^\top (\mathbf{y} - \mathbf{y}^s) = \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{y}^s + \mathbf{y}^{s\top} \mathbf{y}^s \quad \text{and} \quad \lambda(A\mathbf{y}^s)^\top (A\mathbf{y}^s) = \mathbf{y}^{s\top} \lambda A^\top A \mathbf{y}^s$$

---

<sup>12</sup>It should now be easy to understand why matrix  $A$  is the second order difference matrix.

is possible to differentiate and find the first order condition

$$\begin{aligned}\nabla_{\mathbf{y}^s} \left\{ \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{y}^s + \mathbf{y}^{s\top} \left( I + \lambda A^\top A \right) \mathbf{y}^s \right\} &= 0 \\ \Rightarrow -2\mathbf{y} + 2 \left( I + \lambda A^\top A \right) \mathbf{y}^s &= 0\end{aligned}$$

Finally, we arrive to the conclusion that

$$\mathbf{y}^s = \left( I + \lambda A^\top A \right)^{-1} \mathbf{y} \quad (11)$$

$$\mathbf{y}^c \equiv \mathbf{y} - \mathbf{y}^s = \left[ I - \left( I + \lambda A^\top A \right)^{-1} \right] \mathbf{y} \quad (12)$$

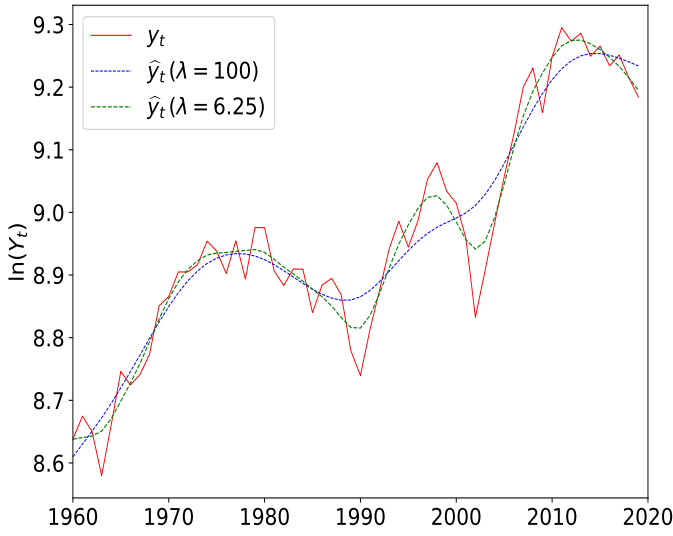


Figure 11: Argentina. HP Filter Trend

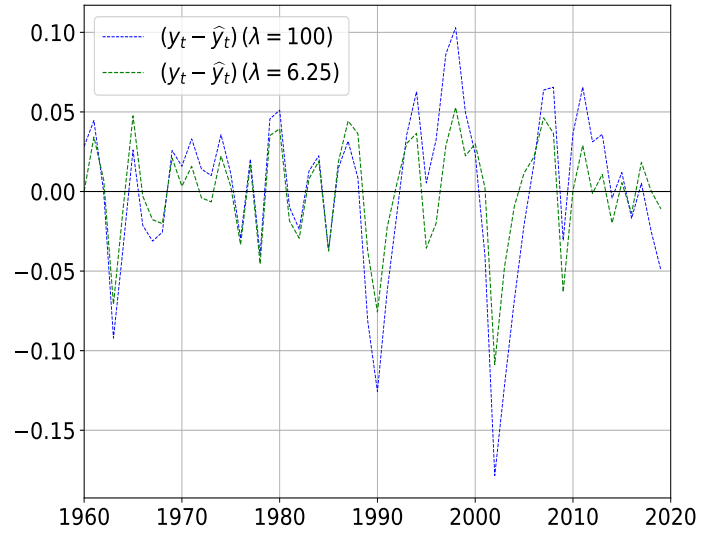


Figure 12: Argentina. HP Filter Cycle

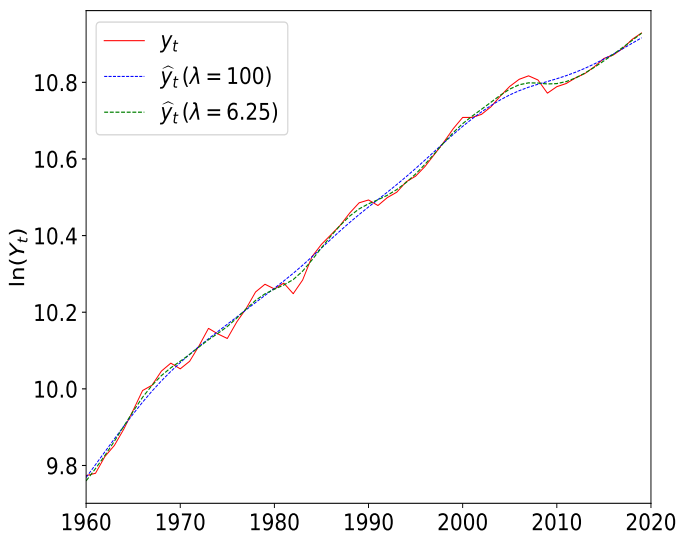


Figure 13: United States. HP Filter Trend

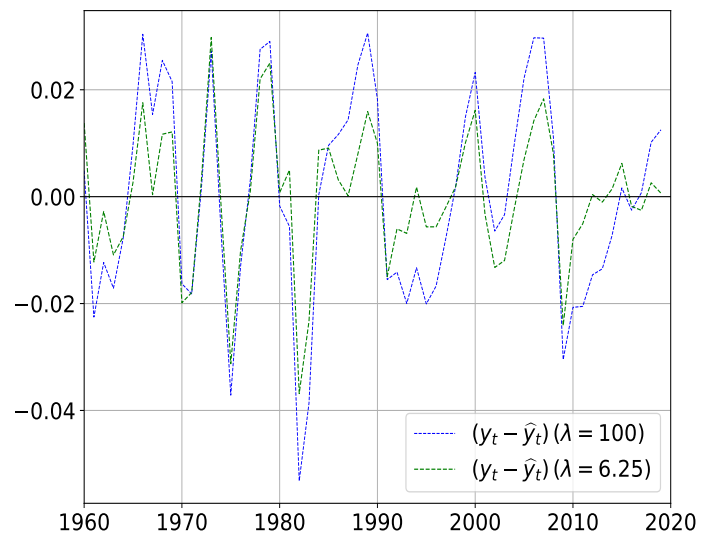


Figure 14: United States. HP Filter Cycle

As we see there is no predetermined ad-hock structure (nor linear or non-linear) for the trend. For the

particular  $\lambda$  of choice we see smaller cycles compared to the previous cases<sup>13</sup>. The discussion of the appropriate  $\lambda$  is a complicated one, particularly since it is not chosen as part of the optimization process. That said, there are commonly used values of  $\lambda$  according to the frequency of the data used. The most common formula is

$$\lambda = 100 \times w^2$$

where  $w$  represents the number of periods in one year.

Data Frequency	$\lambda = 100 \times w^2$	$\lambda$
Yearly	$100 \times 1^2$	100
Quarterly	$100 \times 4^2$	1600
Monthly	$100 \times 12^2$	14400
Weekly	$100 \times 52^2$	270400

Generally, GDP data has either anual or quarterly frequency. Ravn and Uhlig (2002) argue that  $\lambda = 6.25$  should be used annual data. Kaiser and Maravall (1999) propose a  $\lambda = 8$  for annual data, and Pedersen (2001) argues for a value of  $\lambda = 1000$  for quarterly data and for  $\lambda \in [3, 5]$  for annual data. In Bouthevillain et al. (2001) the filter is applied with  $\lambda = 30$  and in Mohr (2001) with  $\lambda = 20$  to annual data. As shown, there is a lot of discussion regarding the adequate value for  $\lambda$ .

## 5 Conclusions

As mentioned in the introduction, the purpose of this series is to present in an intuitive way the mathematical technics that students of economics use, so that they can follow every step of a derivation and not, as is often the case, using the technics as black boxes. In particular, for time series with OLS, students are often confused about the consistency of the application of linear techniques while including in the series terms that depend on  $t^n$  with  $n > 1$ . This problem is a particular case of the general technique of including nonlinear terms in regression-type models where, again, students are often confused about the linearity of the underlying problem and the simultaneous nonlinearities of the terms in the regression. The hope is that this geometric way of presenting the problem, emphasizing the fact that what is really crucial is the linear independence of the resulting columns of the matrix  $X$ , which is in general automatic for different powers of  $t$  (as long as  $n > k$ ), clarifies the point.

---

<sup>13</sup>The Appendix contains the python codes for the minimization problem (10).

## 6 Appendix

### 6.1 Log-Linear Detrending

```
1 #Log-Linear Detrending
2 #@author: Tomas Marinozzi, Leandro Nallar and Sergio Pernice
3
4 import pandas as pd
5 import numpy as np
6 import matplotlib.pyplot as plt
7
8 GDP_Main = pd.read_excel (r'/Users/.../Documents/.../.../GDP_data.xls', sheet_name="USE")
9 GDP_Numbers=GDP_Main.values
10
11 S=np.column_stack((np.transpose(np.matrix(GDP_Numbers[:,0 ])),np.transpose(np.matrix(np.
    log(GDP_Numbers[:,1])))))
12
13 x_1=np.ones((len(S[:,1]),1))
14 x_2=np.transpose(np.matrix(np.linspace(1,len(S[:,1]),len(S[:,1]))))
15 X=np.hstack((x_1,x_2))
16
17 BETA=np.matmul(np.linalg.inv(np.matmul(np.transpose(X),X)),np.matmul(np.transpose(X),S
   [:,1]))
18
19 GDP_trend=np.matmul(X,BETA)
20 GDP_cycle=(S[:,1]-GDP_trend)*100
```

### 6.2 Log-Quadratic Detrending

```
1 #Log-Quadratic Detrending
2 #@author: Tomas Marinozzi, Leandro Nallar, Sergio Pernice
3
4 import pandas as pd
5 import numpy as np
6 import matplotlib.pyplot as plt
7
8 GDP_Main = pd.read_excel (r'/Users/.../Documents/.../.../GDP_data.xls', sheet_name="USE")
9 GDP_Numbers=GDP_Main.values
10
11 S=np.column_stack((np.transpose(np.matrix(GDP_Numbers[:,0 ])),np.transpose(np.matrix(np.
    log(GDP_Numbers[:,2])))))
12 #S = S[~np.isnan(S).any(axis=1)]
13
14 x_1=np.ones((len(S[:,1]),1))
15 x_2=np.transpose(np.matrix(np.linspace(1,len(S[:,1]),len(S[:,1]))))
```

```

16 x_3=np.power(x_2,2)
17 X=np.hstack((x_1,x_2,x_3))
18
19 BETA=np.matmul(np.linalg.inv(np.matmul(np.transpose(X),X)),np.matmul(np.transpose(X),S
    [:,1]))
20
21 GDP_trend=np.matmul(X,BETA)
22 GDP_cycle=(S[:,1]-GDP_trend)*100

```

### 6.3 HP Filter

```

1 #HP Filter
2 #@author: Tomas Marinozzi, Leandro Nallar, Sergio Pernice
3
4 import statsmodels.api as sm
5 import pandas as pd
6 import numpy as np
7 import matplotlib.pyplot as plt
8
9 GDP_Main = pd.read_excel (r'/Users/.../Documents/.../.../GDP_data.xls',sheet_name="USE")
10 GDP_Numbers=GDP_Main.values
11
12 S=np.column_stack((np.transpose(np.matrix(GDP_Numbers[:,0 ])),np.transpose(np.matrix(np.
    log(GDP_Numbers[:,1])))))
13 #S = S[~np.isnan(S).any(axis=1)]
14
15 GDP_cycle_100, GDP_trend_100 = sm.tsa.filters.hpfilter(S[:,1], 100)
16 GDP_cycle_6d25, GDP_trend_6d25 = sm.tsa.filters.hpfilter(S[:,1], 6.25)

```



## References

- Burns, A. (1946). *Burns and mitchell on business cycles*. National Bureau of Economic Research.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.
- Lange, M., Zühlke, D., Holz, O., Villmann, T., & Mittweida, S.-G. (2014). Applications of lp-norms and their smooth approximations for gradient based learning vector quantization. In *Esann* (pp. 271–276).
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Pernice, S. (2019). Intuitive mathematical economics series. linear structures i. linear manifolds, vector spaces and scalar products. *Linear Structures I. Linear Manifolds, Vector Spaces and Scalar Products (March 2019)*. *Serie Documentos de Trabajo-Nro*, 689.
- Phillips, P. C., & Jin, S. (2015). Business cycles, trend elimination, and the hp filter.
- Uribe, M., & Schmitt-Grohé, S. (2017). *Open economy macroeconomics, ch 1*. Princeton University Press.