KU LEUVEN

FACULTY OF SCIENCE

# The Variability of the Belgian Business Survey Indicator and its Predictive Power

**Fabrice VAN BOECKEL**

Co-Supervisor: Prof. G. Molenberghs
KU Leuven

Co-Supervisor: L. Van Belle
National Bank of Belgium

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Statistics

Academic year 2018-2019

# Acknowledgement

Thanks my family and friends - the National Bank of Belgium Laurent, ... -

*"Statistics are the heart of democracy."*
- Simeon Strunsky

# Abstract

This Master Thesis explores the Variance of the Belgian Business Survey. Several finding concerning the nature and properties of the Variance are found as the bounds and relation with the mean.

In a second part, the predictive power of the variance is examined and it's found that ...

It's also the first time that à Markov Switching model is used in this context. It was showed that ...

## Keywords

Business Surveys - Business Barometer - Survey Variance - Markov Switching -

# Contents

# CHAPTER 1

# Introduction

Lemasson (2017)

Business Survey Indicator / Business Barometer / Business Confidence Indicator

A widespread method to predict the evolution of National Economies is the survey-based Business indicator. Belgium have been collecting this indicator for more than 60 years. This long evolution

- Talk about tradition of improving BSB

This Thesis is included in the continuity of a long tradition of papers proposing improvement and ways to add value to the Business Barometer (.......) will propose ways to add information to the Belgian Business Barometer, that could also be applied to others

Since 1968, the National Bank of Belgium publishes each month the national

# CHAPTER 2

# The Business Survey Indicator

This first chapter is a more general description of the Belgian Business Survey Indicator, that we will also call the Business Barometer. We will present it different calculations, the weighting that are applied and

explain the two types of weightings

## 2.1 History

This year, the Belgian Business Survey is celebrating its 65th anniversary.
In 1954, the National Bank of Belgium started this indicator, that since then

- 1972 results are synthesised in an indicator; the business survey indicator

- Wall Street Journal article "Euroland Discovers A Surprise Indicator: Belgian Confidence" (Rhoads, 1999)

- predictive power for the EU Vanhaelen et al. (2000)

changes in 2009 see later (section 2.3)

### 2.1.1 A Business Cycle

## 2.2 Sampling Method

## 2.3 Objective and Methodology

sectionObjective and Methodology In 2009 was published "The National Bank of Belgium's new business survey indicator" (De Greef and Van Nieuwenhuyze)

- only take a limited amount of questions into account, the most relevant ones (3-4 questions)
- inclusion of the services in the global indicator
- lighten smoothing method

**Quality Criterion**

- high correlation with GDP
- fluctuation that's mostly explained by the conjuncture
- predictive power for the futur months

more information can be found in De Greef and Van Nieuwenhuyze (2009)

...

## 2.4 Questionnaire

.... questionnaire can be found in appendix

Questions taken into account for RS975: ....

originally question Q18, 27, 32 and 33, for simplicity numbered here as 1, 2, 3 and 4.

## 2.5 Calculation of the Indicator

This section ...

The calculation of the indicator for a specific question at a specific time can be written as follow;

### 2.5.1 Unweighted Indicator

$$E(X) = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.1}$$

where

$x_i$ is the answer of the respondent i and can each take value -1, 0 and 1

$n$ is the total of respondents

Since x can only take three different values, we can decompose it into

$$E(X) = \frac{\sum_{i=1}^{n} x_{+i} + \sum_{i=1}^{n} x_{Ni} + \sum_{i=1}^{n} x_{-i}}{n} \tag{2.2}$$

where $x_{+i}$, $x_{Ni}$ and $x_{-i}$ are the positive(+), neutral (N) and negative (-) answers of the respondent i

$n$ is the total of respondents

We know that $\sum_{i=1}^{n} x_{Ni} = 0$ so we can write

$$E(X) = \frac{\sum_{i=1}^{n} x_{+i}}{n} + \frac{\sum_{i=1}^{n} x_{-i}}{n} \tag{2.3}$$

$\sum_{i=1}^{n} x_{+i}/n$ is the proportion of positive answers and $\sum_{i=1}^{n} x_{-i}/n$ is the negative proportion of negative answer so for simplicity we write it

$$E(X) = \pi_+ - \pi_- \tag{2.4}$$

where $\pi_+$ and $\pi_-$ are the proportion of respondents answering positive and negative. $\pi$ is use here also in the probabilistic way as it can also be seen as the probability that a respondent answers positive, negative or neutral ($\pi_0$) with $\pi_+ + \pi_0 + \pi_- = 1$.

### 2.5.2 Weighted Indicator

$$E(X) = \frac{\sum_{i=1}^{n} (\omega_i p_i x_i)}{\sum_{i=1}^{n} \omega_i p_i} \tag{2.5}$$

where
$x_i$ is the answer of the respondent i and can each take value -1, 0 and 1
$p_i$ is the weight of the globalisation of the company i
$\omega_i$ is the weight of the company i

**Globalisation procedure**

**Weighting procedure**

### 2.5.3 Properties

E(X) has -1 as lower bound and 1 as upper bound

### 2.5.4 Take different questions into account

The previous calculations where specific to each question. The published indicators are usually taking different survey questions into account. For example the Industry indicator that we will be interested in is composed of four questions that have all the same weight:

$$\text{Industry Business Indicator } = \frac{E(X_{Q18}) + E(X_{Q27}) + E(X_{Q32}) + E(X_{Q33})}{4} \tag{2.6}$$

where $E(X_{Q18})$, $E(X_{Q27})$, $E(X_{Q32})$ and $E(X_{Q33})$ are the different averages for question 18, 27, 32 and 33 (can be weighted or unweighted)

# CHAPTER 3

# Variance of the Indicator

The variance is, with the mean, one of the first tool for Statisticians to study a certain variable.

In the context of the Business Survey, the variance haven't been used much

**difference sampling error and variance**

here variance is a measure of the "dispersion" of the answers.

**difference between nominal and continuous variable variance**

## 3.1 Presentation

The (Cochran, 1977)

$$E(X) = \pi_+ - \pi_- \tag{3.1}$$

As done for the Indicator, two different variance will be take into account here, the weighted and the unweighted variance of the indicator.

### 3.1.1 Variance of the Unweighted Indicator

The main variance

cite

$$
\begin{aligned}
Var(X) &= E\left[(X - E(X))^2\right] \\
&= E\left(X^2\right) - E\left(X\right)^2 \\
&= \left(\frac{\sum_{i=1}^{n} x_{+i}^2}{n}\right) + \left(\frac{\sum_{i=1}^{n} x_{Ni}^2}{n}\right) + \left(\frac{\sum_{i=1}^{n} x_{-i}^2}{n}\right) - E(X)^2 \\
&= \pi_+ + \pi_- - (\pi_+ - \pi_-)^2 \qquad (3.2)
\end{aligned}
$$

Since $\left(\frac{\sum_{i=1}^{n} x_{Ni}^2}{n}\right) = 0$ , $x_{+i}^2 = x_{+i}$, $x_{-i}^2 = x_{-i}$ and $E(X) = \pi_+ - \pi_-$
We then have several different ways to write the previous equation;

$$
\begin{aligned}
Var(X) &= \pi_+ + \pi_- - (\pi_+ - \pi_-)^2 \qquad (3.3) \\
&= \pi_+ + \pi_- - E(X)^2 \qquad (3.4) \\
&= 1 - \pi_n - E(X)^2 \qquad (3.5)
\end{aligned}
$$

Equation 3.3 is interesting Equation 3.4

### 3.1.2 Variance of the Weighted Indicator

$$
Var(X) = \frac{1}{\sum \omega_i p_i} \sum_{i=1}^{N} \omega_i p_i (X_i - \bar{X})^2 \qquad (3.6)
$$

$$
\begin{aligned}
Var(X) &= \pi_+ + \pi_- - (\pi_+ - \pi_-)^2 \qquad (3.7) \\
&= \pi_+ + \pi_- - E(X)^2 \qquad (3.8) \\
&= 1 - \pi_0 - E(X)^2 \qquad (3.9)
\end{aligned}
$$

### 3.1.3 Properties

**Property 1: The variance of X is bounded between -1 and 1**

**Property 2: The variance = A5 and E(X)**

**Property 3: .....**

## 3.2 Take different questions into account

## 3.3 Discussion regarding the 'true variance'

There is another way to calculate the variance that have been ignored for, that is calculating the variance for each lowest group of globalisation, and then combine those calculated variances.

Interestingly, it have been calculated for several Questions of the business barometer, and it is approximately 10 times smaller than the variance based on all the answer a ones.

The reasons why it will not be used here - losing information - weight of globalisation taken into account in the weighted variance

CHAPTER 4

# Indicator of the Evolution of Individual Responses

Best name for this indicator

1. Z indicator
2. Indicator of the Changes in individual answers between t-1 and t

brainstorm evolution volatility change

An issue for this indicator was to find an optimal name for it so that it would be easily understand by the largest number.

| Notation | $x_{t-1}$ | $x_t$ | $z_t$ |
|---|---|---|---|
| $\pi_{--}$ | -1 | -1 | 0 |
| $\pi_{-0}$ | -1 | 0 | 1 |
| $\pi_{-+}$ | -1 | 1 | 2 |
| $\pi_{0-}$ | 0 | -1 | -1 |
| $\pi_{00}$ | 0 | 0 | 0 |
| $\pi_{0+}$ | 0 | 1 | 1 |
| $\pi_{+-}$ | 1 | -1 | -2 |
| $\pi_{+0}$ | 1 | 0 | -1 |
| $\pi_{++}$ | 1 | 1 | 0 |

| | | $t$ | | |
|---|---|---|---|---|
| | | **-** | **0** | **+** |
| $t-1$ | **-** | $\pi_{--}$ | $\pi_{-0}$ | $\pi_{-+}$ |
| | **0** | $\pi_{0-}$ | $\pi_{00}$ | $\pi_{0+}$ |
| | **+** | $\pi_{+-}$ | $\pi_{+0}$ | $\pi_{++}$ |

The Indicator of the evolution of the individual responses can be obtained by

$$E(Z) = \pi_{0+} + \pi_{-0} - \pi_{+0} - \pi_{0-} + 2\pi_{-+} - 2\pi_{+-} \tag{4.1}$$

where
$\pi$ is the proportion/probability of respondent answering (-,0,+) at $t-1$ and (-,0,+) at time $t$

$$E(Z) = \begin{array}{c|ccc} & \textbf{-} & \textbf{0} & \textbf{+} \\ \hline \textbf{-} & 0 & +1 & +2 \\ \textbf{0} & -1 & 0 & +1 \\ \textbf{+} & -2 & -1 & 0 \end{array} \tag{4.2}$$

$$\pi_{++} + \pi_{+0} + \pi_{+-} + \pi_{0+} + \pi_{00} + \pi_{0-} + \pi_{-+} + \pi_{-0} + \pi_{--} = 1 \tag{4.3}$$

# CHAPTER 5

# Variance of the Evolution of Individual Responses / Volatility of Responses

That we will also call the **volatility of the indicator**, in the sens that the variance of the evolution of the indicator account for the dispersion of the difference in answers over a two times period.

In this case, the highers variance of Z, will be obtained when half of the companies went from a negative answer to a positive answer and the other half did the opposite and changed from a positive answer at t-1 to a negative answer at t.

waza see Chapter 5

The idea is that this variance of Z is complementary to the estimation of Z since they have two very interesting but different interpretations. Further interpretation will be

## 5.1 Presentation

$$\pi_{++} + \pi_{+0} + \pi_{+-} + \pi_{0+} + \pi_{00} + \pi_{0-} + \pi_{-+} + \pi_{-0} + \pi_{--} = 1 \tag{5.1}$$

$$\begin{aligned} Var(Z) &= \pi_{0+} + \pi_{-0} + \pi_{+0} + \pi_{0-} + 4\pi_{-+} + 4\pi_{+-} \\ &\quad -(\pi_{0+} + \pi_{-0} - \pi_{+0} - \pi_{0-} + 2\pi_{-+} - 2\pi_{+-})^2 \\ &= \pi_{0+} + \pi_{-0} + \pi_{+0} + \pi_{0-} + 4\pi_{-+} + 4\pi_{+-} - E(Z)^2 \\ &= 1 - \pi_{++} - \pi_{00} - \pi_{--} + 3\pi_{+-} + 3\pi_{-+} - E(Z)^2 \end{aligned} \tag{5.2}$$

$$
\begin{aligned}
Var(Z) &= \left(\begin{array}{c|rrr} & \textbf{-} & \textbf{0} & \mathbf{+} \\ \hline \textbf{-} & 0 & +1 & +4 \\ \textbf{0} & +1 & 0 & +1 \\ \mathbf{+} & +4 & +1 & 0 \end{array}\right) - \left(\begin{array}{c|rrr} & \textbf{-} & \textbf{0} & \mathbf{+} \\ \hline \textbf{-} & 0 & +1 & +2 \\ \textbf{0} & -1 & 0 & +1 \\ \mathbf{+} & -2 & -1 & 0 \end{array}\right)^2 \\
&= \left(\begin{array}{c|rrr} & \textbf{-} & \textbf{0} & \mathbf{+} \\ \hline \textbf{-} & 0 & +1 & +4 \\ \textbf{0} & +1 & 0 & +1 \\ \mathbf{+} & +4 & +1 & 0 \end{array}\right) - (E(Z))^2 \\
&= 1 + \left(\begin{array}{c|rrr} & \textbf{-} & \textbf{0} & \mathbf{+} \\ \hline \textbf{-} & -1 & 0 & +3 \\ \textbf{0} & 0 & -1 & 0 \\ \mathbf{+} & +3 & 0 & -1 \end{array}\right) - (E(Z))^2
\end{aligned}
$$

### 5.1.1 Properties

**Property 1: the variance of Z is bounded between -1 and 1**

**Property 2:**

# CHAPTER 6

# Seasonal Effects

The National Bank, before publishing the Business Survey Indicator, applies a X11 seasonal correction

The literature about seasonal effects is very rich and variate

- NBB developed JDemetra+ and has since been recommended by the ECB and Eurostat for all NSI in Europe.

at the same time the department of Business Survey uses as a X11 adapted method to correct for seasonality because don't want to correct for previous publications.

Methodology
- test for seasonality
- run the analysis without corrections
- apply corrections and see if more accurate

## 6.1 JDemetra+

## 6.2 X11

### 6.2.1 Seasonal correction of the Indicator

### 6.2.2 Seasonal correction of the Variance

### 6.2.3 Seasonal correction of the Indicator of the Evolution

### 6.2.4 Seasonal correction of the Variance of Z

### 6.2.5 Seasonal correction of the Proportions

## 6.3 Limitations

explain the issue of seasonal correction on "future data"

# CHAPTER 7

# Non-Response, Dropout and Attrition

Aside of Seasonal effect, there are three main biases that could arise in the context of the BSI; non-reponse, dropout and attrition
- cor(time, var) = 0.5

## 7.1 Non-Response

## 7.2 Dropout

Non parametric test Das et al. (2011)

## 7.3 Attrition

Attrition / Panel Conditioning Master Thesis done about the Belgian Labor Force Survey, where attrition was studied Priyana Hardjawidjaksana (2019)

Non parametric test Das et al. (2011)

limitation: only some periods of

CHAPTER 8

# Exploratory Analysis

## 8.1 Small vs Large

## 8.2 By Sector

## 8.3 New vs Old participants

## 8.4 Correlation between questions

## 8.5 Correlation with GDP

Belgian industry claims 25% of the labour force in Belgium and have been shown as been the best indicator to predict citeAlain Quartier and Isabelle

### 8.5.1 lag testing

CHAPTER 9

# Linear (Auto-Regressive) Models

lag

## 9.1 Models

### 9.1.1 Month vs Quarterly data

Error to aggregate everything to quaterly -¿ lost of information

Linear Model

$$GDP_t = \mu + \sum_{i=1}^{n} \sum_{j=0}^{q} \beta_{1,j} X_{i,t-j} + \epsilon_t \tag{9.1}$$

Auto-Regressive model

$$GDP_t = \mu + \sum_{j=1}^{p} \phi_j GDP_{t-j} + \sum_{i=1}^{n} \sum_{j=0}^{q} \beta_{1,j} X_{i,t-j} + \epsilon_t \tag{9.2}$$

where

| | |
|---|---|
| $GDP_t$ | GDP growth over the last semester |
| $X_{i,t}$ | monthly predictors |
| $\mu$ | constant |
| $\phi_j$ | auto-regressive coefficients |
| $\beta_{i,j}$ | regression coefficients |

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 4.2653 | 0.2138 | 19.95 | 0.0000 |
| GDP_year_lag1 | 6.1137 | 3.0454 | 2.01 | 0.0470 |
| E_2 | 7.9182 | 4.0531 | 1.95 | 0.0531 |
| E_2_lag1 | -3.6618 | 5.6642 | -0.65 | 0.5192 |

## 9.2 Model

## 9.3 Evaluation

### 9.3.1 R-square

### 9.3.2 Mean Square Prediction Error

### 9.3.3 Diebold-Mariano Test

### 9.3.4 Out-of-Sample performances

### 9.3.5 blabla see slides

# CHAPTER 10

## Markov Switching Models

**Small Introduction + why are we using it**

Since the pioneer work by Hamilton (1989), Markov Switching models have been largely used to model business cycles.

Markov Switching models have been very popular since Hamilton (1989) to model business cycles and predict Turning points (see Duprey and Klaus (2017), ...).

Able to predict the 2008 financial crisis if used MS-VAR model Gadea Rivas and Perez-Quiros (2015)

## 10.1 Model(s) Specification

### 10.1.1 Notation

| | |
|---|---|
| $S_t = \{0, 1\}$ | states |
| $N = 2$ | number of states (2) |
| $T = 372$ | number of observations |
| $x_{t=1...T}$ | (hidden) state at time t |
| $y_{t=1...T}$ | Change of the Industrial production indices at time t |
| $p_{i=1...n,j=1...n}$ | probability of transition from state $i$ to state $j$ |
| $F(y\|\theta)$ | probability distribution of an observation, parametrized on $\theta$ |

### 10.1.2 Model

**Model**

$$y_t = \begin{cases} \mu_0 + \sum_{j=1}^{p} \phi_j GDP_{t-j} + \epsilon_t & \text{if } S_t = 0 \\ \mu_1 + \sum_{j=1}^{p} \phi_j GDP_{t-j} + \epsilon_t & \text{if } S_t = 1 \end{cases} \tag{10.1}$$

where $\epsilon$ is $N(0, \sigma_s)$

| | |
|---|---|
| $\mu_s = \beta_0 = c_s = \alpha_s$ | regime-specific mean |
| $\beta_s = \phi_s$ | regime-specific auto-regressive parameter |
| $\sigma_s$ | regime-specific variance |

**Transition equation/probability**

$$P = P(S_t = s_t \mid S_{t-1} = s_{t-1}) = \begin{bmatrix} 1 - p_t & p_t \\ q_t & 1 - q_t \end{bmatrix} \tag{10.2}$$

We have then,

$P(S_t = 1 \mid S_{t-1} = 1) = p_t$
$P(S_t = 0 \mid S_{t-1} = 1) = 1 - p_t$
$P(S_t = 0 \mid S_{t-1} = 0) = q_t$
$P(S_t = 1 \mid S_{t-1} = 0) = 1 - q_t$

or

$$P = P(S_t = s_t \mid S_{t-1} = s_{t-1}) = \begin{bmatrix} 1 - p_t & p_t\text{..............} \\ q_t\text{............} & 1 - q_t \end{bmatrix} \tag{10.3}$$

# CHAPTER 11

# Conclusion

# CHAPTER 12

# Discussion

## Recruitment procedure and panel data

not real sampling theory

## Z that takes more periods into account

## Limitations

## Further Research

More complex Nowcasting model with Space space models / MIDAS ....

Combine mixed models and Markov Chain for Panel Data (de Haan-Rietdijk et al., 2017)

State Space Model

# Bibliography

Das, M., Toepoel, V., and van Soest, A. (2011). Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods & Research*, 40(1):32–56.

De Greef, I. and Van Nieuwenhuyze, C. (2009). The National Bank of Belgium's new business survey indicator.

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L. B., Allen, N. B., and Hamaker, E. L. (2017). On the Use of Mixed Markov Models for Intensive Longitudinal Data. *Multivariate Behavioral Research*, 52(6):747–767.

Duprey, T. and Klaus, B. (2017). How to predict financial stress? An assessment of Markov switching models. Technical report.

Gadea Rivas, M. D. and Perez-Quiros, G. (2015). The Failure to Predict the Great Recession - A view through the role of credit. *Journal of the European Economic Association*, 13(3):534–559.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357.

Lemasson, C. (2017). *Enquête de conjoncture dans le secteur industriel belge : évaluation de stratégies d'échantillonnage*. PhD thesis.

Priyana Hardjawidjaksana, I. (2019). *Investigating Attrition in Belgian Labor Force Survey*. PhD thesis, KU Leuven.

Rhoads, C. (1999). Euroland Discovers A Surprise Indicator: Belgian Confidence. *Wall Street Journal, Europe; Brussels*, page 1.

Vanhaelen, J.-J., Dresse, L., and De Mulder, J. (2000). The Belgian Industrial Confidence Indicator: Leading Indicator of Economic Activity in the Euro Area? *SSRN Electronic Journal*.

# Appendix

# List of Abbreviations

BSI Business Survey Indicator
GDP

# List of Figures

# List of Tables

# Code

## R code for Seasonal Adjustment

## R code for Creating Lags

## R code for Linear (Auto-Regressive) Models

## R code for Markov Switching Models

**AFDELING**
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be