Article

A New Measure of Complementarity in Market Basket Data

Radosław Puka * and **Stanislaw Jedrusik**

Faculty of Management, AGH University of Science and Technology, 30-067 Cracow, Poland; stj@agh.edu.pl
* Correspondence: rpuka@zarz.agh.edu.pl

Abstract: Modern IT systems collect detailed data on each activity, transaction, forum entry, conversation and many other areas. The availability of large data volumes in the business, industry and research fields opens up new opportunities for the empirical verification of various economic theories and laws. The analysis of big datasets in turn allows us to look at many issues from a new point of view and see the dependencies that are otherwise difficult to derive. In this paper, we propose a new measure for dependencies between goods in market basket data. The introduced measure was inspired by the well-known microeconomic concept of complementarity. Due to its similar properties to those of complementarity, the new measure was called basket complementarity (*b*-complementarity). *B*-complementarity not only measures the strength of dependencies between goods but also measures the direction of these dependencies. The values of the proposed measure can be relatively easily calculated using market basket data. This paper also presents a simple example illustrating this new concept, areas of possible application (e.g., in e-commerce) and preliminary results of searching for goods that meet the criteria of basket complementarity in real market basket data.

Keywords: big data analysis; market basket analysis; basket complementarity; complementary goods; association analysis

1. Introduction

The concept of complementary goods belongs to the realm of microeconomics. Complementary (and substitute) goods first became the subjects of academic discourse in 1894 [1]. Ever since that time, the problem of complementary goods has been constantly present in academic research. A comprehensive description of the development of the concept of complementarity can be found in research conducted by Lenfant [2,3]. According to the common understanding, complementarity is a relationship involving mutual supplementation of goods. In microeconomics, the concept of complementarity can be defined in a number of different ways. The theory of supply and demand defines complementarity by means of a cross-price elasticity of demand, which measures the responsiveness of demand for one good to a change in the price of another good (ceteris paribus). Two goods are considered to be complementary when the cross-price elasticity of demand is negative for these goods.

A very intuitive definition of complementarity is also provided by the theory of consumer choice. The definition of complementarity is based on the correlations between utility and demand. For complementary goods, the marginal utility of one good increases along with the increase in demand for another good.

The definitions of complementarity that can be found in the literature are rather difficult to apply in practice, mainly due to the insufficient amount of data describing demand fluctuations resulting from price changes, and to the necessity of making quite restrictive assumptions about the type of utility function. One of the possible ways to overcome these difficulties is to take advantage of the opportunities provided by association analysis.

Association analysis, also referred to as market basket analysis, investigates the correlations between goods based on purchases made by consumers [4]. If two goods are

considered to be complementary, then they will be more frequently purchased together. So, the more frequently a particular combination of goods appears in a basket, the greater the probability is that these goods are complementary.

The notion of complementarity (but also substitutability) lies at the heart of recommendation systems. Interesting examples of using product complementarity in creating such a system can be found in [5–7].

The main objective of this paper is to define the concept of complementarity of goods based on association analysis. The proposed measure is called basket complementarity (or *b*-complementarity). It provides the possibility to consider asymmetric cases, i.e., cases when a consumer purchasing good *x* gains greater satisfaction if he/she additionally buys good *y* but good *y* may also be purchased without any connection with good *x*. Such properties can be of great importance, e.g., in building pricing strategies or in recommender systems. The 'basket' definition also enables identification of the relations of complementarity that exist regardless of other goods, as well as relations that are a derivative of complementarity with other goods.

The paper consists of an introduction and three sections. Section 1 is devoted to the issue of complementarity of goods in microeconomics. It presents two main approaches to the mathematical formalization of this concept. Section 2 describes the well-known market basket analysis concepts. In Section 3, the definition of complementarity of goods is introduced based on individual shopping baskets. The section also provides a numerical example illustrating the possibilities of using market basket analysis in order to identify complementary goods. The paper ends with concluding remarks. The conclusions point out the advantages provided by the definition of complementarity based on the properties of shopping baskets in comparison with the classic approach, as well as the potential of using association analysis in building marketing strategies.

2. Complementarity of Goods in Microeconomics

The idea of complementarity was first introduced by Auspitz and Lieben in 1894 [1]. Two goods are considered to be complementary if $d^2u/dx_1dx_2 > 0$, where u is the utility function while x_1 and x_2 represent the demand for the first and the second good, respectively. A doubtless advantage of this definition is its very intuitive interpretation. It suffices to observe that du/dx_1 is the marginal utility of the first good and du/dx_2 is the marginal utility of the second good. The complementarity condition means that the marginal utility of the first good increases as more of the second good is purchased. Another advantage of Auspitz and Lieben's definition is the symmetry. If the first good is complementary to the second good, then the second good is also complementary to the first good.

However, the definition by Auspitz and Lieben has an essential disadvantage. From the utility function theory, it follows that monotonic transformations do not change the customer's preferences. Bearing the above in mind, it should be expected that monotonic transformations will not change the classification of goods, i.e., complementary goods prior to the transformation will be complementary after the monotonic transformation too. Auspitz and Lieben's definition meets this expectation but exclusively for linear transformations.

The contemporary definition of complementarity was formed as a result of many economists' work. However, the greatest contribution was from Hicks and Allen [8], hence it is most often ascribed to these two researchers. In contrast to the definition by Auspitz and Lieben, which is classified among the concepts of the so-called theory of cardinal utility, the definition by Hicks and Allen is a part of the ordinal utility theory which assumes that the measurement of utility in absolute scale is neither possible nor necessary. It suffices for the consumer to be able to determine the preference relation between the goods as well as the rate of exchange of one good for other ones.

Within the ordinal utility theory, complementary goods are defined based on correlations between demand and price. Two goods are complementary if $dx_1/dp_2 > 0$, where x_1 is the function of demand for the first good and p_2 is the price of the second good. Indifference curves for complementary goods have a characteristic hyperbolic shape. The

stronger relation of complementarity, the more the shape of the indifference curve is close to the letter L. It should be mentioned that the Allen-Hicks definition occurs in two variants, which are derived from Slutsky's theorem. The effect of the change in demand resulting from the change of a price consists of the substitution and the income effects. If the total demand is taken into account, then under Allen-Hicks's definition it is termed gross complementarity. If demand resulting solely from a change in price relations (substitution effect) is taken into account, then it is termed net complementarity. The net complementarity relation is symmetric. The specific impact of the income effect implies that the gross complementarity relation may not be symmetric. Good x may be gross complementary to good y but good y need not be gross complementary to good x. Chris De Jaegher [9] argues that the asymmetry of the gross complementarity relation need not be disadvantageous at all.

In contemporary research, the emphasis is placed on distinguishing different types of complementarity. Berry et al. [10] distinguish 7 major types:

- quantitative complementarity: when an increase in the quantity of one good leads to an increase in the value of another, e.g., right and left shoe;
- qualitative complementarity: when an increase in the quality of one good leads to an increase in the marginal value of quality of another good, e.g., a suit paired with a tie;
- within a category complementarity: when a basket of goods within the same category is selected in such manner as to best suit the customer's current needs, e.g., a home film library;
- cross-category complementarity: when goods from different categories are related with each other in order to achieve a greater value for the consumer, e.g., milk and cornflakes, or software and hardware;
- provider-driven complementarity: independent goods become complementary if they are delivered by the same provider (often within a brand or series), e.g., banking services and brokerage services;
- dynamic complementarity: substitute goods in static conditions become complementary in dynamic conditions, e.g., free vs. paid software version;
- complementarity across individual agents: if consumers (agents) interact with each other, then their choices are complementary, e.g., a Facebook friend.

It follows from the above classification that complementarity is a complex and heterogeneous relation between goods. It can be considered from the point of view of both single and multiple consumers. Also, time can play an important role (dynamic context vs. static context), i.e., some goods may become complementary in the dynamic context even though they were not so before in the static context.

3. Basic Terminology in Association Analysis

This section reviews the basic terminology used in association analysis.

Let I be the set of all available items (goods) and T be the set of all transactions. Any set $X \subseteq I$ is termed itemset. Each transaction t contains a subset of items in I. An association rule is defined as an implication of the form $X \implies Y$, where X and Y are disjoint itemsets that are respectively called the antecedent and consequent of the rule. Such a rule may be interpreted as:

If a basket contains itemset X then, with a specified probability, this basket will also contain itemset Y.

The relative number of occurrences of itemset X in the transaction set T is called the support of itemset and is calculated according to the following formula:

$$supp(X) = \frac{|t \in T : X \subseteq t|}{|T|} \qquad (1)$$

where the symbol $|\cdot|$ denotes the number of elements in a set.

J. Theor. Appl. Electron. Commer. Res. **2021**, 16

673

The level of support signifies the percentage of transactions in which specified goods were purchased at the same time in relation to all the transactions concerned.

Let *min_supp* denote the minimum level of support. The itemset with support less than min_supp will be considered as insignificant. If a support for a given itemset is greater than *min_supp*, then the set will be referred to as a frequent itemset. Association rules are created based on frequent itemsets, and the support of an association rule equals to the support of the itemsets.

Each rule is also characterized by its level of confidence, calculated according to the following formula:

$$conf(X \Longrightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}. \qquad (2)$$

Confidence indicates the probability of occurrence of the itemset, which is the consequent of the association rule in the basket if the basket also contains the itemset being the antecedent of the rule. Let *min_conf* be the minimum confidence level that a rule must achieve to be considered as significant. Rules with confidence less than *min_conf* will be rejected.

An essential feature of confidence is its asymmetry. This implies that the value of *conf*($X \Longrightarrow Y$) may, but need not, be equal to the value of *conf*($Y \Longrightarrow X$). This property will be used further on in this paper.

Another measure used to evaluate association rules is the lift [11], which is calculated according to the following formula:

$$lift(X \Longrightarrow Y) = \frac{conf(X \Longrightarrow Y)}{supp(Y)}. \qquad (3)$$

Lift measures the correlation between itemsets from antecedent and consequent of a rule. If *lift* = 1 itemsets X and Y are independent. When *lift* > 1, itemsets X and Y are positively correlated. For *lift* < 1, itemsets X and Y are negatively correlated.

The aforementioned measures of association rules are the most popular but not the only ones appearing in the literature. More examples of measures describing the rules can be found e.g., in [12].

4. The Concept of Basket Complementarity

Association analysis was developed to support investigation on relations between goods in a shopping basket. Hence the alternative name Market Basket Analysis was also developed. It was developed taking advantage of the computerization of sales processes at supermarkets, which enabled collecting data on purchases made by customers. Agrawal, Imiliński and Swami [13] made a significant contribution to the development of the discipline. The main purpose of association analysis is to discover patterns in the sets of empirical data, such as data on shopping transactions. *B*-complementarity between goods may be regarded as one of such patterns.

This section is organized as follows. First, the concept of *b*-complementarity will be formally introduced, along with all of its variants. In the next part, the most important properties of introduced concept will be presented. At the end of the section, a simple example illustrating each concept will be shown. The following types of *b*-complementarity may occur between any two goods:

I. One-sided *b*-complementarity:

Definition 1. *Any two goods x, y ∈ I are one-sided complementary if, at a given level of min_conf, the following conditions hold true: conf({x} $\Longrightarrow$ {y}) ≥ min_conf, conf({y} $\Longrightarrow$ {x}) < min_conf and supp({x, y}) > supp({x})∗supp({y}).*

One-sided *b*-complementarity concerns the cases when good *y* is usually purchased together with good *x* but good *x* is seldom purchased together with good *y*. This prompts

the conclusion that good y is purchased more frequently than good x. The condition involving the level of support is intended to ensure that items x and y are not independent, and the relations between them are not merely a matter of coincidence.

As an example of one-sided b-complementarity, one might consider the relation between cornflakes (x) and milk (y). It is not hard to observe that cornflakes are usually bought together with milk while milk is often bought regardless of whether cornflakes are bought or not.

A special example of one-sided b-complementarity is perfectly one-sided b-complementarity, which occurs when: $conf(\{x\} \implies \{y\}) \to 1$ and $conf(\{y\} \implies \{x\}) \to 0$. This is when one item (y) serves as a complement to another item (x) and may be, e.g., indispensable for its functioning. At the same time, item y is so universal that it can be used (purchased) with other items as well. Examples of items with possible relation of perfectly one-sided b-complementarity are: a remote control (x) and batteries (y), a mobile phone (x) and a charger (y).

II. *B-complementarity (two-sided b-complementarity):*

Definition 2. *Any two goods x, $y \in I$ are b-complementary (two-sided b-complementary) if, at a given level of min_conf, the following conditions hold true: $conf(\{x\} \implies \{y\}) \geq min\_conf$, $conf(\{y\} \implies \{x\}) \geq min\_conf$ and $supp(\{x, y\}) > supp(\{x\}) * supp(\{y\})$.*

B-complementarity occurs when both good x is purchased with good y and good y is purchased with good x. This is also subject to the condition that the relation between the items is not a matter of coincidence, i.e., that the items are not independent.

An example of goods with possible relation of b-complementarity might be bread and ham, where bread is usually bought with ham and, in many cases, a person buying ham also buys bread.

The relation of b-complementarity need not be a symmetric relation, i.e., the fact that good x is b-complementary to y need not imply that good y will be to the same extent b-complementary to x. If the strength of the relation between goods x and y is equal to the strength of the relation between y and x, i.e., $conf(\{x\} \implies \{y\}) = conf(\{y\} \implies \{x\})$, then this relation will be called **perfectly symmetric b-complementarity**. Because such an ideal match may seldom occur in real life, it is proposed to introduce the permissible margin of deviation between the values of b-complementarity, denoted as Mrg. The value of the Mrg factor should be interpreted as the maximum difference between the level of b-complementarity of two goods whose relation can be termed as symmetric. Thus, if a relation of b-complementarity between goods x and y satisfies the following condition: $|conf(\{x\} \implies \{y\}) - conf(\{y\} \implies \{x\})| < Mrg$, then there occurs a relation of **symmetric b-complementarity** between the goods. If the difference in the strengths of relations between goods x and y exceeds the adopted value of Mrg, then this relation will be termed asymmetric.

Figure 1 presents the relations of b-complementarity described above.

Both points A and B marked in the figure represent correlations between goods x and y and so they are positioned symmetrically with respect to the straight line denoting perfectly symmetric b-complementarity. The coordinates of the points, for example of point A, have been determined as follows:

- x-axis: $conf(\{x\} \implies \{y\})$;
- y-axis: $conf(\{y\} \implies \{x\})$.

The following part of the article presents two examples using mock data, the purpose of which is to illustrate the described definitions.
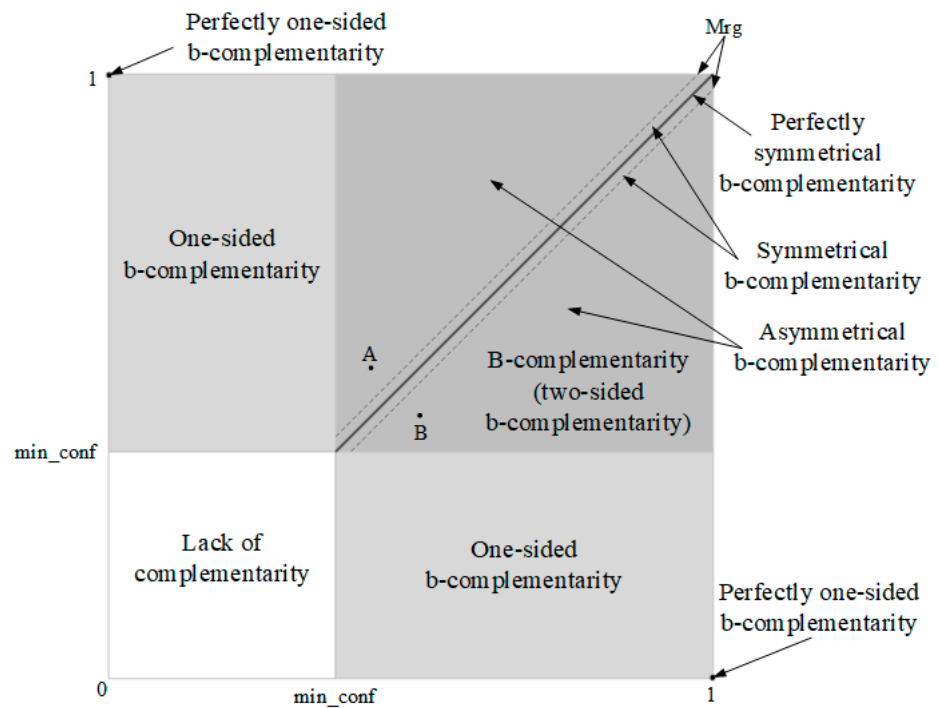
Figure 1. Types of *b*-complementarity for any pair of goods.

Example 1. *Table 1 presents a set of 15 transactions. It is assumed that the minimum level of support min_supp = 10% and min_conf = 40%.*

Table 1. Set of 15 transactions.

Transaction Number	Purchased Goods
1	Milk, Cornflakes
2	Beer
3	Milk, Cornflakes
4	Peanuts
5	Milk, Bread Roll
6	Milk, Bread Roll, Butter
7	Cornflakes
8	Milk, Bread Roll, Butter
9	Chocolate
10	Milk, Bread Roll, Butter
11	Bread Roll, Butter
12	Bread Roll, Butter
13	Butter
14	Butter
15	Mineral water

Then:

$$supp(\{Milk\}) = 40\%$$
$$supp(\{Cornflakes\}) = 20\%$$
$$supp(\{Beer\}) = 6.7\%$$
$$supp(\{Bread\ Roll\}) = 40\%$$
$$supp(\{Butter\}) = 46.7\%$$
$$supp(\{Chocolate\}) = 6.7\%$$
$$supp(\{Mineral\ water\}) = 6.7\%$$
$$supp(\{Peanuts\}) = 6.7\%$$

and:

$$conf(\{\text{Milk}\} \implies \{\text{Cornflakes}\}) = \tfrac{2}{6} = 33.3\%$$
$$conf(\{\text{Cornflakes}\} \implies \{\text{Milk}\}) = \tfrac{2}{3} = 66.7\%$$
$$\left.\begin{array}{r} supp(\{\text{Milk, Cornflakes}\}) = 13.3\% \\ supp(\{\text{Milk}\}) * supp(\{\text{Cornflakes}\}) = 8\% \end{array}\right\}$$
$$\implies supp(\{\text{Milk, Cornflakes}\})$$
$$> supp(\{\text{Milk}\}) * supp(\{\text{Cornflakes}\})$$
$$conf(\{\text{Milk}\} \implies \{\text{Bread Roll}\}) = \tfrac{4}{6} = 66.7\%$$
$$conf(\{\text{Bread Roll}\} \implies \{\text{Milk}\}) = \tfrac{4}{6} = 66.7\%$$
$$\left.\begin{array}{r} supp(\{\text{Milk, Bread Roll}\}) = 26.7\% \\ supp(\{\text{Milk}\}) * supp(\{\text{Bread Roll}\}) = 16\% \end{array}\right\}$$
$$\implies supp(\{\text{Milk, Bread Roll}\})$$
$$> supp(\{\text{Milk}\}) * supp(\{\text{Bread Roll}\})$$
$$conf(\{\text{Butter}\} \implies \{\text{Bread Roll}\}) = \tfrac{5}{7} = 71.4\%$$
$$conf(\{\text{Bread Roll}\} \implies \{\text{Butter}\}) = \tfrac{5}{6} = 83.3\%$$
$$\left.\begin{array}{r} supp(\{\text{Butter, Bread Roll}\}) = 33\% \\ supp(\{\text{Butter}\}) * supp(\{\text{Bread Roll}\}) = 18.7\% \end{array}\right\}$$
$$\implies supp(\{\text{Butter, Bread Roll}\})$$
$$> supp(\{\text{Butter}\}) * supp(\{\text{Bread Roll}\})$$
$$conf(\{\text{Butter}\} \implies \{\text{Milk}\}) = \tfrac{3}{7} = 42.9\%$$
$$conf(\{\text{Milk}\} \implies \{\text{Butter}\}) = \tfrac{3}{6} = 50\%$$
$$\left.\begin{array}{r} supp(\{\text{Butter, Milk}\}) = 20\% \\ supp(\{\text{Milk}\}) * supp(\{\text{Butter}\}) = 18.7\% \end{array}\right\} \implies supp(\{\text{Milk, Butter}\})$$
$$> supp(\{\text{Milk}\}) * supp(\{\text{Butter}\})$$

Based on the calculations presented above, it can be concluded that:

- Cornflakes are complementary to Milk and this is one-sided *b*-complementarity;
- Milk is complementary to a Bread Roll and this is symmetric *b*-complementarity;
- Butter is complementary to a Bread Roll and this is asymmetric *b*-complementarity;
- Butter is complementary to Milk and this is asymmetric *b*-complementarity.

The relations described are presented in Figure 2. The items in the figure are marked as follows:

- Bread Roll—BR,
- Milk—Ml,
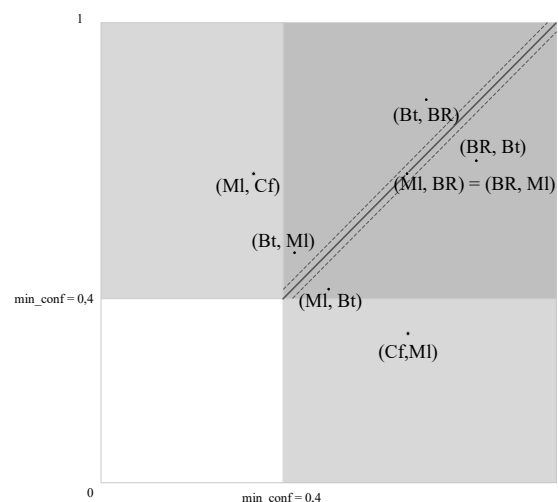- Cornflakes—Cf,
- Butter—Bt.

Figure 2. *B*-complementarity of goods based on Table 1.

For three goods, there may occur a relationship of **weak *b*-complementarity**. Weak *b*-complementarity should be understood as a relation of *b*-complementarity between goods being the result of the occurrence of a relation between other goods. Formally, a weak *b*-complementarity is defined as follows:

Definition 3. *Three goods are provided: x, y, z ∈ I. Goods x and y are one- or two-sided b-complementary. Between goods y and z, there is also a relation of b-complementarity (two-sided or one-sided). If between goods x and z there is a relation of b-complementarity (two-sided or one-sided) and at least one of the following conditions is met:* $conf(\{x\} \Longrightarrow \{z\}) \leq conf(\{x\} \Longrightarrow \{y\})*conf(\{y\} \Longrightarrow \{z\})$ *or* $conf(\{z\} \Longrightarrow \{x\}) \leq conf(\{z\} \Longrightarrow \{y\})*conf(\{y\} \Longrightarrow \{x\})$, *then the relation between goods x and z will be called **weak b-complementarity**.*

Weak *b*-complementarity may occur when two goods, *x* and *y*, are strongly *b*-complementary to each other and one of them, e.g., *y*, is strongly *b*-complementary to z. In such a case, even though there is a relation of *b*-complementarity between goods *x* and *z*, this relation may result from the relation between goods *x* and *y* or between *y* and *z*. Therefore, such a relation is termed weak *b*-complementarity.

Example 2. *Utilizing the data from Example 1 (see Table 1), the following calculations have been carried out:*

$$\left. \begin{array}{c} conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\}) * conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\}) = \frac{4}{6} * \frac{5}{6} = \frac{20}{36} \\ conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\}) = \frac{3}{6} = \frac{18}{36} \end{array} \right\}$$
$$\Longrightarrow conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\})$$
$$\leq conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\})$$
$$*conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\})$$

$$\left. \begin{array}{c} conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\}) * conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Milk}\}) = \frac{5}{7} * \frac{4}{6} = \frac{20}{42} \\ conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\}) = \frac{3}{7} = \frac{18}{42} \end{array} \right\}$$
$$\Longrightarrow conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\})$$
$$\leq conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\})$$
$$*conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Milk}\})$$

$$\left. \begin{array}{c} conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\}) * conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\}) = \frac{3}{7} * \frac{4}{6} = \frac{12}{42} \\ conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\}) = \frac{5}{7} = \frac{30}{42} \end{array} \right\}$$
$$\Longrightarrow conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\})$$
$$> conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\})$$
$$*conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\})$$

$$\left. \begin{array}{c} conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Milk}\}) * conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\}) = \frac{4}{6} * \frac{3}{6} = \frac{12}{36} \\ conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\}) = \frac{5}{6} = \frac{30}{36} \end{array} \right\}$$
$$\Longrightarrow conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\})$$
$$\leq conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Milk}\})$$
$$*conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\})$$

$$\left. \begin{array}{c} conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\}) * conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\}) = \frac{3}{6} * \frac{5}{7} = \frac{15}{42} \\ conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\}) = \frac{4}{6} = \frac{28}{42} \end{array} \right\}$$
$$\Longrightarrow conf(\{\text{Milk}\} \Longrightarrow \{\text{Bread Roll}\})$$
$$> conf(\{\text{Milk}\} \Longrightarrow \{\text{Butter}\})$$
$$*conf(\{\text{Butter}\} \Longrightarrow \{\text{Bread Roll}\})$$

$$\left. \begin{array}{c} conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\}) * conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\}) = \frac{5}{6} * \frac{3}{7} = \frac{15}{42} \\ conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Milk}\}) = \frac{4}{6} = \frac{28}{42} \end{array} \right\}$$
$$\Longrightarrow conf(\text{Bread Roll} \Longrightarrow \text{Milk})$$
$$> conf(\{\text{Bread Roll}\} \Longrightarrow \{\text{Butter}\})$$
$$*conf(\{\text{Butter}\} \Longrightarrow \{\text{Milk}\})$$

Based on the obtained results, it can be concluded that the relation of *b*-complementarity between Milk and Butter is a weak relation. A high level of *b*-complementarity between

J. Theor. Appl. Electron. Commer. Res. **2021**, *16*

678

these goods results from the very strong relation between Bread Roll and Butter, as well as the strong relation between Milk and Bread Roll.

5. Empirical Example

The presented concept has been verified on empirical data from "The Instacart Online Grocery Shopping Dataset 2017" [14]. The database has 49,677 items and 3,214,874 transactions. It was imported to Microsoft SQL Server 2017. The number of occurrences of each product was determined using a query that counts the number of transactions in which a given product was purchased. Table 2 shows a summary of how many items occurred in a given number of transactions (defined by ranges). The average number of transactions for the items is 653 with a standard deviation of 4792.

Table 2. The number of items depending on the range of number of transactions.

Range of Number of Transactions	Number of Items
[1, 9]	7165
[10, 31]	11,343
[32, 99]	11,102
[100, 499]	11,777
≥500	8,290

As shown in Table 2, it may be concluded that there are 7165 items that have been bought fewer than 10 times, which makes up to 14.5% of all the items. Conversely, 16.7% (8290) items were bought 500 times at the minimum. Over two thirds of all the items were bought either more than 10 times or less than 500 times. The described dependencies are presented in Figure 3.
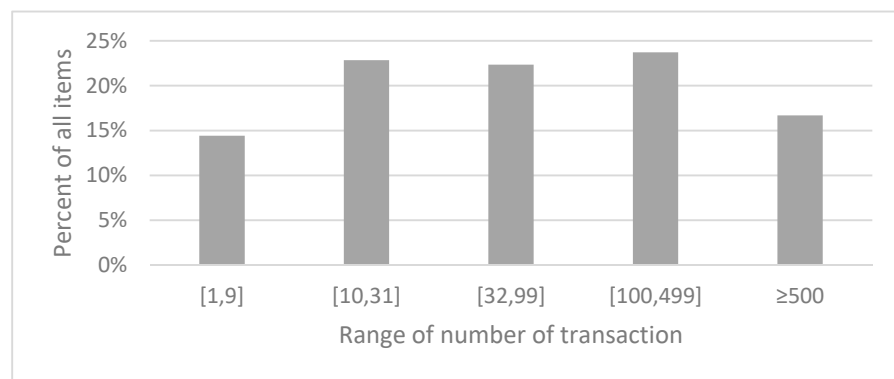

Figure 3. The percentage share of the number of items from a given interval of transactions number in all item numbers.

A minimum support level was set to 0.001%, the value of which corresponds to a minimum number of 32 transactions. The number of items meeting this condition was 31,169. On this basis, 961,000 two-element itemsets were determined that met the condition of a minimum level of support. After rejecting itemsets that did not meet the second requirement $(supp(\{x, y\}) > supp(\{x\}) * supp(\{y\}))$ regarding item independence, 877,000 two-element itemsets remained.

Figure 4 presents the visualization of confidence levels between the items ({x, y}) that forming a given itemset. In order to increase transparency, each pair of items was marked as exactly one point—in contrast to Figure 2, where each pair of items corresponds to two points located symmetrically to the straight line, indicating perfectly symmetric *b*-complementarity.
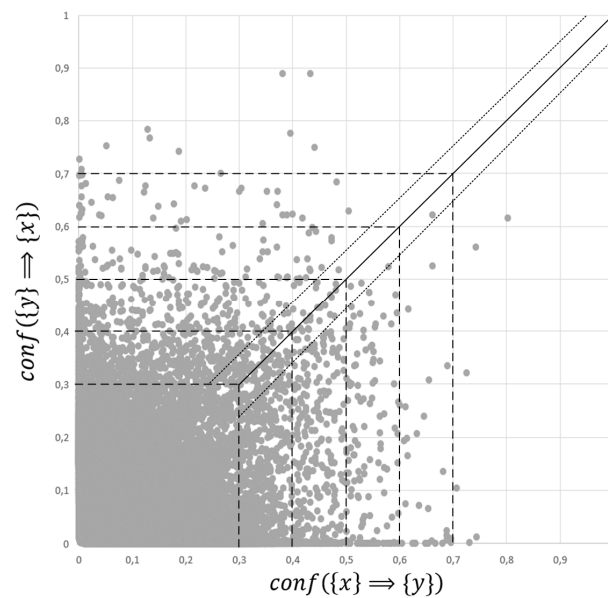
J. Theor. Appl. Electron. Commer. Res. **2021**, 16

679

Figure 4. *B*-complementarity of goods based on "The Instacart Online Grocery Shopping Dataset 2017".

Vertical and horizontal dashed lines indicate support values, which correspond to the Min_conf value from Table 3. The continuous line indicates perfect symmetry, and the diagonal dotted line—the Mrg coefficient, which is used to determine symmetric *b*-complementarity—equals 5%.

Table 3. The number of two-element itemsets of *b*-complementary goods depending on the level of *min_conf* and *Mrg* parameters.

Min_conf	One-Sided *B*-Complementarity	Two-Sided *B*-Complementarity	Symmetric *B*-Complementarity (*Mrg*)		
			0%	1%	5%
30%	3403	278	5	31	107
40%	957	70	3	10	27
50%	310	7	1	1	2
60%	90	3	1	1	2
70%	18	1	1	1	1

Table 3 shows the number of detected one- and two-sided *b*-complementarities depending on the assumed minimum confidence level. The number of cases of perfectly symmetric *b*-complementarity, which depends on the assumed value of the *Mrg* and *min_conf* parameters, is also presented.

Based on the data from Table 3, it can be concluded that one-sided *b*-complementarity between goods occurs much more often than two-sided *b*-complementarity. For symmetric *b*-complementarities, depending on the value of the *Mrg* parameter, the number of occurrences ranges from several (perfectly symmetric *b*-complementarity for the *Mrg* = 0%) to over 100 times. It is worth noting that regardless of the level of minimal support, at least 1/4 of two-sided *b*-complementarities are symmetric for Mrg when it is equal to 5%.

There were no perfectly one-sided *b*-complementarity relationships in the data set under consideration. The strongest one-sided relations $conf(\{x\} \implies \{y\}) > 70\%$ and $conf(\{y\} \implies \{x\}) < 1\%$ occurred in the case of 3 itemsets.

The highest confidence levels for the detected weak *b*-complementarity relationships ranged from 10% to 15% (3 relationships including 1 two-sided relationship) and were therefore rejected due to the low confidence level.

The analysis of sales data leads to an interesting conclusion regarding the relationship between the number of one and two-sided complementary goods. It turns out that the

number of cases of one-sided *b*-complementarity (for all analyzed levels of minimal support) is an order of magnitude greater than the number of cases of two-sided *b*-complementarity. Such an observation may be the basis for reconsidering the assumptions underlying the classic definition of complementarity between goods.

6. Conclusions

Within the theory of supply and demand, complementarity is defined on the basis of relations between price and demand. Goods are deemed to be complementary if the cross-price elasticity of demand is negative. Estimation of cross-price elasticity of demand requires data on demand at different prices, ceteris paribus. Data of this kind are typically difficult to access, their quantity is largely insufficient and the ceteris paribus assumption is in clear contradiction to the size of the sample. The definition of *b*-complementarity introduced in the paper makes use of data on transactions made by customers and does not require information on the prices, although such information can be taken into account as well in future research. The accessibility of this kind of data (e.g., in the e-commerce industry) is incomparably greater than price and demand data, which means that opportunities for automatic identification of complementary goods are greater too.

The definition is based on the assumption that if goods *b*-complement each other, then they will be more frequently purchased together. Thus, goods deemed to be *b*-complementary are goods for which the frequency of occurrence in one basket is at an appropriately high level. There is no doubt that this is compliant with the intuitive understanding of complementarity in microeconomics. At the present stage of the investigation, it is still unknown whether there is a closer relation between the *b*-complementarity as introduced by the authors and complementarity in the sense of the definitions provided by Auspitz and Lieben [1] or by Hicks and Allen [8]. Both the classic definitions and the one introduced in the paper describe the same kind of relation between goods. It can be suspected that for the majority of goods, the same results will be obtained, i.e., complementary goods in the sense of the classic definitions will also be *b*-complementary.

When analyzing the relation of *b*-complementarity, it is easy to notice that there are goods for which this relation occurs in both directions, but there are some where the relation is one-sided. The definition introduced in the paper distinguishes between these types of *b*-complementarity relations by introducing the concepts of one- and two-sided *b*-complementarity. Examples provided in the paper indicate that one-sided relations of *b*-complementarity are not a rare phenomenon. Neglecting them or classifying them in the same category as two-sided relations is a major simplification; in some situations, it can even be an error. Furthermore, the introduced definition of *b*-complementarity makes it possible to distinguish if a relation of *b*-complementarity between goods being observed exists regardless of other goods or is a consequence of other relations. The proposed method is based on a basket analysis, which requires the digitization of a company's transaction data. Too short a period of analyzed data (not taking into account seasonality) may distort the obtained results. Another problem may be the use of alternative sources of supply by customers and the sharing of orders between different suppliers. The measures of complementarity proposed in the article require that the goods under analysis should be purchased within one transaction.

The main objective of this paper was an attempt at defining the concept of *b*-complementarity. The paper focused on the formal aspects of all definition of *b*-complementarity. Nevertheless, it turned out to be possible to prepare a representative example explaining the introduced concepts. An important part of this paper was empirical studies on a large sample of sales data. These empirical studies showed the presence of all types of *b*-complementarity between goods. It also turned out that the number of cases of one-sided *b*-complementarity is by an order of magnitude greater than the number of two-sided *b*-complementarity cases. The results of empirical research are so promising that it is worth considering developing a 'basket' definition of substitution goods. An interesting direction of development also appears in the field of recommendation systems.

The concept of complementarity and substitutability of products could be the basis for the construction of many recommendation systems [5–7,15–18]. There is no obstacle preventing the application of *b*-complementarity for making recommendations in such a system.

The concept of *b*-complementarity can be used to measure all of the types of complementarity introduced by Berry, et al. [10]. Contrary to the known measures of complementarity created for recommendation systems [5–7,15–18], it distinguishes between many different aspects of complementarity. It is also better focused on measuring complementarity than measures of complementarity from recommendation systems.

Author Contributions: Conceptualization, R.P. and S.J.; methodology, R.P. and S.J.; formal analysis, R.P. and S.J.; investigation, R.P. and S.J.; resources, R.P. and S.J.; writing—original draft preparation, R.P. and S.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Auspitz, R.; Lieben, R. *Untersuchungen Über Die Theorie des Preises*; Duncker & Humblot: Berlin, Germany, 1889. [CrossRef]
2. Lenfant, J.S. In search of complementarity. In Proceedings of the 2003 History of Economics Society annual meeting, Durham, NC, USA, 4–7 July 2003.
3. Lenfant, J.S. Complementarity and Demand Theory: From the 1920s to the 1940s. *Hist. Political Econ.* **2006**, *38* (Suppl. 1), 48–85. [CrossRef]
4. Aguinis, H.; Forcum, L.E.; Joo, H. Using market basket analysis in management research. *J. Manag.* **2013**, *39*, 1799–1824. [CrossRef]
5. McAuley, J.; Pandey, R.; Leskovec, J. Inferring Networks of Substitutable and Complementary Products. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '15, Sydney, Australia, 10–13 August 2015.
6. Moosavi, S.; Nematbakhsh, M.; Farsani, H.K. A semantic complement to enhance electronic market. *Expert Syst. Appl.* **2009**, *36*, 5768–5774. [CrossRef]
7. Khalaji, M.; Mansouri, K.; Mirabedini, S.J. Improving Recommender Systems in E-Commerce Using Similar Goods. *J. Softw. Eng. Appl.* **2012**, *5*, 96–101. [CrossRef]
8. Hicks, J.R.; Allen, R.G.D. A Reconsideration of the Theory of Value. Part II. A Mathematical Theory of Individual Demand Functions. *Economica* **1934**, *1*, 196. [CrossRef]
9. De Jaegher, K. Asymmetric substitutability: Theory and some applications. *Econ. Inq.* **2009**, *47*, 838–855. [CrossRef]
10. Berry, S.; Khwaja, A.; Kumar, V.; Musalem, A.; Wilbur, K.C.; Allenby, G.; Anand, B.N.; Chintagunta, P.K.; Hanemann, W.M.; Jeziorski, P. Structural models of complementary choices. *Mark. Lett.* **2014**, *25*, 245–256. [CrossRef]
11. Brin, S.; Motwani, R.; Silverstein, C. Beyond market baskets. *ACM Sigmod Rec.* **1997**, *26*, 265–276. [CrossRef]
12. Lenca, P.; Meyer, P.; Vaillant, B.; Lallich, S. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *Eur. J. Oper. Res.* **2008**, *184*, 610–626. [CrossRef]
13. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. *ACM Sigmod Rec.* **1993**, *22*, 207–216. [CrossRef]
14. The Instacart Online Grocery Shopping Dataset. 2017. Available online: https://www.instacart.com/datasets/grocery-shopping-2017 (accessed on 11 November 2020).
15. Zhao, T.; McAuley, J.; Li, M.; King, I. Improving recommendation accuracy using networks of substitutable and complementary products. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3649–3655.
16. Yu, H.; Litchfield, L.; Kernreiter, T.; Jolly, S.; Hempstalk, K. Complementary Recommendations: A Brief Survey. In Proceedings of the International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Shenzhen, China, 9–11 May 2019; pp. 73–78.
17. Zhang, M.; Bockstedt, J. Complements and substitutes in product recommendations: The differential effects on consumers' willingness-to-pay. *CEUR Workshop Proc.* **2016**, *1679*, 36–43.
18. Osadchiy, T.; Poliakov, I.; Olivier, P.; Rowland, M.; Foster, E. Recommender system based on pairwise association rules. *Expert Syst. Appl.* **2019**, *115*, 535–542. [CrossRef]