Automated Breakdown Analysis

A case study of the cost price of French fries potatoes

Jamal Roskam[*]   Paul van Leeuwen[†]

3 June 2021

Abstract

[*] Wageningen Economic Research, jamal.roskam@wur.nl
[†] Wageningen Economic Research, paul2.vanleeuwen@wur.nl

Contents

1 Introduction

@Jamal, wil hier iets over schrijven?

2 Methodology

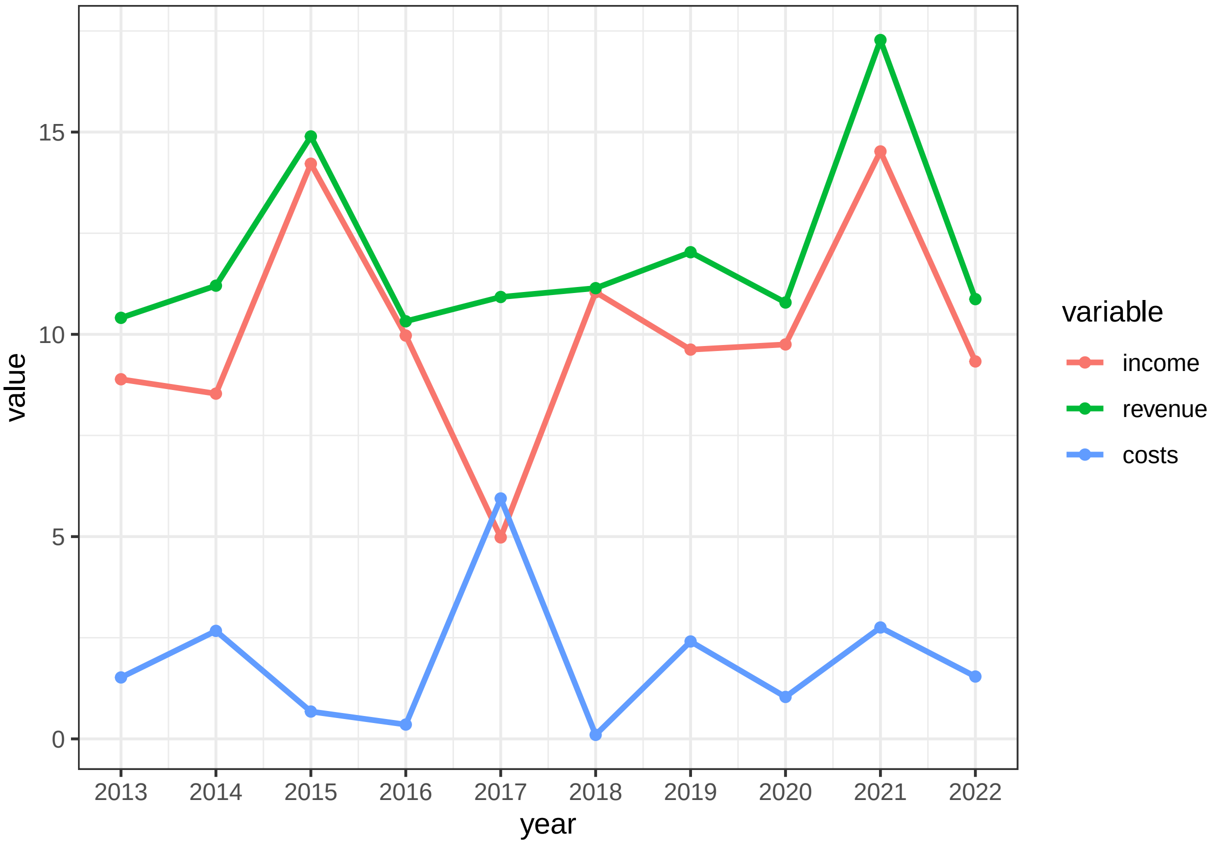
We describe three approaches to analyse an arbitrary time series: analysis of change, sensitivity analysis, and variance decomposition analysis.

2.1 Analysis of Change

Given a time series variable $\boldsymbol{y} \in \mathbb{R}^T$ to be studied and defined as the dependent variable. $\boldsymbol{y}$ is the result of a mapping from $m \geq 1$ other variables that are gathered in the matrix $\boldsymbol{X} \in \mathbb{R}^{Tm}$ with $T \geq 1$ and $\boldsymbol{X} = (\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_m)'$ where $\boldsymbol{x}_j$ is the j^{th} column of $\boldsymbol{X}$. The mapping $f \colon \mathbb{R}^m \mapsto \mathbb{R}$ is defined as $\boldsymbol{y}_t = f(x_1, \ldots, x_m)$ where $t = 1, \ldots, T$ and $j = 1, \ldots, m$ unless indicated otherwise.

Example
Suppose $\boldsymbol{y}$ is the average annual income of a farmer over a period of T years. $\boldsymbol{y}$ is constructed as $\boldsymbol{y}_t = f(\boldsymbol{r}_t, \boldsymbol{c}_t) = \boldsymbol{r}_t - \boldsymbol{c}_t$ with $\boldsymbol{X} = (\boldsymbol{r} \quad \boldsymbol{c})$, $\boldsymbol{r} \in \mathbb{R}_+^T$ the revenue, and $\boldsymbol{c} \in \mathbb{R}_+^T$ the costs. Furthermore, let $T = 10$, $\ln \boldsymbol{r}_t \sim \mathcal{N}(10, 1)$, and $\ln \boldsymbol{r}_t \sim \mathcal{N}(0, 1)$. See Figure 1 for a visualisation of the income and its decomposition into revenue and costs.

Figure 1: Visualisation of the income and its decomposition into revenue and costs for the period 2013 - 2022.



2.1.1 Change With Respect to Smallest Time Granularity

The change in $\boldsymbol{y}$ over the periods can be explained by the change in the underlying variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ over the same period. More mathematically, for $T \geq 2$ let the change in $\boldsymbol{y}$ from period $t-1$ to period t be denoted by

$$\Delta(\boldsymbol{y}_t) = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}, \quad t = 2, \ldots, n$$

Then $\Delta(\boldsymbol{y}_t)$ is explained by a change in $\boldsymbol{x}_j$ from period $t-1$ to period t as

$$\tilde{\nabla}_{tj} = \begin{cases} 0 & \text{when } \Delta(\boldsymbol{y}_t) = 0 \\ \dfrac{\boldsymbol{y}_t - f(\boldsymbol{X}_{t1} \ldots, \boldsymbol{X}_{t,j-1}, \boldsymbol{X}_{t-1,j}, \boldsymbol{X}_{t,j+1}, \ldots, \boldsymbol{X}_{tm})}{\Delta(\boldsymbol{y}_t)} & \text{when } \Delta(\boldsymbol{y}_t) \neq 0 \end{cases} \tag{1}$$

with $\boldsymbol{X}_{tj}$ the element of $\boldsymbol{X}$ in row t and column j. In other words, for variable $\boldsymbol{x}_j$ only the t^{th} observation $\boldsymbol{X}_{tj}$ is replaced by its one-period lagged version $\boldsymbol{X}_{t-1,j}$. $\tilde{\nabla}_{tj}$ is defined as zero when $\Delta(\boldsymbol{y}_t)$ is zero because there is no change in $\boldsymbol{y}$ from period $t-1$ to period t. Hence, the change from $\boldsymbol{y}_{t-1}$ to $\boldsymbol{y}_t$ cannot be related to a change of $\boldsymbol{X}_{t-1,j}$ to $\boldsymbol{X}_{tj}$ by only considering $\boldsymbol{y}_{t-1}$, $\boldsymbol{y}_t$, $\boldsymbol{X}_{t-1,j}$, $\boldsymbol{X}_{tj}$, and the mapping f.[1]

To enable comparison of the importance of a change in one of the underlying variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ over time, the $\tilde{\nabla}_{tj}$ need to be positive and they add up to a constant value. Without loss of generality, we require the $\tilde{\nabla}_{tj}$ to add up to one.

Only when f is an affine combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ and $\Delta(\boldsymbol{y}_t) \neq 0$ the $\tilde{\nabla}_{tj}$ add up to one. This is shown as follows. Let f be defined as

$$f(x_1, \ldots, x_m) = a + \sum_{j=1}^{m} b_j x_j, \quad a \in \mathbb{R} \quad \text{and} \quad b_j \in \mathbb{R} \tag{2}$$

Then

$$\tilde{\nabla}_{tj} = \frac{a + \sum\limits_{k=1}^{m} b_k \boldsymbol{X}_{tk} - \left(b_j \boldsymbol{X}_{t-1,j} + a + \sum\limits_{k=1, k \neq j}^{m} b_k \boldsymbol{X}_{tk} \right)}{\Delta(\boldsymbol{y}_t)} = b_j \frac{\boldsymbol{X}_{tj} - \boldsymbol{X}_{t-1,j}}{\Delta(\boldsymbol{y}_t)}$$

resulting in

$$\sum_{j=1}^{m} \tilde{\nabla}_{tj} = \frac{a + \sum_{j=1}^{m} b_j \boldsymbol{X}_{tj} - a - \sum_{j=1}^{m} b_j \boldsymbol{X}_{t-1,j}}{\Delta(\boldsymbol{y}_t)} = \frac{\Delta(\boldsymbol{y}_t)}{\Delta(\boldsymbol{y}_t)} = 1$$

When, in addition, the sign of all b_j is equal to the sign of $\Delta(\boldsymbol{y}_t)$ the $\tilde{\nabla}_{tj}$ are non-negative and therefore $0 \leq \tilde{\nabla}_{tj} \leq 1$.

When f is not an affine combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ while $\Delta(\boldsymbol{y}_t) \neq 0$ the $\tilde{\nabla}_{tj}$ do not add up to one and therefore the $\tilde{\nabla}_{tj}$ need to be normalised:

$$\nabla_{tj} = \begin{cases} 0 & \text{when } \tilde{\nabla}_{tj} = 0 \text{ for all } j \in \{1, \ldots, m\} \\ \dfrac{\left|\tilde{\nabla}_{tj}\right|}{\sum_{j=1}^{m} \left|\tilde{\nabla}_{tj}\right|} & \text{when } \tilde{\nabla}_{tj} \neq 0 \text{ for at least one } j \in \{1, \ldots, m\} \end{cases} \tag{3}$$

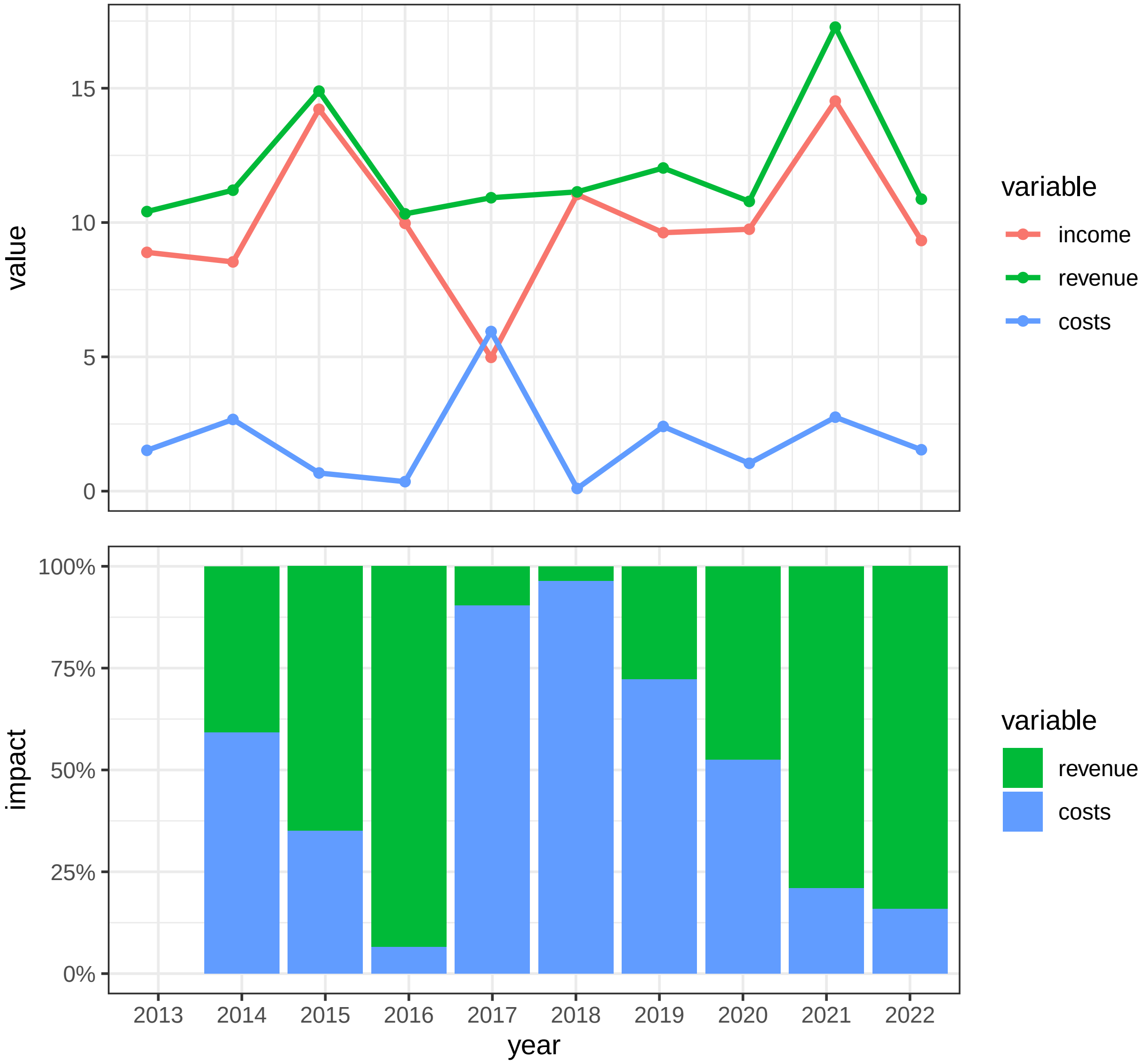
where the absolute value of each $\tilde{\nabla}_{tj}$ is taken to ensure the impact of the change in $\boldsymbol{x}_j$ is not cancelled out.

[1] Note that a non-zero change of $\boldsymbol{X}_{t-1,j}$ to $\boldsymbol{X}_{tj}$ could lead to $\Delta(\boldsymbol{y}_t) = 0$ but that requires an analysis of the mapping f and not just Equation (1). For example, when $f(x_1, x_2) = x_1 - x_2$ and $x_1 = x_2$ for all periods we have $\Delta(\boldsymbol{y}_t) = 0$ although x_1 and x_2 are not necessarily zero. The use of the derivative of f with respect to the underlying variables is described below.

Note that, in general, ∇_{tj} does not explain the fraction of $\Delta(\boldsymbol{y}_t)$. More specifically, they show the impact on the change of $\boldsymbol{y}_t$ due to a change in one of its underlying variables. Alos note that ∇_{tj} only explains the fraction of $\Delta(\boldsymbol{y}_t)$ when $\tilde{\nabla}_{tj} = \Delta(\boldsymbol{y}_t)$.

Example (continued)
In the example above we see a maximal change in income of in 2018. Of that change 3.61% is explained by a change in revenue (+0.22) and 96.39% by a change in costs (-5.84). The decrease in costs is larger than the increase in revenue, hence the most relevant change is in the costs. In this manner an arbitrary function can be analysed with respect to its underlying variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ as shown in Equations (1) and (3). Figure 1 extended with the analysis of this section is shown in Figure 2. For the sake of convenience, Figure 1 is shown again.

Figure 2: Visualisation of the income and its decomposition into revenue and costs (upper panel) and a visualisation of the impact of a change in revenue and costs with respect to income (lower panel). Both panels are for the time window 2013 - 2022. Note that the first year is omitted in the lower panel as change can only be calculated from the second year on. Impact is ∇_{tj} from Equation (3).



2.1.2 (Experimental) Alternative derivation: Taylor approximation

The Taylor approximation of an n-times differentiable function $g\colon \mathbb{R} \mapsto \mathbb{R}$ at $x = x_0$ could be expressed as

$$g(x) = \frac{(x-x_0)^n}{n!}\, g^{(n)}(x)\Big|_{x=x_0} + \sum_{k=0}^{n-1} \frac{(x-x_0)^k}{k!}\, g^{(k)}(x)\Big|_{x=x_0}$$

for some $\xi \in]x, x_0[$. Let $x = \boldsymbol{X}_{tj}$ for variable j to be analysed, $x_0 = \boldsymbol{X}_{t-1,j}$, and $\Delta = \boldsymbol{X}_{tj} - \boldsymbol{X}_{t-1,j}$. Subtracting $g(x_0) = g\left(\boldsymbol{X}_{j,t-1}\right)$ from both sides leads to

$$\tilde{\nabla}_{tj} = g\left(\boldsymbol{X}_{j,t}\right) - g\left(\boldsymbol{X}_{j,t-1}\right) = \frac{\Delta^n}{n!}\, g^{(n)}(x)\Big|_{x=\xi} + \sum_{k=1}^{n-1} \frac{\Delta^k}{k!}\, g^{(k)}(x)\Big|_{x=\boldsymbol{X}_{j,t-1}} \tag{4}$$

In other words, a change from $\boldsymbol{X}_{j,t-1}$ to $\boldsymbol{X}_{jt}$ causes a change from $g\left(\boldsymbol{X}_{j,t-1}\right)$ to $g\left(\boldsymbol{X}_{jt}\right)$. The latter change is calculated by the right-hand side of Equation (4).

To assess the impact of a change from $\boldsymbol{X}_{j,t-1}$ to $\boldsymbol{X}_{jt}$ on the change from $g\left(\boldsymbol{X}_{j,t-1}\right)$ to $g\left(\boldsymbol{X}_{jt}\right)$, we need to divide $g\left(\boldsymbol{X}_{jt}\right) - g\left(\boldsymbol{X}_{j,t-1}\right)$ by $\boldsymbol{X}_{jt} - \boldsymbol{X}_{j,t-1}$, i.e.

$$\tilde{\nabla}_{tj} = \frac{g\left(\boldsymbol{X}_{jt}\right) - g\left(\boldsymbol{X}_{j,t-1}\right)}{\boldsymbol{X}_{jt} - \boldsymbol{X}_{j,t-1}}$$

Example

Suppose $f(x_1, x_2) = x_1 + x_2$. Then the first derivative is one for both x_1 and x_2. The second derivative of f and beyond evaluates to zero. Then, for $j = 1, 2$, Equation (4) evaluates to

$$\tilde{\nabla}_{t,1} = \boldsymbol{X}_{t,1} - \boldsymbol{X}_{t-1,1} \quad \text{and} \quad \tilde{\nabla}_{t,2} = \boldsymbol{X}_{t,2} - \boldsymbol{X}_{t-1,2}$$

Intuitively, this decomposition is appealing as the sum over all $\tilde{\nabla}_{tj}$ for all j leads to

$$\sum_{j=1}^{m} \tilde{\nabla}_{tj} = \Delta(\boldsymbol{y}_t)$$

For a function f that is different than a polynomial function with non-negative powers the Taylor approximation may lead to an exploding error term $\frac{\Delta^n}{n!} g^{(n)}(\xi)$.

Vind een functie die aan de axiomen voldoet: geen verandering in x is nabla gelijk aan 0 en alle nabla tellen op tot 1 en zijn niet-negatief.

2.2 Sensitivity Analysis

Suppose f is differentiable once on its domain.[2] Analogously to Section 2.1.1, Equation (3) in particular, we can analyse the instantaneous change of f with respect to the underlying variables x_j,

$$d_{tj} = \begin{cases} 0 & \text{when } \tilde{d}_{tk} = 0 \text{ for all } k \in \{1, \ldots, m\} \\ \dfrac{\left|\tilde{d}_{tj}\right|}{\sum_{k=1}^{m} \left|\tilde{d}_{tk}\right|} & \text{when } \tilde{d}_{tk} \neq 0 \text{ for at least one } k \in \{1, \ldots, m\} \end{cases} \tag{5}$$

with $\tilde{d}_{tj} = \dfrac{\partial}{\partial x_j} f(x_1, \ldots, x_m)\Bigg|_{x_1 = \boldsymbol{X}_{t1}, \ldots, x_m = \boldsymbol{X}_{tm}}$. As with Equation (3), the $\tilde{d}_k$ need to be normalised to enable comparison of $\tilde{d}_k$ over time.

[2] An example of a function that is differentiable once but not twice on $\mathbb{R}$ is $f(x) = x|x|$.

In the example of Section 2.1 the derivative of the income with respect to revenue (costs) is 1 (-1) and therefore $d_{tj} = \frac{1}{2}$ for all t and j. A more interesting example of a mapping f would be when d_{tj} is not a constant for all t and j.

Example
Let $f\colon \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$ and take $\boldsymbol{y}$ to be the income per annual working unit, $\boldsymbol{x}_1$ the income as before, and $\boldsymbol{x}_2$ the number of annual working units (awu), i.e. $f(x_1, x_2) = x_1/x_2$. Furthermore, let $T = 10$ and the number of awu is sampled from $1 + \dfrac{2}{1+\exp(a)}$ with $a \sim \mathcal{N}(0,1)$. See Figure 3 for a visualisation of the number of the income per awu for different values of income and the number of awu. See Figure 4 for a visualisation over time of the income per awu (upper panel) and its decomposition into income and the number of awu (lower panel). Working out the mathematics of Equation (5) yields

$$\tilde{d}_{t,1} = \frac{1}{\boldsymbol{X}_{t,2}} \quad \text{and} \quad \tilde{d}_{t,2} = -\frac{\boldsymbol{X}_{t,1}}{\boldsymbol{X}_{t,2}^2}$$

leading to

$$d_{t,1} = \frac{\left|\frac{1}{\boldsymbol{X}_{t,2}}\right|}{\left|\frac{1}{\boldsymbol{X}_{t,2}}\right| + \left|\frac{\boldsymbol{X}_{t,1}}{\boldsymbol{X}_{t,2}^2}\right|} = \frac{\boldsymbol{X}_{t,2}}{\boldsymbol{X}_{t,2} + \boldsymbol{X}_{t,1}} \quad \text{and} \quad d_{t,2} = \frac{\left|\frac{\boldsymbol{X}_{t,1}}{\boldsymbol{X}_{t,2}^2}\right|}{\left|\frac{1}{\boldsymbol{X}_{t,2}}\right| + \left|\frac{\boldsymbol{X}_{t,1}}{\boldsymbol{X}_{t,2}^2}\right|} = \frac{\boldsymbol{X}_{t,1}}{\boldsymbol{X}_{t,2} + \boldsymbol{X}_{t,1}} \tag{6}$$

Note the absence of the absolute signs as all values of income and the number of awu are strictly positive.
As becomes apparent from Figure 4, the sensitivity is mostly determined by the variable the number of awu. Intuitively this can be explained by the observation that income per awu is more sensitive to a change in the number of awu than a change in income since the division by the number of awu has more impact than the multiplication by income. More mathematically, an infinite small change in the number of awu has a larger impact on the income per awu compared to an infinite small change in income. This follows from the observation that the number of awu resides in the interval $[1.2, 2.8]$ while income resides in the interval $[5, 15]$. Since $d_{t,1}$ and $d_{t,2}$ have the same denominator, the d_{tj} with the largest absolute value of the numerator is by definition the largest sensitivity factor.

Misschien ook nog richting per onderliggend variabele aangeven? JR: ja, als optie in d interactiev applicatie

2.2.1 Change With Respect to Groups

Suppose $\boldsymbol{y}$ is given for a group. In general, when categorical variable $\boldsymbol{x}_j$ has $c \in \mathbb{N}$ categories with the value of category $\ell = 1, \ldots, c$ defined by $\boldsymbol{X}_{tj\ell}$ and $\boldsymbol{X}_{tj} = \sum_{\ell=1}^{c} \boldsymbol{X}_{tj\ell}$. Note that $d_{tj} = 1/m$ for this type of variable.

Figure 3: Visualisation of the income per awu and its decomposition into income and the number of awu. The sensitivity is maximal for a high income and a low number of awu. That is partially because, when the number of awu is low, an increase in the number of awu for a certain income leads to a much larger decrease in income per awu compared to the same change in the number of awu when the number of awu is higher. The other reason is that, for a given number of awu, a higher income has a larger effect on the income per awu compared to a lower income.

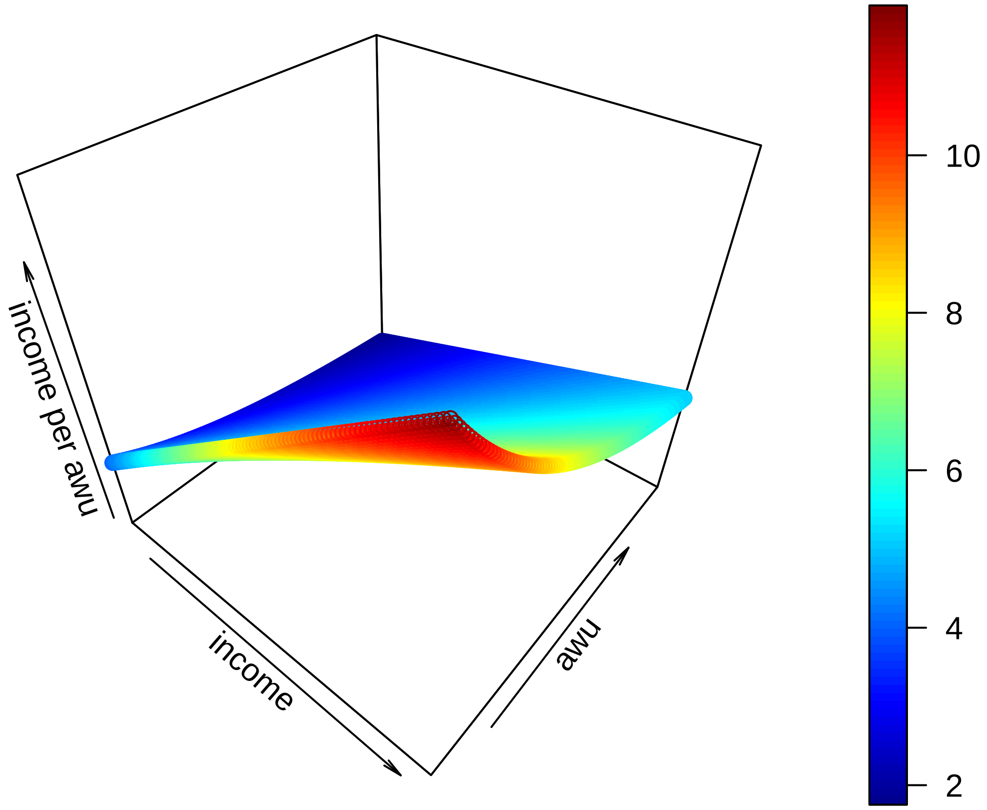
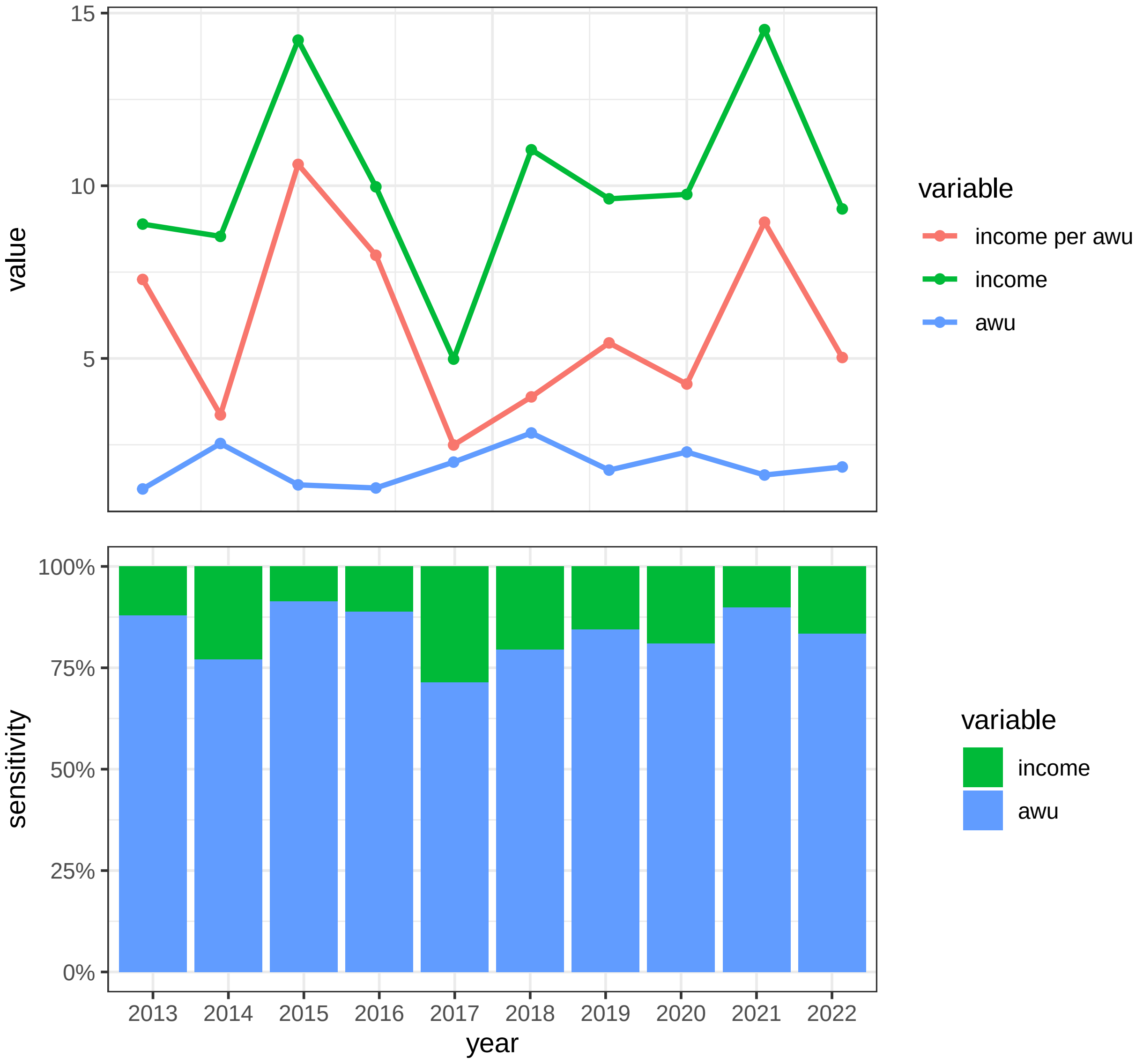


Figure 4: Visualisation of the income per awu and its decomposition into income and the number of awu (upper panel) and a visualisation of the sensitivity with respect to income and the number of awu (lower panel). Both panels are for the time window 2013 - 2022. Sensitivity is d_{tj} from Equation (6).



Example

Let $\boldsymbol{y}$ be again the average annual income of farmers and $\boldsymbol{y}$ is given for different types of soil. More mathematically, the soil type is denoted in the $(n \times 1)$ vector $\boldsymbol{x}$ where $\boldsymbol{x}_t$ corresponds to $\boldsymbol{y}_t$ and $\boldsymbol{x}_t$ can take the values

$$\{\text{bog}, \text{clay}, \text{loess}, \text{sand}\}$$

with the corresponding income for each soil type given by $\boldsymbol{y}_t^{\text{bog}}$, $\boldsymbol{y}_t^{\text{clay}}$, $\boldsymbol{y}_t^{\text{loess}}$, and $\boldsymbol{y}_t^{\text{sand}}$, respectively. Furthermore,

$$\boldsymbol{y}_t^{\text{bog}} \geq 0, \quad \boldsymbol{y}_t^{\text{clay}} \geq 0, \quad \boldsymbol{y}_t^{\text{loess}} \geq 0, \quad \boldsymbol{y}_t^{\text{sand}} \geq 0, \quad \boldsymbol{y}_t = \boldsymbol{y}_t^{\text{bog}} + \boldsymbol{y}_t^{\text{clay}} + \boldsymbol{y}_t^{\text{loess}} + \boldsymbol{y}_t^{\text{sand}}$$

Then the following mapping in conjunction with Equations (3) and (5) can be used to analyse the change in $\boldsymbol{y}_t$ with respect to the change in the underlying variables $\boldsymbol{y}_t^{\text{bog}}$, $\boldsymbol{y}_t^{\text{clay}}$, $\boldsymbol{y}_t^{\text{loess}}$, and $\boldsymbol{y}_t^{\text{sand}}$:

$$f(x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4 \quad \text{with} \quad x_1 = \boldsymbol{y}_t^{\text{bog}}, \quad x_2 = \boldsymbol{y}_t^{\text{clay}}, \quad x_3 = \boldsymbol{y}_t^{\text{loess}}, \quad x_4 = \boldsymbol{y}_t^{\text{sand}}$$

2.3 Variance Decomposistion Analysis

Up to now we only considered a time series for a single entity, e.g. a single farm. When the time series of multiple entities are aggregated into a single time series, dispersion among the measured variable could arise. For example, income is likely to be different over time for different farmers. The corresponding variance can be decomposed into the variance of the underlying variables. More formally, for $n \geq 1$ entities, the variable to be decomposed $\boldsymbol{Y} \in \mathbb{R}^{nT}$ is now of length nT instead of T as before.[3] In other words, $\boldsymbol{Y}$ is the $(nT \times 1)$ vector of stacked values of the dependent variable $\boldsymbol{y}$ for each entity. Suppose entity i has $\boldsymbol{y}$ as dependent variable, then at time t the corresponding element of $\boldsymbol{Y}$ is $\boldsymbol{Y}_{T(i-1)+t}$.

The aggregation of $\boldsymbol{Y}$ in period t is defined to be a weighted average of the associated values for the entities, i.e. $\boldsymbol{y}_t = \sum_{i=1}^{n} \boldsymbol{W}_{T(i-1)+t} \boldsymbol{Y}_{T(i-1)+t}$ with weight $\boldsymbol{W}_{T(i-1)+t} \in [0, 1]$ and, for all t, $\boldsymbol{W}_{T(i-1)+t} > 0$ for at least one i. As with $\boldsymbol{Y}$, for entity i at time t the corresponding element of $\boldsymbol{W}$ is $\boldsymbol{W}_{T(i-1)+t}$.

As the underlying variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ could exhibit dispersion they become random variables relative to the information set $\{\boldsymbol{y}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ and are denoted by, respectively, $X_1, \ldots, X_m$. The variance of $f(X_1, \ldots, X_m)$ can be decomposed into the variances of $X_1, \ldots, X_m$ as follows. For now we assume that f is an affine mapping, i.e. f is defined as in Equation (2). Then

$$\text{Var}(f) = \sum_{j=1}^{m} a_j^2 \text{Var}(X_j) + 2 \sum_{k=1, k \neq j}^{m} a_j a_k \text{Cov}(X_j, X_k)$$

with $\text{Var}(X)$ the variance of the random variable X,

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \tag{7}$$

and $\text{Cov}(X, Y)$ the covariance between X and Y,

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \tag{8}$$

Let

$$\tilde{v}_k = a_k^2 \text{Var}(X_k) + \sum_{k=1, k \neq j}^{m} a_j a_k \text{Cov}(X_j, X_k)$$

[3] Note that we implicitly assume that the dependent variable $\boldsymbol{y}$ is observed for each period t and each entity i. Whether or not this assumption is applicable, is not of relevance for the analysis in this section.

and

$$v_k = \begin{cases} 0 & \text{when } \tilde{v}_k = 0 \text{ for all } k \in \{1, \dots, m\} \\ \dfrac{|\tilde{v}_k|}{\sum_{k=1}^m |\tilde{v}_k|} & \text{when } \tilde{v}_k \neq 0 \text{ for at least one } k \in \{1, \dots, m\} \end{cases}$$

When desired, one could break up v_k into the variance and the covariance part,

$$\tilde{v}_k^{\text{Var}} = a_k^2 \text{Var}(X_k) \quad \text{and} \quad \tilde{v}_k^{\text{Cov}} = \sum_{k=1, k \neq j}^{m} a_j a_k \text{Cov}(X_j, X_k)$$

leading to

$$v_k^{\text{Var}} = \begin{cases} 0 & \text{when } \tilde{v}_k^{\text{Var}} = 0 \text{ and } \tilde{v}_k^{\text{Cov}} = 0 \text{ for all } k \in \{1, \dots, m\} \\ \dfrac{\left|\tilde{v}_k^{\text{Var}}\right|}{\sum_{k=1}^m |\tilde{v}_k^{\text{Var}}| + |\tilde{v}_k^{\text{Cov}}|} & \text{when } \tilde{v}_k^{\text{Var}} \neq 0 \text{ or } \tilde{v}_k^{\text{Cov}} \neq 0 \text{ for at least one } k \in \{1, \dots, m\} \end{cases}$$

and

$$v_k^{\text{Cov}} = \begin{cases} 0 & \text{when } \tilde{v}_k^{\text{Var}} = 0 \text{ and } \tilde{v}_k^{\text{Cov}} = 0 \text{ for all } k \in \{1, \dots, m\} \\ \dfrac{\left|\tilde{v}_k^{\text{Cov}}\right|}{\sum_{k=1}^m |\tilde{v}_k^{\text{Var}}| + |\tilde{v}_k^{\text{Cov}}|} & \text{when } \tilde{v}_k^{\text{Var}} \neq 0 \text{ or } \tilde{v}_k^{\text{Cov}} \neq 0 \text{ for at least one } k \in \{1, \dots, m\} \end{cases}$$

For more than one period v_k, v_k^{Var}, and v_k^{Cov} become time-dependent and the dependency on time is denoted by, respectively, v_{tk}, v_{tk}^{Var}, and v_{tk}^{Cov}. As with Equation (3), the $\tilde{v}_{tk}$ (and $\tilde{v}_{tk}^{\text{Var}}$ and $\tilde{v}_{tk}^{\text{Cov}}$ when desired) need to be normalised to enable comparison of $\tilde{v}_{tk}$ over time.

Apart from very specific data-generating processes the variance (Equation (7)) and covariance (Equation (8)) can only be estimated from the observed values, i.e. for a random variable X with $n \geq 2$ observed values $(X_1, \dots, X_n)$, we have sample variance

referentie
invoegen

$$S_n = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and only for $n \to \infty$ we have that S_n equals $\text{Var}(X)$. Since n is finite we observe variance in S_n as well. That is,

$$\text{Var}(S_n) = \frac{\mu^4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}, \quad \mu^4 = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right]$$

Similar for the covariance

Kan ik nu
de sample
momenten
nemen?

Wishart o
https://st
of-a-
sample-
covariance
for-norma
variables?

Example

Suppose we have as dependent variable the annual income of a group of farmers. f is defined as $f(A,B,C) = A+2B-3C$ with $A \sim \mathcal{N}\left(\mu_A, \sigma_A^2\right)$ the income of product A, $B \sim \mathcal{N}\left(\mu_B, \sigma_B^2\right)$ the income of product B which is doubled because of government subsidies, and $C \sim \mathcal{N}\left(\mu_C, \sigma_C^2\right)$ the net investments not depending on A and B. The correlation between products X and Y is denoted by $\rho_{XY} \in [-1,1]$ for $X \in \{A,B,C\}$ and $Y \in \{A,B,C\}$. For A, B, and C the variances are $(\sigma_A^2, \sigma_B^2, \sigma_C^2) = (3,2,1)$ and the correlations are $(\rho_{AB}, \rho_{AC}, \rho_{BC}) = (0.5, -0.2, 0)$. Then, for $k \in \{A,B,C\}$,

$$\text{Var}(f) = a_A^2\sigma_A^2 + a_B^2\sigma_B^2 + a_C^2\sigma_C^2 + 2a_Aa_B\rho_{AB}\sigma_A\sigma_B + 2a_Aa_B\rho_{AC}\sigma_A\sigma_C + 2a_Ba_C\rho_{BC}\sigma_B\sigma_C = 26.9774$$

with $(a_A, a_B, a_C) = (1, 2, -3)$. The variance of f decomposed into the variances of A, B, and C is given by

$$v_A = 0.2405, \quad v_B = 0.3873, \quad v_C = 0.3721$$

Indeed, as required, $v_A + v_B + v_C = 1$. Although C has a larger variance contribution to $\text{Var}(f)$ then B ($a_C^2\sigma_C^2 = 9$ vs. $a_B^2\sigma_B^2 = 8$), its overall share is smaller as A and C are negatively correlated while A and B are positively correlated. The consequence is that $a_Aa_B\rho_{AC}\sigma_A\sigma_C$ is negative implying a smaller v_C while $a_Aa_B\rho_{AB}\sigma_A\sigma_B$ is positive implying a larger v_B. When such effects need to be analysed as well, we need the decomposition of v_k into v_k^{Var} and v_k^{Cov}:

$$v_A^{\text{Var}} = 0.1112,\ v_A^{\text{Cov}} = 0.1293,\ v_B^{\text{Var}} = 0.2965,\ v_B^{\text{Cov}} = 0.0908,\ v_C^{\text{Var}} = 0.3336,\ v_C^{\text{Cov}} = 0.0385$$

Note that μ_A, μ_B, and μ_C do not play a role in this variance decomposition.

2.3.1 (Experimental) Alternative decomposition: derivative

Let

$$\tilde{v}_k = \frac{\partial \text{Var}(f(X_1, \ldots, X_m))}{\partial \text{Var}(X_k)}$$

Example

Suppose we have as dependent variable the annual income of farmers. f is defined as $f(A,B) = A+B$ with $A \sim \mathcal{N}\left(\mu_A, \sigma_A^2\right)$ the fixed income and $B \sim \mathcal{N}\left(\mu_B, \sigma_B^2\right)$ the variable income. The correlation between A and B is $\rho \in [-1,1]$. Then

$$\text{Var}(f) = \sigma_A^2 + \sigma_B^2 + 2\rho\sigma_A\sigma_B$$

See Figure 5 for a visualisation of the income and its decomposition into revenue and costs for a certain choice of $\sigma_A^2 = 2$, $\sigma_B^2 = 1$, and $\rho = 0$. Note that μ_A and μ_B do not play a role in the variance decomposition. The variance of f decomposed into the variances of A (σ_A^2) and B (σ_B^2) are given by

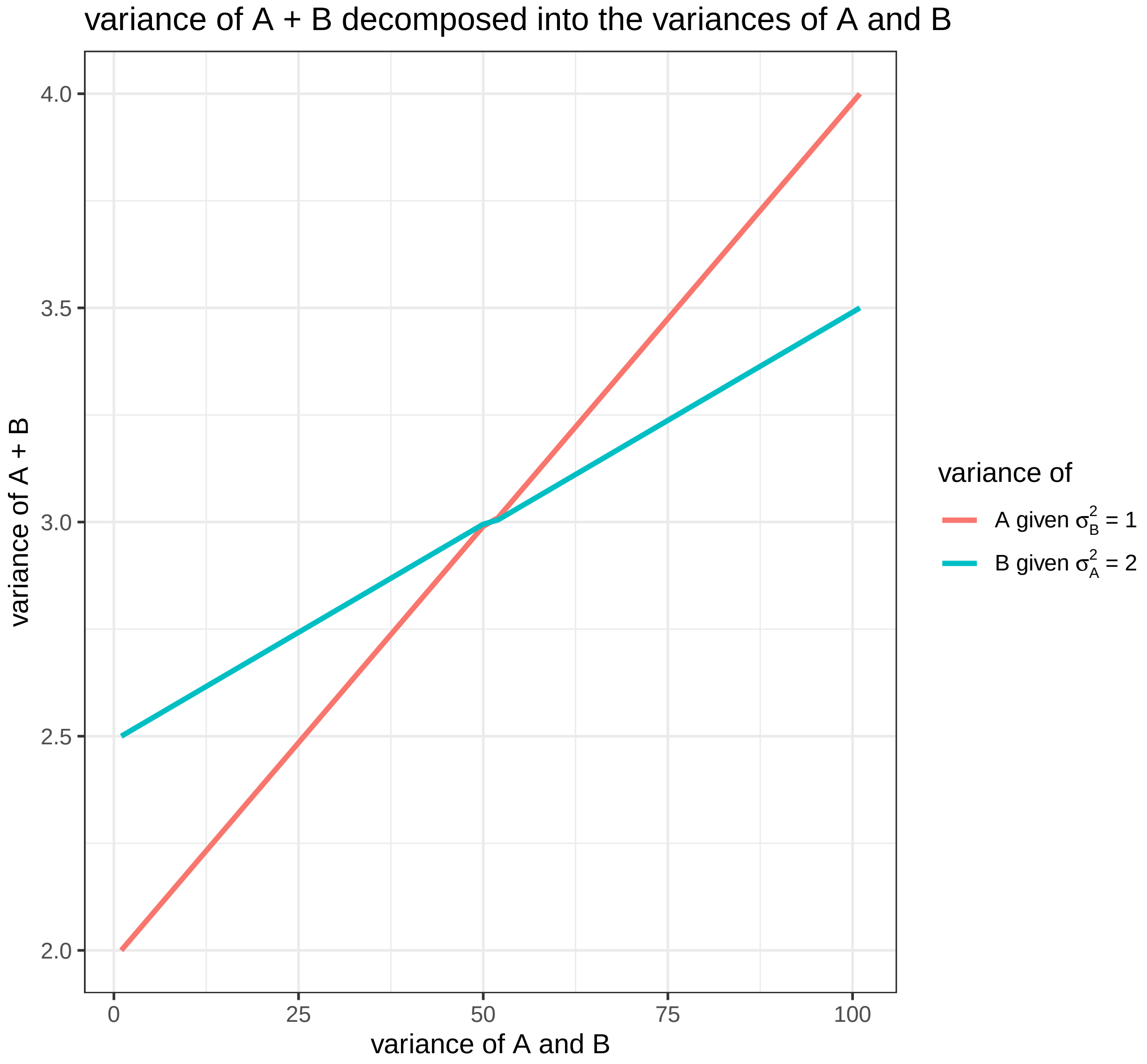
$$v_1 = v_A = \frac{\left|1 + \rho\frac{\sigma_B}{\sigma_A}\right|}{\left|1 + \rho\frac{\sigma_B}{\sigma_A}\right| + \left|1 + \rho\frac{\sigma_A}{\sigma_B}\right|} \quad \text{and} \quad v_2 = v_B = \frac{\left|1 + \rho\frac{\sigma_A}{\sigma_B}\right|}{\left|1 + \rho\frac{\sigma_B}{\sigma_A}\right| + \left|1 + \rho\frac{\sigma_A}{\sigma_B}\right|}$$

The justification for taking the derivative is the following.

Let $g(\sigma_X^2, \sigma_Y^2)$ be the variance of the function $f(X,Y) = X+Y$ with variances of X and Y denoted by,

Probleem met $A +$ want met $\rho = 0$ is d afgeleide 1 en dus gelijk voo A en B

Figure 5: Visualisation of the breakdown of the variance of the income into the variance of the fixed income (A) and the variance of the variable income (B).



respectively, σ_X^2 and σ_Y^2. Now

$$\mathrm{Var}(f) = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y, \quad \rho \in [-1, 1]$$

For $\sigma_X = \sigma_Y$ the variance of f equals $2\sigma_X^2(1+\rho)$ and the contribution to the $\mathrm{Var}(f)$ of X and Y is the same:

$$v_X = \frac{\tilde{v}_X}{\tilde{v}_X + \tilde{v}_Y} = \frac{\tilde{v}_X}{\tilde{v}_X + \tilde{v}_X} = \frac{1}{2} = v_Y$$

For $\rho = 0$ the variance of f equals $\sigma_X^2 + \sigma_Y^2$ and the contribution to the $\mathrm{Var}(f)$ of X and Y results in

$$v_X = \frac{\sigma_X}{\sigma_X + \sigma_Y} \quad \text{and} \quad v_Y = \frac{\sigma_Y}{\sigma_X + \sigma_Y}$$

For $\sigma_X \neq \sigma_Y$ and $\rho \neq 0$ we proceed as follows. The Taylor approximation of $\mathrm{Var}(f)$ in *only* σ_X^2 around $\sigma_X^2 = \sigma_0^2$, for $\sigma_0^2 > 0$, is given by

$$\tilde{v}_X = g\left(\sigma_X^2, \sigma_Y^2\right) = \sigma_0^2 + \sigma_Y^2 + 2\rho\sigma_0\sigma_Y + \left(\sigma_X^2 - \sigma_0^2\right)\left(1 + \rho\frac{\sigma_Y}{\sigma_0}\right) + \frac{\left(\sigma_X^2 - \sigma_0^2\right)^2}{2}\left(1 - \frac{\rho}{2}\frac{\sigma_Y}{\sigma_0^{3/2}}\right) + \ldots$$

Written out this yields

$$\tilde{v}_X = g\left(\sigma_X^2, \sigma_Y^2\right) = \sigma_0^2 + \sigma_Y^2 + 2\rho\sigma_0\sigma_Y + \sum_{j=1}^{k}\frac{\left(\sigma_X^2 - \sigma_0^2\right)^j}{j!}\left((-)^{j+1}\alpha_j\frac{\sigma_Y}{\sigma_0^{\beta_j}} + 1\right)$$

with, for $j = 1, \ldots, k$, $\beta_0 = 0$, $\beta_j = \beta_{j-1} + 1$, $\alpha_0 = 1$, $\alpha_j = \beta_{j-1}\alpha_{j-1}$. The restriction $\sigma_0^2 > 0$ is required because g needs to be differentiable at σ_0^2, which is only the case for $\sigma_0^2 > 0$ as g is not defined on the real line for $\sigma_0^2 < 0$.

For $\sigma_0^2 = \sigma_Y^2$ we have

$$\mathrm{Var}(f) = g\left(\sigma_X^2, \sigma_Y^2\right) = 2\sigma_Y^2(1+\rho) + \left(\sigma_X^2 - \sigma_Y^2\right)\left(1 + \rho\frac{\sigma_X}{\sigma_Y}\right) + \frac{\left(\sigma_X^2 - \sigma_Y^2\right)^2}{2}\left(1 - \frac{\rho}{2}\frac{\sigma_X}{\sigma_Y^2}\right) + \ldots$$

3 Automated Breakdown

Given the analyses results of Section 2.1, Section 2.2, and Section 2.3, we can automatically suggest the next (underlying) variable to be analysed. More formally, for a dependent variable $\boldsymbol{y}$ and underlying variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ we have the impact ∇_{tj} starting from the second period, the sensitivity d_{tj} starting from the first period, and variance decomposition v_{tk} starting from the first period.

	d_{tj} is low	d_{tj} is high
∇_{tj} is low		
∇_{tj} is high		

4 Case Study

5 Implementation

6 Conclusion and Recommendations