Acta Univ. Sapientiae, Economics and Business, **10** (2022) 199–227

DOI 10.2478/auseb-2022-0012

Topic Classification of Central Bank Monetary Policy Statements: Evidence from Latent Dirichlet Allocation in Lesotho

Moeti DAMANE
Central Bank of Lesotho, P.O. Box 1184, Maseru 100, Lesotho
e-mail: damane.moeti@gmail.com

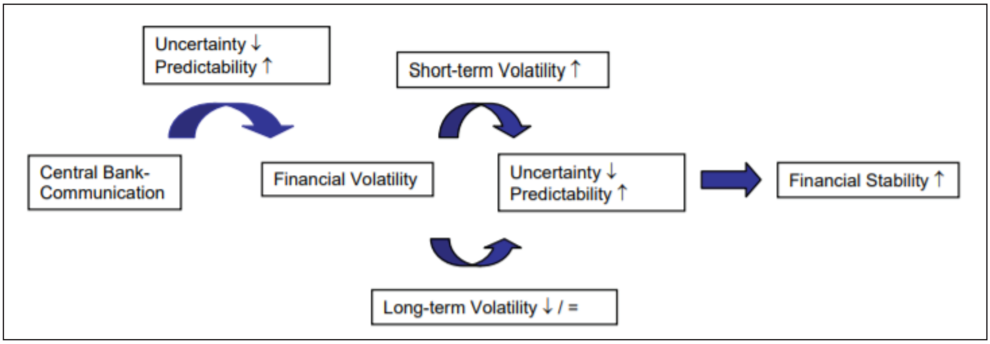
Abstract: This article develops a baseline on how to analyse the statements of monetary policy from Lesotho's Central Bank using a method of topic classification that utilizes a machine learning algorithm known as Latent Dirichlet Allocation. To evaluate the changes in the policy distribution, the classification of topics is performed on a sample of policy statements spanning from February 2017 to January 2021. The three-topic Latent Dirichlet Allocation model extracted topics that remained prominent throughout the sample period and were most closely reflective of the functions of the Central Bank of Lesotho Monetary Policy Committee. The topics identified are: (i) International Monetary and Financial Market Conditions; (ii) Monetary Policy Committee and International Reserves; (iii) Regional and International Economic Policy Conditions. The three-topic Latent Dirichlet Allocation model was determined as the most appropriate model through which a consistent analysis of topic evolution in Central Bank of Lesotho Monetary Policy Statements can be performed.

Keywords: monetary policy statement, topic modelling, central bank, Lesotho, Latent Dirichlet Allocation

JEL Classification: E5, E52, E58

1. Introduction

The primary mandate of the Central Bank of Lesotho (CBL) is to achieve and maintain price stability in the country. This is achieved through the one-to-one peg between the South African Rand and Lesotho's currency, the Loti (Damane, 2019). The CBL communicates its policy stance through its Monetary Policy Statement, a policy document released at the end of every meeting of the Monetary Policy Committee (MPC). The advantages of a clear policy communication from institutions such as central banks cannot be overstated. Central Bank Monetary Policy Statements shape expectations about future policy actions. Although the statements usually carry quantitative measures and indicators, they also include detailed qualitative descriptions of current economic conditions and outlook, which can affect current and future financial market sentiment and conditions (Knütter et al., 2011; Vílchez-Román et al., 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021). *Figure 1* is an illustration of how central bank communication might affect financial stability.[1]



Source: Knütter et al., 2011

Figure 1. *The role of central bank communication in financial stability*

From *Figure 1*, clear and transparent communication of policy by central banks helps to maintain financial stability through the provision of clear, relevant, and credible information. When market participants better understand policies of monetary and financial stabilization, as communicated by the central bank (through its policy statements), this might reduce financial sector uncertainty and volatility and create a conducive environment for safe and stable financial activity. Although short-term volatility is possible, volatility in the long-term is muted, given clear and predictable central bank communication. In this case, asset prices will be aligned with overall monetary policy objectives (Knütter et al., 2011; Vílchez-Román et al., 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021).

There is currently no baseline way of how to analyse CBL MPC statements and concisely characterize the thematic or topical trends of the MPC's discussions over time. This is primarily due to the unstructured[2] nature of the text data.

1 Financial stability exists when there is confidence in the general functioning of key financial institutions and markets in the economy, such that real or financial asset price movements do not undermine the economy's monetary stability and levels of employment (Foot, 2003 – in Knütter et al., 2011).

2 Unstructured data is data often contained in business documents, reports, news articles, etc. It is free-form text data that cannot be immediately obtained or extracted from an electronic file or relational database with well-defined (structured) rows (records) and columns (fields) ("structured data", n. d.).

Recent years have seen marked advances in artificial intelligence (AI),[3] big-data analytics, machine learning,[4] and the digitalization and digitization of global economies and their data. These technological strides have made it possible for economic and financial data analysts to better infer policy guidance from qualitative information contained in policy documents such as Monetary Policy Statements. Using techniques such as natural language processing and topic modelling, analysts can extract "latent" or hidden patterns and relationships between text data, which would otherwise be a challenge to isolate and identify. This enables the assessment of the tone[5] and/or sentiment and discussion trends in Monetary Policy Statements (Edison and Carcel, 2021; Doh et al., 2021). Topic models are Bayesian statistical models that structure large quantities of discrete and unstructured textual data, known as "corpus" (from a document set), in terms of latent themes usually referred to as "topics". They facilitate analysis of how discussions around particular topics evolve and change through time (Blei et al., 2003; Shirota et al., 2015; Zhao et al., 2015; Schwarz, 2018; Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Reisenbichler and Reutterer, 2019; Rabindranath, 2020; Buenaño-Fernandez et al., 2020; Edison and Carcel, 2021; Doh et al., 2021).

This paper's objective is to analyse the monetary policy statements of Lesotho's Central Bank from February 2017 to January 2021 and to identify the evolutionary trends of key themes or topics discussed by members of the CBL MPC over the sample period using the Latent Dirichlet Allocation (LDA) topic modelling technique. LDA is an unsupervised machine learning algorithm developed by Blei and Jordan (2003) and Blei et al. (2003). It is the most widely used probabilistic method for modelling large quantities of corpus. The model's main assumption is that a series of topics can be derived from a combination of textual documents using all the documents' words (i.e. document vocabulary). The model's purpose is to use machine learning to uncover the latent/hidden (unobserved) topics in the corpus and how they are exhibited in each document. Topic development leverages a fixed vocabulary distribution of words that co-occur in a recurring pattern. In this way, a probability distribution over a set of topics can represent each document, while the distribution of probability over words in the corpus[6]

3 AI is defined by Merriam-Webster as "a branch of computer science dealing with the simulation of intelligence behaviour in computers [and] the capability of a machine to imitate intelligent human behaviour" ("Artificial Intelligence", n. d.).

4 Machine learning is a component of AI that designs a sequence of actions known as algorithms for purposes of solving a problem or identifying patterns in datasets that are often large and complex. The algorithms automatically optimize through experience and require little to no human intervention to do so – i.e. supervised learning, reinforcement learning, unsupervised learning (Financial Stability Board, 2017).

5 Despite its quantitative attributes, the tone of the monetary policy statement can be assessed to be either pessimistic or optimistic (Doh et al., 2021).

6 A policy statement could exhibit multiple topics and not fit just neatly into one, while all policy statements taken separately could each show varying proportions of the same topics (Reisenbichler and Reutterer, 2019).

can represent each topic. The LDA model is therefore from the family of mixed membership models. However, unlike classical clustering methods with binary variable membership, LDA allows every word to partially occupy all topics with varying probabilities. A vector of continuous non-negative latent variables that adds up to 1 represents the word membership across topics. Similarly, all topics extracted from the corpus are partial members of all documents with different probabilities (Blei et al., 2003; Shirota et al., 2015; Zhao et al., 2015; Schwarz, 2018; Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Reisenbichler and Reutterer, 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021).

To our best knowledge, this the first study of its kind in Lesotho. Most studies that employ topic classification to analyse recent discussion trends in central bank policy documents (i.e. Monetary Policy Statements and/or Minutes) are mostly found in developed countries (see: Shirota et al., 2015; Schwarz, 2018; Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021). This study adds to the monetary policy communication literature, especially in the developing world, by offering readers of the CBL Monetary Policy Statements a baseline of how to analyse the statements and to gather key information from their topical trends. Central bank communication that is clear and well understood by all relevant stakeholders can benefit policymakers through the pooling of knowledge and the management of expectations by market participants (Buenaño-Fernandez et al., 2020).

The remainder of the paper is organized such that an overview of the CBL Monetary Policy Committee with a focus on the monetary policy statements is offered in Section 2. Subsequently, Section 3 reviews the literature. An outline of the data and methodology is provided in Section 4. Section 5 presents the results and analysis. Section 6 concludes and gives policy recommendations, and Section 7 provides the areas for further study.

2. CBL Monetary Policy Committee and Statements

The CBL MPC's principal goal is to formulate and monitor the implementation of monetary policy towards price stability as outlined in sections 5, 6(c), and 6(d) of the CBL Act, 2000. According to CBL (2020), the functions of the MPC are four-fold, namely: (i) reviewing and formulating appropriate monetary policy responses to achieve and maintain price stability; (ii) reviewing the likely impact of international and domestic economic developments on the Bank's ability to achieve and maintain price stability (iii); to ensure that the Bank has adequate levels of foreign reserves required to maintain the one-to-one peg of the loti/rand exchange rate, through a decision on the targeted floor of Net International Reserves; (iv) regular review of the Bank's monetary policy framework to inform the adoption of changes as and when necessary. Communication of MPC decisions is made through a Monetary Policy Statement that is placed on the CBL's website and sent to members of the press through the Corporate Communications Office. The frequency of MPC meetings is bi-monthly. This implies that, ordinarily, a total of six MPC Monetary Policy Statements can be expected in any one year.

3. Review of Literature

An overview of the common assumptions and relative costs associated with the different types of text classification methods is provided in this section. The section also discusses a handful of recent studies that use LDA and other topic modelling techniques to analyse topic trends in economic and financial publications and central bank policy documents (i.e. Monetary Policy Statements and/or Minutes, speeches, etc.).

Overview of Text Classification Methods

There are generally five ways in which discrete text data can be classified. These comprise topic modelling, supervised learning, dictionaries, human coding, and reading. *Table 1* presents the assumptions and costs that are associated with each of these methods. Asmussen and Møller (2019) explain that a comparison of the associated costs and assumptions across methods can add value to the decision of which method to use given the task at hand.

Table 1. *Discrete text categorization methods – common assumptions and relative costs*

	Method				
A. Assumptions	***Reading***	***Human Coding***	***Dictionaries***	***Supervised Learning***	***Topic Model***
Categories known	No	Yes	Yes	Yes	No
Category nesting is known	No	Yes	Yes	Yes	No
Features of relevant text are known	No	No	Yes	Yes	Yes
Mapping known	No	No	Yes	No	No
Ability to automate coding	No	No	Yes	Yes	Yes

B. Costs	*Reading*	*Human Coding*	*Dictionaries*	*Supervised Learning*	*Topic Model*
Pre-analysis Costs					
Conceptualization time – in person hours	Low	High	High	High	Low
Substantive knowledge requirement	Moderate/ High	High	High	High	Low
Costs of Analysis					
Time spent per text – in person hours	High	High	Low	Low	Low
Substantive knowledge requirement	Moderate/ High	Moderate	Low	Low	Low
Post-analysis Costs					
Time spent interpreting – in person hours	High	Low	Low	Low	Moderate
Substantive knowledge requirement	High	High	High	High	High

Source: Quinn et al. (2010)

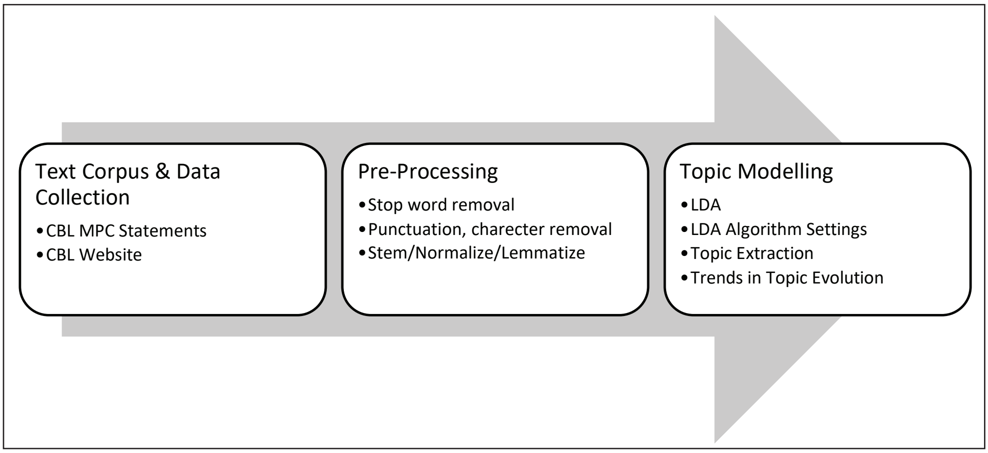
From *Table 1*, the topic modelling approach is relatively much more convenient compared to the rest of the other methods. It does not require the prior knowledge of categories for classification. It can also be automated (reducing person hours), and it offers low costs across the analytic process.

Empirical Literature

A detailed summary of each study is presented in *Appendix A1*. Unsurprisingly, given the dearth of these kinds of studies in the developing world, the empirical evidence reviewed is from developed countries. The studies focus on the topic classification of economic policy documents (i.e. Monetary Policy Statements and/ or Minutes) (see: Edison and Carcel, 2021; Doh et al., 2021; Vílchez-Román et al., 2019; Shirota et al., 2015) as well as customer comments that evaluate products or services the customers consume (see: Sperkova, 2018; Lee et al., 2018). Before the corpus is put through the LDA model, it is pre-processed (i.e. cleared of stop words, normalized, lemmatized, etc.). An assessment of model clustering (topic) output is evaluated at various stages through an iterative process of running the model again but with different Dirichlet priors or a diverse number of topics. In addition, topic selection is critiqued based on authors' discretion and understanding of the subject matter. The choice is also supported by consistency to topic selection with significant co-occurring economic and/or financial developments across the sample period. In all cases, the LDA model can successfully identify evolutionary trends in the topics and thus provide evidence of the importance of the qualitative rationale that usually accompanies the quantitative aspect of policy communication. For instance, if policy documents include information on how policy institutions intend to interpret incoming data, this can have significant impact on the statement's overall tone (i.e. optimistic vs. pessimistic).

4. Data and Methodology

This section outlines the data and methodology used in the study. *Figure 2* provides an illustrative depiction of the research design and how this section's discussion is structured.



Source: author's own illustration

Figure 2. *Diagram depicting research design*

Text Corpus and Data Collection

The study uses publicly available monetary policy statements of the CBL from February 2017 to January 2021 as the corpus (i.e. text input). The Monetary Policy Statements were obtained from the CBL website[7] on 27 February 2021. The study did not employ any inclusion/exclusion criteria. All the available Monetary Policy Statements available on the CBL website at the time of the study were used. The statements were grouped by the year of publication from 2017 to 2021. The total number of Monetary Policy Statements collected at the time of the study was twenty-six, such that in 2017 (n = 6), 2018 (n = 6), 2019 (n = 6), 2020 (n = 7), and 2021 (n = 1). This makes a total of twenty-six .pdf files. The extent of the text corpus and the study timeline were dictated by the number of publicly available monetary policy statements during the time of the study.

7 https://www.centralbank.org.ls/index.php/monetary-policy/mpc-statements.

Pre-Processing

The pre-processing of textual raw data consists in transforming it and readying it for algorithmic use by removing unnecessary words and characters. There are generally four steps in standard text mining pre-processing. These include: (1) lowercasing the corpus and thus preventing a word with differing capitalization from being misconstrued as two different words; (2) removal of typical stop words, such as “a”, “an, “and”, “the”, etc., since they do not have any value to add to the analysis and could impair result accuracy; (3) stemming/normalizing or lemmatizing the text to remove pluralization or other suffixes that interfere with the uniqueness of the word (e.g. “growing” and “growth” are transformed into “grow”); (4) numbers, punctuation characters, and white noise removal to mitigate against any interference with model efficacy (Blei et al., 2003; Shirota et al., 2015; Zhao et al., 2015; Schwarz, 2018; Dwivedi, 2018; Cedervall and Jansson, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Reisenbichler and Reutterer, 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021).

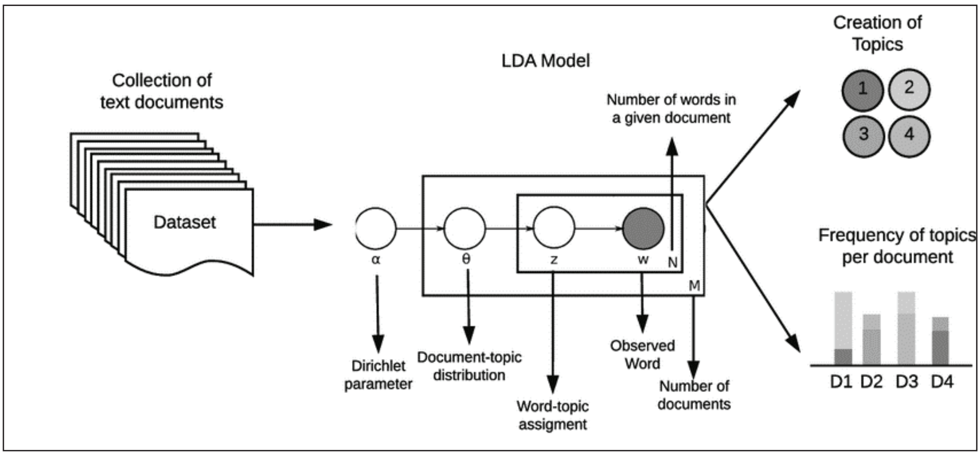
Topic Modelling

The LDA technique can be traced back to the work of Blei and Jordan (2003) and Blei et al. (2003). It is a machine learning algorithm that facilitates the topic modelling of unclassified copious collections of text data. It allows its user to group any number or type of documents into clusters with a similar content, often referred to as “topics”, as defined by the user. The algorithm only leverages text data. Thus, the clustering is independent of the language used in the text or the number of documents being considered. The LDA’s approach to text clustering is probabilistic. The model bases itself on the fact that same-topic discussions usually use words that are similar and often co-occur. The co-occurrence of such words describes a topic as a distribution of probability over words and each document as a distribution of probability over topics. Through LDA, previously unused text data can be made useful for research. Documents and topics can be compared for similarities, and their previously hidden relationships can be unlocked (Popa and Brandabur, 2020; Sperkova, 2018; Borah et al., 2018; Lee et al., 2018; Bendle and Xin, 2016). A major advantage of LDA over other possible alternatives, such as Latent Semantic Analysis, lies in its flexibility. It can easily compare two or more documents, and it is not restricted in any way by the kind of language (e.g. English, Spanish, etc.) used in the text. LDA consists of two parts, namely: (i) a probability model that describes text as a likelihood function and (ii) an approximate inference algorithm that uses Gibbs sampling to find the optimal topic assignment instead of simply maximizing the likelihood function, since this would be computationally unfeasible (Blei et al., 2003; Shirota et al., 2015; Zhao et al., 2015; Schwarz, 2018;

Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Reisenbichler and Reutterer, 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021).

The LDA's Probabilistic Model

Figure 3 provides an illustration of the LDA algorithm schematic. The schematic can be used to explain the model. It considers that each document m out of the total number of documents M is a probabilistic combination of T topics.



Source: Buenaño-Fernandez et al. (2020)

Figure 3. *Schematic of LDA algorithm*

A probability vector θ_m of length T captures the probabilities. The value of T (i.e. the number of topics) is arbitrary and depends on the LDA user and the precision required. The LDA outcome is a $M\,x\,T$ matrix of θ with probabilities $p(t_t \mid m_m)$, where $\theta_1, \ldots, \theta_M$ are $1\ x\ T$ vectors. The probability being measured is that of document belonging to topic. It is defined as:

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_M \end{pmatrix} = \begin{pmatrix} p(t_1|m_1) & \cdots & p(t_T|m_1) \\ \vdots & \ddots & \vdots \\ p(t_1|m_m) & \cdots & p(t_T|m_m) \end{pmatrix} \tag{1}$$

The set of words (of size $t \in T$) in all the documents constitutes the vocabulary used in the LDA. Each topic V is determined as a probabilistic distribution over the text vocabulary. In this case, a topic can be used to determine the likelihood of a word associated with it. In the same way as documents can be clustered as probabilistic distributions over topics, words can also be represented as probabilistic vectors of each topic in a $V x T$ matrix:

$$\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_M \end{pmatrix} = \begin{pmatrix} p(w_1|t_1) & \cdots & p(w_1|t_T) \\ \vdots & \ddots & \vdots \\ p(w_v|t_1) & \cdots & p(w_v|t_T) \end{pmatrix} \tag{2}$$

In matrix 2, the probabilities $p(w_v|t_T)$ reflect the probability that a word v can be detected from the vocabulary conditional on topic t. According to Blei et al. (2003), Lee et al. (2018), Schwarz (2018), Edison and Carcel (2021), and Doh et al. (2021), this means that φ can be used to decide on the content appropriateness of each topic and its ultimate name.[8] Blei et al. (2003) explain that, given the LDA model parameters θ and φ, the model considers the data text in the corpus as being created by a two-part process: (i) a word probability distribution $\varphi \sim Dir(\beta)$ and (ii) topic proportions $\theta_d \sim Dir(\alpha)$, for each document m in the text. For each of the N_m words w_M, the topic assignment is drawn such that $z_{m,n} \sim Multinomial(\theta_m)$, while each word $w_{m,n}$ is drawn from $p(w_{m,n}|z_{m,n},\varphi)$. In the two-part process described above, α and β are known as Dirichlet priors, or hyperparameters, that are both greater than zero and are necessary for the Gibbs sampling process explained later. According to Blei et al. (2003), the corpus likelihood with respect to the model parameters is:

$$\prod_{m=1}^{M} P(\theta_m|\alpha) \left(\prod_{n=1}^{N_M} \sum_{z_{m,n}} P\left(z_{m,n}|\theta_M\right) P\left(w_{m,n}|z_{m,n,}\varphi\right)\right), \tag{3}$$

where $P(\theta_m|\alpha)$ measures the likelihood of observing the topic distribution θ_m of document m conditional on α. $P(z_{m,n}|\theta_M)$ describes how likely the topic assignment $z_{m,n}$ of word n in document m is conditional on the topic distribution of the document. Last, $P(w_{m,n}|z_{m,n},\varphi)$ measures the probability of detecting a specific word conditional on the word's topic assignment and the word possibilities of given topics contained in φ. When the sum-up of all possible topic assignments $\sum_z$ is made, along with the product of all N_m words in a document $\prod_{n=1}^{N_M}$, as well as the product of all documents in the corpus $\prod_{m=1}^{M}$, this gives the likelihood of observing the words in the documents. This means that the LDA is concerned with optimal topic assignment $z_{m,n}$ for each word in each document, coupled with the optimal word probabilities φ for each topic that maximizes this likelihood (Blei et al., 2003; Shirota et al., 2015; Schwarz, 2018; Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021). Summing over all possible topic assignments for all words in all the documents is a computationally unfeasible exercise. The alternative is to

8 LDA does not provide names to the topics but rather allows the user to decide on appropriate names given their knowledge of the subject matter under study (Schwarz, 2018; Edison and Carcel, 2021).

approximate the likelihood function using Gibbs sampling as developed by Griffiths and Steyvers (2004).

Approximate Inference Using Gibbs Sampling

According to Schwarz (2018) and Edison and Carcel (2021), the Gibbs sampler is based on a Markov Chain Monte Carlo (MCMC) algorithm that repeatedly draws new samples conditional on all the available data. In the case of the LDA, the Gibbs sampler updates the topic assignment of words conditional on the topic assignments of all other words in the corpus. The technique is Bayesian. As such, it relies on values being identified for the hyperparameters α and β that were discussed earlier. Both α and β lie in the unit interval. The α prior is chosen based on the number of topics T, while the β prior is dependent on the size of the vocabulary. When the Gibbs sampler is run for a burn-in of several hundred iterations, the Markov chain converges towards a maximum of the likelihood function (Schwarz, 2018). Although the LDA is the most widely used topic modelling technique, it is not free from limitations. According to Reisenbichler and Reutterer (2019), the LDA often requires extensive parameter optimization before running as well as the likelihood that topics are arranged more to fit the needs of the researcher (self-selection) than capturing what is in the corpus.

LDA Implementation

In the same way as Edison and Carcel (2021) as well as Vílchez-Román et al. (2019) proceeded, we implement the LDA algorithm using "ldagibbs", a community-contributed command for STATA developed by Schwarz (2018). The ldagibbs package implements LDA through a two-part process. First, the document is divided into single word tokens. Each word is then randomly assigned with equal probability to one of the T topics. Second, new topic assignments for each of the word tokens are sampled on the basis of the following probability model of word token to topic t assignment:

$$P\left(z_{m,n} = t \middle| w_{m,n}, \varphi\right) \ \propto \ P\left(w_{m,n} \middle| z_{m,n} = t, \varphi\right) \ . \ P(z_{m,n} = t) \tag{4}$$

Schwarz (2018) explains that the topic assignment of all other word tokens is used by the Gibbs sampler to acquire approximate values for $P(z_{m,n}= t \mid w_{m,n},\varphi)$ and $P(z_{m,n}= t)$ in equation 4. The $P(w_{m,n} \mid z_{m,n}= t,\varphi)$ is given by the number of words identical to $w_{m,n}$ that are assigned to topic t divided by the total number of words assigned to that topic.

LDA Algorithmic Settings

In our study, three topic models with five, four, and three respective topics were ran. The most ideal topic choice was based on the speed of convergence of each topic model as indicated by the likelihood ratio. In a similar way to Schwarz (2018), Edison and Carcel (2021), and Vílchez-Román et al. (2019), the following settings are applied to each topic model: $\alpha = 0.25$, $\beta = 0.1$, with a burn-in period of 1,000 (i.e. the number of iterations the Gibbs sampler should run). An assessment of model clustering (topic) output can be evaluated at various stages and could be done by running the same model either with varying Dirichlet priors or with a different number of topics (Blei et al., 2003; Shirota et al., 2015; Zhao et al., 2015; Schwarz, 2018; Dwivedi, 2018; Mahanty et al., 2019; Böök, 2019; Vílchez-Román et al., 2019; Rabindranath, 2020; Edison and Carcel, 2021; Doh et al., 2021).

5. Results and Analysis

The results and analysis of the topic modelling exercise are presented in this section. The discussion that follows is divided into two parts, namely: (i) topic extraction and formulation and (ii) topic distribution over time.

5.1 Topic Extraction and Average Topic Trends

To extract topics from CBL Monetary Policy Statements, the study ran three versions of the LDA model: a three-topic, a four-topic, and a five-topic LDA model. The study also performed robustness iterations, where the Dirichlet priors were systematically altered in each model. This was done to allow for an assessment of model validity in the process of topic extraction. The choice of topic number in each model was guided by the four functions of the CBL MPC, as stated in the CBL MPC Charter 2020. The sections that follow discuss the topic extraction (and formulation) as well as the average topic trends under each version of the LDA model.

Results of Three-Topic LDA

Appendix A2 presents the word probability matrix generated from the Three-Topic LDA. Each topic's word probability vectors are a description of the likelihood of seeing a word conditional on a topic. For this reason, the word probability sum in a single word probability vector is 1. The most frequent words in each topic are sorted in increasing order of word probability to reveal the top five words in each topic. This allows for the adding of topic labels. *Table 2* shows the top five words per topic from the Three-Topic LDA.

Table 2: *Top five words per topic in the Three-Topic LDA*

Topic 1	Topic 2	Topic 3
intern	committe	intern
condit	monetari	econom
monetary	intern	polici
financi	reserv	condit
market	polici	region

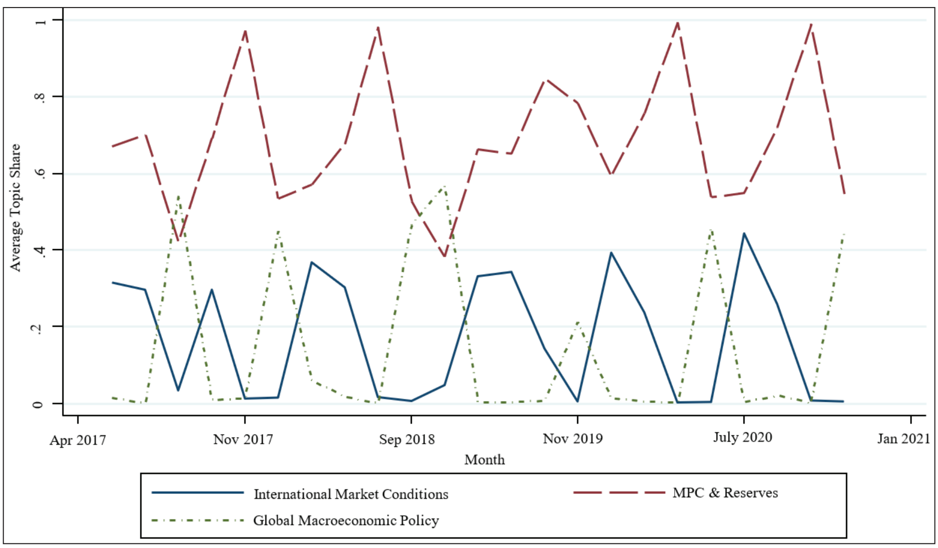
Source: author's classification

Table 3 shows the Three-Topic LDA topic labels. *Figure 4* illustrates the average topic trends over the sample period. Upon close observation, the extracted topics are closely reflective of the functions of the CBL Monetary Policy Committee, as discussed in Section 2.

Table 3. *Three-Topic LDA topic labels*

Topic No.	Topic Label	Shorter Variant
Topic One	International Monetary and Financial Market Conditions	International Market Conditions
Topic Two	Monetary Policy Committee and International Reserves	Monetary Policy and Reserves
Topic Three	Regional and International Economic Policy Conditions	Global Macroeconomic Policy

Source: author's classification



Source: author's classification

Figure 4. *CBL Monetary Policy Statement average topic trends – Three-Topic LDA*

From *Figure 4*, the Monetary Policy Committee (MPC) and Reserves topic has the highest average topic share of all the three topics plotted. This implies that it is the topic that occupied the most discussion during the review period. This makes intuitive sense, seeing that the ultimate work of the CBL MPC is to ensure that there are enough international reserves to maintain the one-to-one fixed exchange rate between Lesotho's currency and that of South Africa.

Results of Four-Topic LDA

Like the Three-Topic LDA model, the four-topic word probability matrix in the Four-Topic LDA model is presented in *Appendix A3*. Comparing *Appendix A2* to *A3*, the word probabilities relative to the top five words in the Four-Topic LDA model are less than those in the Three-Topic LDA word probability matrix. The top five words in the Four-Topic LDA word probability matrix, with the highest probability of belonging to the topics, are selected and presented in *Table 4*. They are used to develop the respective topic labels.

Table 4. *Top words in Four-Topic LDA*

Topic 1	Topic 2	Topic 3	Topic 4
economic	quarter	global	economic
COVID-19	growth	sector	risks
global	economic	these	global
expected	domestic	likely	policy
measures	quarter	financial	quarter

Source: author's classification

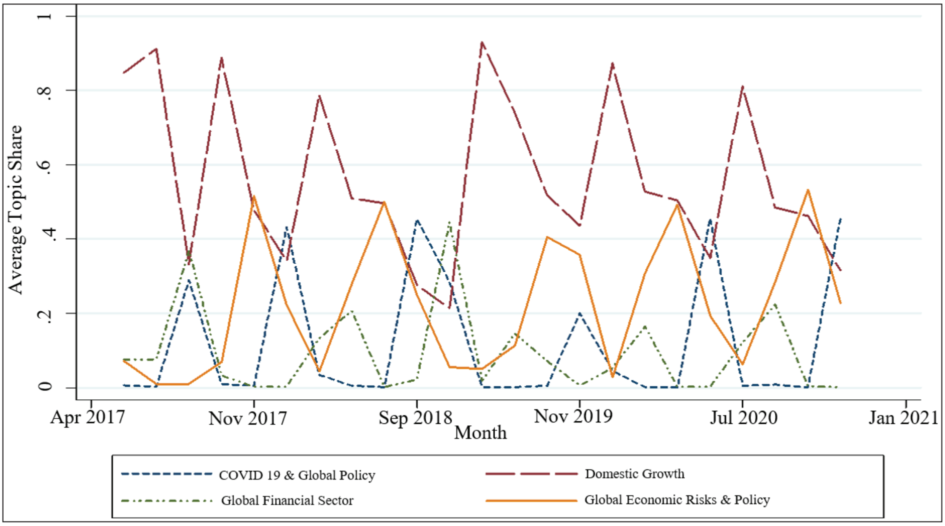
Although there are some similarities in the words per topic in the Four-Topic LDA relative to the Three-Topic LDA, the Four-Topic LDA reveals other aspects of discussion that featured in the CBL Monetary Policy Statements, such as the COVID-19 global pandemic and associated global economic measures. *Table 5* shows the Four-Topic LDA topic labels.

Table 5. *Four-Topic LDA topic labels*

Topic No.	Topic Label	Shorter Variant
Topic One	COVID-19 and Global Economic Measures	COVID-19 and Global Policy
Topic Two	Domestic Economic Growth	Domestic Growth
Topic Three	Global Financial Sector	Financial Sector
Topic Four	Global Economic Risks and Policy	Economic Risks and Policy

Source: author's classification

Figure 5 is an illustration of the average topic trends over the sample period. Like the Three-Topic LDA model, topics that most closely related to the functions of the CBL MPC were the most prominent; that is, the discussions on domestic economic growth as well as global economic risks and policy had the highest average topic shares.



Source: author's classification

Figure 5. *CBL Monetary Policy Statement average topic trends – Four-Topic LDA*

Results of Five-Topic LDA

Appendix A4 presents the world probability matrix from the Five-Topic LDA model. The words with the highest probability of belonging to the topic are used to develop the respective topics, as can be inferred from *tables 6–7*.

Table 6. *Top words in Five-Topic LDA*

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
risks	quarter	performance	economic	likely
global	economic	estimated	covid-19	global
price	growth	first	pandemic	expected
remained	domestic	credit	measures	these
measured	developments	positive	global	banks

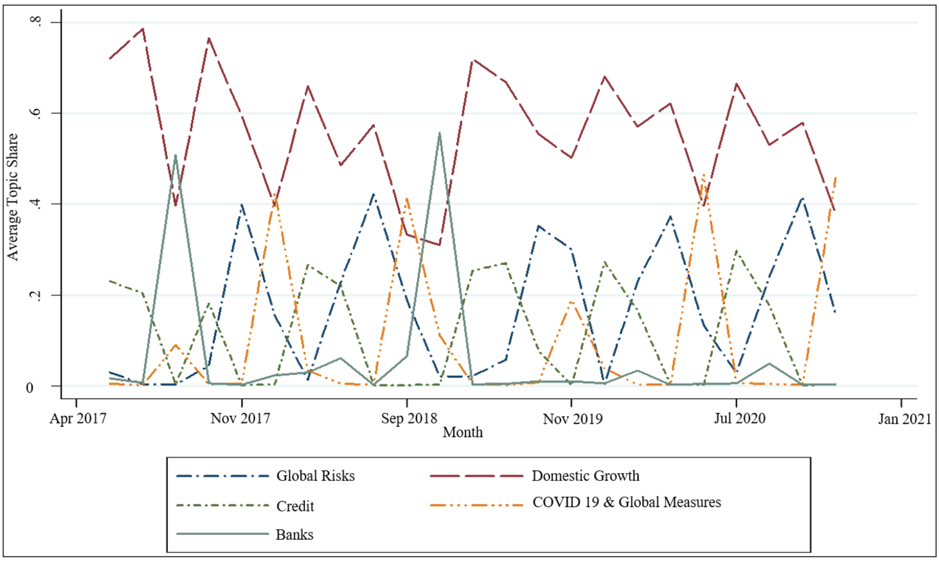
Source: author's classification

Table 7. *Five-Topic LDA topic labels*

Topic No.	Topic Label	Shorter Variant
Topic One	Global Risks and Prices	Global Risks and Prices
Topic Two	Domestic Economic Growth	Domestic Growth
Topic Three	Credit Performance	Credit
Topic Four	COVID-19 and Global Economic Measures	COVID-19 and Global Measures
Topic Five	Global Banks	Banks

Source: author's classification

The average topic share in the Five-Topic LDA model is presented in *Figure 6*. The trend is broadly similar to that observed in the Three- and Four-Topic LDA models, that is, the most prominent topics of discussion involved domestic economic growth and global policy measures.



Source: author's classification

Figure 6. *CBL Monetary Policy Statement average topic trends – Five-Topic LDA*

6. Conclusions and Recommendations

This paper's objective was to use Latent Dirichlet Allocation (LDA) to analyse the monetary policy statements of Lesotho's Central Bank from February 2017 to January 2021 and to identify the evolutionary trend of key themes or topics discussed by members of the CBL Monetary Policy Committee (MPC) over the sample period. In the same way as Edison and Carcel (2021), Doh et al. (2021), Vílchez-Román et al. (2019), and Shirota et al. (2015), our study shows that the LDA technique can extract topics from corpora without any prior knowledge and determine their importance and appropriateness based on the probability of occurrence. However, according to Sperkova (2018), if the technique used is unsupervised, as is the case in our study, it may contain words that are not immediately interpretable. In this respect, it requires a high substantive knowledge of the subject being analysed and classified.

In the current study, the Three-Topic LDA model extracted topics that remained prominent throughout the sample period and were most closely reflective of the functions of the CBL Monetary Policy Committee. The topics identified were: (i) International Monetary and Financial Market Conditions, (ii) Monetary Policy Committee and International Reserves, and (iii) Regional and International Economic Policy Conditions. Owing to its high word-to-topic probability association, the Three-Topic LDA model was determined as the most appropriate model through which a consistent analysis of topic evolution in CBL Monetary Policy Statements can be performed.

The findings of the current study provide a knowledge base for analysing the monetary policy statements of the CBL. As in Doh et al. (2021), the findings indicate that the evolution of MPC discussions and the tone of the MPC statements over time are influenced by incoming economic data. The results can be used to guide future research. To this end, the study recommends the continued use of LDA, especially the Three-Topic LDA model, in the analysis of topic evolution in CBL Monetary Policy Statements. This can help reveal useful relationships in key policy topic priorities of discussions and their evolution over time.

7. Areas for Further Study

The study has some limitations. First, the study timeline ranges from February 2017 to January 2021, and the corpus comprises texts from twenty-six MPC statements. The choice of timeline and the number of MPC statements was underpinned by the presence of publicly available data during the time of the study. Second, the study is of a single-country perspective, meaning that it fails to leverage on regional monetary policy dynamics to add a layer of robustness in the findings. Future extensions to the study could consider examining a longer timeline and a larger corpus of text from an increased number of MPC statements. Follow-up studies could also consider using topic classification to evaluate the monetary policy stance of the CBL and comparing it to that of other central banks within the same region across the same time period.

References

"Artificial Intelligence", Merriam-Webster.com. (n. d). Available online: https://www.merriam-webster.com/dictionary/artificial%20intelligence. (accessed on: 15 April 2021).

Asmussen, C. B.; Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data* 6(1): 1–18.

Bendle, N. T.; Wang, X. S. (2016). Uncovering the message from the mess of big data. *Business Horizons* 59(1): 115–124.

Blei, D. M.; Jordan, M. I. (2003, July). Modeling annotated data. In: *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 127–134.

Blei, D. M.; Ng, A. Y.; Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3: 993–1022.

Borah, A.; Wang, X. S.; Ryoo, J. H. J. (2018). Understanding influence of marketing thought on practice: An analysis of business journals using textual and Latent Dirichlet Allocation (LDA) analysis. *Customer Needs and Solutions* 5(3): 146–161.

Böök, A. (2019). Sentiment analysis on FOMC statements. Available online: https://www.linkedin.com/pulse/sentiment-analysis-fomc-statements-arthur-b%C3%B6%C3%B6k. (accessed on: 11 March 2021).

Buenaño-Fernandez, D.; González, M.; Gil, D.; Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access* 8: 35318–35330.

CBL (Central Bank of Lesotho). (2020). Monetary Policy Committee Charter – November 2020 Revision. *Central Bank of Lesotho*.

Cedervall, A.; Jansson, D. (2018). Topic classification of monetary policy minutes from the Swedish Central Bank. Available online: https://www.diva-portal.org/smash/get/diva2:1272108/FULLTEXT01.pdf. (accessed on: 11 March 2021).

Central Bank of Lesotho Act (2000). Available online: https://www.centralbank.org.ls/images/Legislation/Principal/CBL_Act_2000.pdf. (accessed on: 16 March 2021).

Damane, M. (2019). The effects of regional cross listing on firm value and financial performance: Drawing lessons for Lesotho. *African Journal of Accounting, Auditing and Finance* 6(3): 201–235.

(2020). Investigating the structural stability of the theoretical money multiplier model in Lesotho: A flexible least squares approach. *International Journal of Statistics and Economics* 21(2).

Doh, T.; Kim, S.; Yang, S. K. (2021). How you say it matters: Text analysis of FOMC statements using natural language processing. *Economic Review* 106(1): 5–20. Federal Reserve Bank of Kansas City.

Dwivedi, P. (2018). NLP: Extracting the main topics from your dataset using LDA in minutes. *Towards data science*. Available online: https://towardsdatascience.com/nlp-extracting-the-main-topics-from-your-dataset-using-lda-in-minutes-21486f5aa925. (accessed on 11 March 2021).

Edison, H.; Carcel, H. (2021). Text data analysis using Latent Dirichlet Allocation: An application to FOMC transcripts. *Applied Economics Letters* 28(1): 38–42.

Financial Stability Board (2017). Artificial intelligence and machine learning in financial services. Available online: https://www.fsb.org/wp-content/uploads/P011117.pdf. (accessed on: 14 April 2021).

Gartner. (2020). Hype cycle for emerging technologies. Available online: https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/. (accessed on: 21 April 2021).

Griffiths, T. L.; Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences* 101: 5228–5235.

Knütter, R.; Mohr, B.; Wagner, H. (2011). The effects of central bank communication on financial stability: A systematization of the empirical evidence. Available online: https://www.fernuni-hagen.de/ls_wagner/download/publikationen/dp463.pdf. (accessed on: 12 April 2021).

Lee, H.; Seo, H.; Geum, Y. (2018). Uncovering the topic landscape of product-service system research: From sustainability to value creation. *Sustainability* 10(4): 911.

Mahanty, S.; Boons, F.; Handl, J.; Batista-Navarro, R. (2019, November). Studying the evolution of the 'circular economy' concept using topic modelling. In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Cham. 259–270.

Popa, Alina; Brandabur, Raluca-Ecaterina. (2019). Evolution of integrated marketing communication research through Latent Dirichlet Allocation (LDA) analysis. *Journal of Emerging Trends in Marketing and Management* 1(1): 61–70.

Quinn, K. M.; Monroe, B. L.; Colaresi, M.; Crespin, M. H.; Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* 54(1): 209–228.

Rabindranath, G. (2020). Modeling topic trends in FOMC meetings: How to use LDA topic modeling to analyze content trends. *Towards data science.* Available online: https://highdemandskills.com/topic-trends-fomc/. (accessed on: 11 March 2021).

Reisenbichler, M.; Reutterer, T. (2019). Topic modeling in marketing: Recent advances and research opportunities. *Journal of Business Economics* 89(3): 327–356.

Schwarz, C. (2018). ldagibbs: A command for topic modeling in Stata using latent Dirichlet allocation. *The Stata Journal* 18(1): 101–117.

Shirota, Y.; Yano, Y.; Hashimoto, T.; Sakura, T. (2015, July). Monetary policy topic extraction by using LDA: Japanese monetary policy of the second ABE cabinet term. In: *2015 IIAI 4th International Congress on Advanced Applied Informatics*. IEEE. 8–13.

Šperková, L. (2018). Review of Latent Dirichlet Allocation methods usable in voice of customer analysis. *Acta Informatica Pragensia* 7(2): 152–165.

"Structured data". Your dictionary. (n. d.). Available online: https://www.yourdictionary.com/structured-data. (accessed on: 15 April 2021).

Vílchez-Román, C.; Huamán-Delgado, F.; Sanguinetti-Cordero, S. (2019). Topic modeling applied to business research: A Latent Dirichlet Allocation (LDA)-based classification for organization studies. *Communications in Computer and Information Science* 212–219. DOI: 10.1007/978-3-030-11680-4_21.

Zhao, W.; Chen, J. J.; Perkins, R.; Liu, Z.; Ge, W.; Ding, Y.; Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics* 16(3): 1–10. BioMed Central.

Appendices

A1. Summary of empirical literature

Author(s)	Objective	Data	Methodology	Results
Edison and Carcel (2021)	Use topic classification to explore evolution in topics discussed by Federal Open Market Committee members between 2003 and 2012.	A total of 80 meeting transcripts of the Federal Open Market Committee during 2003–2012.	Identification of source of text corpus – data collection. Pre-processing. Topic classification with Latent Dirichlet Allocation (LDA). Assessment of model clustering (topic) output is evaluated at various stages through an iterative process. Topic selection is critiqued on the basis of authors' discretion and understanding of the subject matter.	The study identified four topics that were as follows: economic modelling, banking, economic activity, and communication. Discussions on economic modelling greatly featured during the Great Financial Crisis, while discussions on the banking system followed in subsequent years. Communication gained prominence as a discussion topic later in the sample.

Author(s)	Objective	Data	Methodology	Results
Doh et al. (2021)	To identify the tone of post-meeting statements of the Federal Open Market Committee.	A total of 87 Federal Open Market Committee statements from March 2004 through December 2014.	Identification of source of text corpus – data collection. Pre-processing. Natural language processing tools.	Information about the Federal Open Market Committee's quantitative decision on the target policy rate is as important as an assessment of risk. If information on how the Federal Open Market Committee interprets incoming data is included into the statement, this has a significant impact on the statement's overall tone.
Vílchez-Román et al. (2019)	Use topic classification to identify primary topics in organization studies between 1970 and 2015.	Academic document abstracts under the business and management subject research during 1970–2015.	Identification of source of text corpus – data collection. Pre-processing. Topic classification with Latent Dirichlet Allocation (LDA).	The study identified that a three-topic-based classification system was more stable instead of a four- or five-topic-based one. The most popular topics were organizational behaviour, organizational management, and service management. Document classification under the three extracted topics showed that the first 10 documents in the dataset were most represented under the topic: organizational management.
Sperkova (2018)	To use topic modelling in voice of customer textual analysis for performing tasks of emotion, personality, and sentiment detection.	Customer reviews in the form of textual comments.	Identification of source of text corpus – data collection. Pre-processing. Topic classification with Latent Dirichlet Allocation (LDA).	LDA allows dynamics to be evaluated at a highly granular temporal level over time.

Author(s)	Objective	Data	Methodology	Results
Lee et al. (2018)	To delineate the thematic landscape of the product-service system research between 2000 and 2016.	1,229 product-service system publications between 2000 and 2016.	Identification of source of text corpus – data collection. Pre-processing. Topic classification with Latent Dirichlet Allocation (LDA).	Ten product-service system topics are selected using LDA. Results reflect changes in focus in product-service system topics over time.
Shirota et al. (2015)	To use topic extraction to analyse the monetary policy minutes of the Bank of Japan, under the second Cabinet term of Prime Minister Abe.	Minutes of the Central Bank of Japan Monetary Policy Meeting from January 2013 to June 2014.	Identification of source of text corpus – data collection. Pre-processing. Topic classification with Latent Dirichlet Allocation (LDA). Topic choice is supported by consistency to topic selection with significant co-occurring economic and/or financial developments across the sample period.	The study identified five topics that ranged from domestic to international matters. Extracted topics showed that the most prominent discussions were around monetary policy easing and consumption tax increase that prevailed during the sample period.

A2. Three-topic word probability matrix

i. Sorted by increasing order in Word_prob1

words	word_prob1	word_prob2	word_prob3
intern	0.10204082	0.07482993	0.09289617
condit	0.08843537	0.04761905	0.07103825
monetari	0.07482993	0.10204082	0.03825137
financi	0.06802721	0.02040816	0.05464481
market	0.06802721	0.03401361	0.04371585
domest	0.06802721	0.04081633	0.03825137
polici	0.05442177	0.06802721	0.08196721
region	0.05442177	0.01360544	0.07103825
develop	0.05442177	0.04081633	0.04918033
reserv	0.05442177	0.07482993	0.02185792
consid	0.04761905	0.03401361	0.06010929
outlook	0.04081633	0.03401361	0.06557377

words	word_prob1	word_prob2	word_prob3
lesotho	0.04081633	0.04761905	0.05464481
stanc	0.04081633	0.02721088	0.01639344
committe	0.03401361	0.17687075	0.06557377
central	0.03401361	0.04761905	0.06010929
april	0.03401361	0.04081633	0.01092896
determin	0.02721088	0.04081633	0.01639344
econom	0.01360544	0.03401361	0.08743169
SUM	**1.00000002**	**1.00000003**	**0.99999999**

ii. Sorted by increasing order in Word_prob2

words	word_prob1	word_prob2	word_prob3
committe	0.03401361	0.17687075	0.06557377
monetari	0.07482993	0.10204082	0.03825137
intern	0.10204082	0.07482993	0.09289617
reserv	0.05442177	0.07482993	0.02185792
polici	0.05442177	0.06802721	0.08196721
condit	0.08843537	0.04761905	0.07103825
lesotho	0.04081633	0.04761905	0.05464481
central	0.03401361	0.04761905	0.06010929
domest	0.06802721	0.04081633	0.03825137
develop	0.05442177	0.04081633	0.04918033
april	0.03401361	0.04081633	0.01092896
determin	0.02721088	0.04081633	0.01639344
market	0.06802721	0.03401361	0.04371585
consid	0.04761905	0.03401361	0.06010929
outlook	0.04081633	0.03401361	0.06557377
econom	0.01360544	0.03401361	0.08743169
stanc	0.04081633	0.02721088	0.01639344
financi	0.06802721	0.02040816	0.05464481
region	0.05442177	0.01360544	0.07103825
SUM	**1.00000002**	**1.00000003**	**0.99999999**

iii. Sorted by increasing order in Word_prob3

words	word_prob1	word_prob2	word_prob3
intern	0.10204082	0.07482993	0.09289617
econom	0.01360544	0.03401361	0.08743169
polici	0.05442177	0.06802721	0.08196721
condit	0.08843537	0.04761905	0.07103825
region	0.05442177	0.01360544	0.07103825
committe	0.03401361	0.17687075	0.06557377
outlook	0.04081633	0.03401361	0.06557377
central	0.03401361	0.04761905	0.06010929
consid	0.04761905	0.03401361	0.06010929
lesotho	0.04081633	0.04761905	0.05464481
financi	0.06802721	0.02040816	0.05464481
develop	0.05442177	0.04081633	0.04918033
market	0.06802721	0.03401361	0.04371585
monetari	0.07482993	0.10204082	0.03825137
domest	0.06802721	0.04081633	0.03825137
reserv	0.05442177	0.07482993	0.02185792
determin	0.02721088	0.04081633	0.01639344
stanc	0.04081633	0.02721088	0.01639344
april	0.03401361	0.04081633	0.01092896
SUM	**1.00000002**	**1.00000003**	**0.99999999**

A3. Four-topic word probability matrix

i. Sorted by increasing order in Word_prob1

words	word_prob1	word_prob2	word_prob3	word_prob4
economic	0.03987342	0.02282416	0.00054007	0.02572554
covid-19	0.02359033	0.00002011	0.00010801	0.00008561
global	0.01962025	0.00478604	0.01717434	0.02508347
expected	0.0164557	0.00673665	0.00723698	0.00025683
measures	0.01639816	0.00012066	0.00010801	0.00008561
pandemic	0.0159954	0.00004022	0.00032404	0.0000428
likely	0.01409666	0.00034186	0.01404191	0.00222584
outlook	0.01352129	0.01132159	0.00032404	0.00226864

words	**word_prob1**	**word_prob2**	**word_prob3**	**word_prob4**
decline	0.01329114	0.00553008	0.00043206	0.00059926
spread	0.0132336	0.00002011	0.00010801	0.00008561
growth	0.01058688	0.02582046	0.00054007	0.00517935
relative	0.01058688	0.00020109	0.00043206	0.00068487
contract	0.00966628	0.00018098	0.00064809	0.0000428
economy	0.00926352	0.00528877	0.0036725	0.00025683
		(output omitted)		
SUM	**1.00000044**	**1.00000116**	**0.99999793**	**0.99999858**

ii. Sorted by increasing order in Word_prob2

words	**word_prob1**	**word_prob2**	**word_prob3**	**word_prob4**
quarter	0.00126582	0.02765042	0.00010801	0.01780669
growth	0.01058688	0.02582046	0.00054007	0.00517935
economic	0.03987342	0.02282416	0.00054007	0.02572554
domestic	0.00701956	0.01972732	0.00032404	0.00933139
developments	0.00143843	0.01697233	0.00626485	0.003724
monetary	0.00017261	0.01325209	0.00108015	0.00350997
increase	0.00080552	0.01319176	0.00162022	0.00124133
economies	0.00069045	0.013051	0.00097213	0.0113004
committee	0.00005754	0.01278958	0.00216029	0.00029963
during	0.0004603	0.01264881	0.0044286	0.00068487
inflation	0.00126582	0.01226673	0.00108015	0.01121479
outlook	0.01352129	0.01132159	0.00032404	0.00226864
compared	0.00034522	0.0102759	0.00097213	0.00077048
deficit	0.00005754	0.01003459	0.00010801	0.00012841
		(output omitted)		
SUM	1.00000044	1.00000116	0.99999793	0.99999858

iii. Sorted by increasing order in Word_prob3

words	**word_prob1**	**word_prob2**	**word_prob3**	**word_prob4**
global	0.01962025	0.00478604	0.01717434	0.02508347
sector	0.00011507	0.00832529	0.01458198	0.0018406
these	0.00017261	0.00006033	0.01436595	0.00021402
likely	0.01409666	0.00034186	0.01404191	0.00222584
financial	0.00040276	0.00368002	0.01263772	0.00415204
africa	0.00086306	0.00012066	0.01231367	0.01108638
negative	0.00241657	0.00006033	0.01036941	0.00012841
number	0.00195627	0.00032175	0.00972132	0.00059926
banks	0.00063291	0.00004022	0.00939728	0.00017122
import	0.00028769	0.00378057	0.00918125	0.00068487
remain	0.00143843	0.00287564	0.00864118	0.00196901
major	0.00069045	0.00237291	0.0081011	0.00295351
expected	0.0164557	0.00673665	0.00723698	0.00025683
result	0.00011507	0.00378057	0.00702095	0.00081329
		(output omitted)		
SUM	**1.00000044**	**1.00000116**	**0.99999793**	**0.99999858**

iv. Sorted by increasing order in Word_prob4

words	**word_prob1**	**word_prob2**	**word_prob3**	**word_prob4**
economic	0.03987342	0.02282416	0.00054007	0.02572554
risks	0.00005754	0.00004022	0.00021603	0.02521188
global	0.01962025	0.00478604	0.01717434	0.02508347
policy	0.00005754	0.00969273	0.00021603	0.02135947
quarter	0.00126582	0.02765042	0.00010801	0.01780669
activity	0.00420023	0.00892857	0.00129618	0.01605171
price	0.00057537	0.00008044	0.00086412	0.01592329
remained	0.00005754	0.00780245	0.00021603	0.01553805
measured	0.00132336	0.00032175	0.00021603	0.01511001
international	0.00166858	0.00583172	0.00097213	0.01395428
second	0.00011507	0.00154842	0.00054007	0.01369746
declined	0.00040276	0.00010055	0.00054007	0.01271295
rates	0.00069045	0.0008647	0.00043206	0.01262734
other	0.00017261	0.00094514	0.00043206	0.01155723
		(output omitted)		
SUM	**1.00000044**	**1.00000116**	**0.99999793**	**0.99999858**

A4. Five-topic word probability matrix

i. Sorted by increasing order in Word_prob1

words	word_prob1	word_prob2	word_prob3	word_prob4	word_prob5
risks	0.0312308	0.00025517	0.0006152	0.00102459	0.00015004
global	0.02474999	0.00897011	0.00010253	0.0161373	0.02580645
price	0.01972177	0.00051033	0.00051266	0.00064037	0.00015004
remained	0.01927482	0.00700728	0.00512663	0.00006404	0.00030008
measured	0.01854852	0.00058885	0.00041013	0.001729	0.00030008
economic	0.01849265	0.03018823	0.00071773	0.03425973	0.0036009
declined	0.0149729	0.00060848	0.00020507	0.00025615	0.00150038
policy	0.01486117	0.01332758	0.0003076	0.00198514	0.00090023
trade	0.0148053	0.00023554	0.00010253	0.00108863	0.00150038
activity	0.01396726	0.01171806	0.00051266	0.00307377	0.00075019
rates	0.01346444	0.00188431	0.00010253	0.00096055	0.00030008
first	0.01301749	0.00147212	0.01445709	0.00032018	0.00015004
tensions	0.01089446	0.00001963	0.00020507	0.00038422	0.00015004
include	0.01078273	0.00009814	0.0003076	0.00147285	0.00015004
(output omitted)					
SUM	**1.00000114**	**0.9999998**	**0.99999941**	**1.00000127**	**1.0000025**

ii. Sorted by increasing order in Word_prob2

words	word_prob1	word_prob2	word_prob3	word_prob4	word_prob5
quarter	0.00536343	0.03344652	0.00102533	0.00012807	0.00045011
economic	0.01849265	0.03018823	0.00071773	0.03425973	0.0036009
growth	0.00128499	0.02930496	0.00010253	0.00384221	0.00270068
domestic	0.00553104	0.02155181	0.00010253	0.00787654	0.00060015
developments	0.00122912	0.01856832	0.00010253	0.00044826	0.00585146
economies	0.0039667	0.01666438	0.0003076	0.00038422	0.00090023
inflation	0.00363149	0.01578111	0.00010253	0.00198514	0.00060015
monetary	0.0008939	0.0143875	0.00010253	0.00006404	0.00060015
increase	0.00027935	0.0133472	0.00184559	0.00025615	0.0012003
policy	0.01486117	0.01332758	0.0003076	0.00198514	0.00090023
outlook	0.00067043	0.01240505	0.00071773	0.01299949	0.00015004
during	0.00005587	0.01183583	0.00779247	0.0005123	0.00105026
committee	0.00055869	0.01179657	0.00512663	0.00006404	0.00045011
activity	0.01396726	0.01171806	0.00051266	0.00307377	0.00075019
(output omitted)					
SUM	**1.00000114**	**0.9999998**	**0.99999941**	**1.00000127**	**1.0000025**

iii. Sorted by increasing order in Word_prob3

words	**word_prob1**	**word_prob2**	**word_prob3**	**word_prob4**	**word_prob5**
performance	0.0014526	0.00113844	0.01814826	0.00083248	0.00015004
estimated	0.00022348	0.00047108	0.01599508	0.00057633	0.00030008
first	0.01301749	0.00147212	0.01445709	0.00032018	0.00015004
credit	0.00016761	0.00563331	0.01394443	0.00012807	0.00105026
positive	0.00005587	0.00029442	0.01281657	0.00275359	0.00015004
review	0.00050282	0.00117769	0.00984313	0.00006404	0.00135034
which	0.00094977	0.00054959	0.00974059	0.00076844	0.00045011
official	0.00027935	0.00047108	0.00974059	0.00064037	0.00015004
unchanged	0.00335214	0.00157026	0.0085102	0.00006404	0.00015004
registered	0.00011174	0.00449487	0.00830514	0.00064037	0.00045011
during	0.00005587	0.01183583	0.00779247	0.0005123	0.00105026
compared	0.00055869	0.00891122	0.00779247	0.00019211	0.00030008
above	0.00217889	0.00098141	0.00758741	0.00006404	0.00015004
african	0.00033521	0.00084401	0.00748488	0.01018186	0.00060015
(output omitted)					
SUM	**1.00000114**	**0.9999998**	**0.99999941**	**1.00000127**	**1.0000025**

iv. Sorted by increasing order in Word_prob4

words	word_prob1	word_prob2	**word_prob3**	**word_prob4**	**word_prob5**
economic	0.01849265	0.03018823	0.00071773	0.03425973	0.0036009
covid-19	0.00005587	0.00005888	0.00010253	0.02593494	0.00075019
pandemic	0.00005587	0.00001963	0.00010253	0.017354	0.00165041
measures	0.00061456	0.00019628	0.0003076	0.01639344	0.00225056
global	0.02474999	0.00897011	0.00010253	0.0161373	0.02580645
outlook	0.00067043	0.01240505	0.00071773	0.01299949	0.00015004
spread	0.00005587	0.00001963	0.00010253	0.01223105	0.00615154
relative	0.00094977	0.00007851	0.0006152	0.01191086	0.00030008
recovery	0.00055869	0.00054959	0.00645955	0.01043801	0.00015004
african	0.00033521	0.00084401	0.00748488	0.01018186	0.00060015
expected	0.00044695	0.00702691	0.00020507	0.00998975	0.02565641
likely	0.00167607	0.00123658	0.00010253	0.00998975	0.02925731
remains	0.00005587	0.00049071	0.0047165	0.00966957	0.00030008
contract	0.00011174	0.00017665	0.0003076	0.00851691	0.00570143
(output omitted)					
SUM	**1.00000114**	**0.9999998**	**0.99999941**	**1.00000127**	**1.0000025**

iv. Sorted by increasing order in Word_prob5

words	word_prob1	word_prob2	word_prob3	word_prob4	word_prob5
likely	0.00167607	0.00123658	0.00010253	0.00998975	0.02925731
global	0.02474999	0.00897011	0.00010253	0.0161373	0.02580645
expected	0.00044695	0.00702691	0.00020507	0.00998975	0.02565641
these	0.00005587	0.00005888	0.00143546	0.00012807	0.01875469
banks	0.00016761	0.0001374	0.00082026	0.00012807	0.01275319
given	0.00005587	0.00009814	0.00051266	0.00268955	0.01230308
affected	0.00022348	0.00007851	0.00020507	0.00019211	0.01230308
major	0.00134086	0.00339569	0.00010253	0.00012807	0.01125281
decline	0.00033521	0.00632029	0.0003076	0.00806865	0.01020255
especially	0.00810101	0.00015703	0.00010253	0.00230533	0.00975244
result	0.00022348	0.0039649	0.00051266	0.00006404	0.00945236
negative	0.0007263	0.0007655	0.00123039	0.0012167	0.00930233
short	0.00050282	0.00007851	0.0006152	0.00435451	0.00870218
under	0.00134086	0.0006281	0.00133292	0.00147285	0.00795199
(output omitted)					
SUM	**1.00000114**	**0.9999998**	**0.99999941**	**1.00000127**	**1.0000025**