Chapter-3

Database, Methodology and Construction of Variables

To attain the set objectives and test the related hypotheses, different mathematical and econometric models have been used under a structured framework. The present study is based upon the use of both primary and secondary data sources. The analysis of bankers’ and customers’ perspective on bancassurance is important only if the banks are having substantial gains from bancassurance. If bancassurance doesn’t offer a significant improvement in the performance of Indian commercial banking then said analysis carry no worth. Therefore, before the analysis of the bankers’ and customers’ perspective on bancassurance, the gains in performance of sampled banks have been worked out in the form of efficiency improvements. The efficiency analysis has been performed using secondary data published by the sampled banks in their Annual Reports. However, in remaining analysis, the primary data has been collected using well-structured questionnaires. In order to evaluate extent of awareness regarding bancassurance amongst bank customers, factors affecting level of awareness, and overall satisfaction from bancassurance, the Ordered Probit model has been used. However, the technique of Structured Equations Modeling (SEM) has been utilized to analyze factors affecting the choice of product among bank customers; the drivers of cross buying intention amongst bank customers, sources of customers’ satisfaction, and sources of motivation for bank employees to initiate bancassurance.

To explain the methodology used in detail, the chapter has been divided into four sections. Section-3.1 offers the discussion over data collection and classification. The next Section-3.2 provides the sample characteristicson the basis of demographic and model categories such as City, Gender, Education, Age, etc. Section-3.3 provides the methodology used to test the set hypotheses. A summary of the different econometric methods namely, Data Envelopment Analysis (DEA), OrderedProbit model, and Structure Equations Modeling (SEM) is the feature of the said section. The last section concludes the chapter.

3.1. Data Collection

The data has been collected from three cities of Punjab; Amritsar, Jalandhar and Ludhiana. The population of Punjab is geographically divided in three regions; Majha, Doaba and Malwa. From each region, one city has been selected and the choice of cities is
convenience based. However, the banks have been selected following cluster based sampling. The Indian banking sector is dominated by the two major categories of commercial banks namely, public and private banks. Further, all these banks are either following Agency or Joint Venture (JV) model of bancassurance. Under Agency category, banks are offering agency services to an insurance company, whereas, under JV model they are selling insurance as diversified product. The banks have been selected in such a manner so as to ensure 10 banks from each sector and model. The banks reporting highest average bancassurance income have been selected in each category defined. Table 3.1 represents the category and model of bancassurance followed by each sampled bank.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Bank</th>
<th>Category of Bank</th>
<th>Model of Bancassurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>State Bank of India</td>
<td>Public Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>2</td>
<td>Canara Bank</td>
<td>Public Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>3</td>
<td>Oriental Bank of Commerce</td>
<td>Public Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>4</td>
<td>Bank of India</td>
<td>Public Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>5</td>
<td>Union Bank of India</td>
<td>Public Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>6</td>
<td>State Bank of Hyderabad</td>
<td>Public Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>7</td>
<td>United Bank</td>
<td>Public Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>8</td>
<td>Punjab and Sind Bank</td>
<td>Public Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>9</td>
<td>State Bank of Bikaner and Jaipur</td>
<td>Public Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>10</td>
<td>Indian Bank</td>
<td>Public Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>11</td>
<td>Federal Bank</td>
<td>Private Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>12</td>
<td>Jammu and Kashmir Bank</td>
<td>Private Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>13</td>
<td>Karnataka Bank</td>
<td>Private Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>14</td>
<td>ICICI</td>
<td>Private Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>15</td>
<td>HDFC</td>
<td>Private Sector Bank</td>
<td>Joint Venture</td>
</tr>
<tr>
<td>16</td>
<td>Kotak Mahindra Bank</td>
<td>Private Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>17</td>
<td>Yes Bank</td>
<td>Private Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>18</td>
<td>Axis Bank</td>
<td>Private Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>19</td>
<td>ING Vyasya</td>
<td>Private Sector Bank</td>
<td>Agency</td>
</tr>
<tr>
<td>20</td>
<td>Indusind Bank</td>
<td>Private Sector Bank</td>
<td>Agency</td>
</tr>
</tbody>
</table>

**Source:** Author’s Elaborations
### Table 3.2: Obtaining Number of Branches to be Sampled from Each Bank out of Three Cities Using Maximum Entropy Principle

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Bank</th>
<th>Observed Frequency</th>
<th>Expected Frequency for 1195</th>
<th>Expected Frequency for 200</th>
<th>Adjusted Expected Frequency</th>
<th>Column Total (Cj)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bank</td>
<td>ASR JAL LDH R_i</td>
<td>ASR JAL LDH R_i</td>
<td>ASR JAL LDH R_i</td>
<td>ASR JAL LDH R_i</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>SBI</td>
<td>55 70 74 199</td>
<td>60 62 77 199</td>
<td>10 10 13 33</td>
<td>10 10 13 33</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Canara Bank</td>
<td>19 39 22 80</td>
<td>24 25 31 80</td>
<td>4 4 5 13</td>
<td>4 4 5 13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>OBC</td>
<td>39 41 55 135</td>
<td>41 42 53 135</td>
<td>7 7 9 23</td>
<td>7 7 9 23</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Bank of India</td>
<td>21 19 45 85</td>
<td>26 26 33 85</td>
<td>4 4 6 14</td>
<td>4 4 6 14</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>UBI</td>
<td>14 19 32 65</td>
<td>20 20 25 65</td>
<td>3 3 4 11</td>
<td>3 3 4 11</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>SB Hyderabad</td>
<td>1 1 2 4</td>
<td>1 1 2 4</td>
<td>0 0 0 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>United Bank of India</td>
<td>1 3 3 7</td>
<td>2 2 3 7</td>
<td>0 0 0 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Punjab and Sind Bank</td>
<td>86 47 67 200</td>
<td>60 62 78 200</td>
<td>10 10 13 33</td>
<td>10 10 13 33</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>SB of Bikaner &amp; Jaipur</td>
<td>1 2 5 8</td>
<td>2 2 3 8</td>
<td>0 0 1 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Indian Bank</td>
<td>1 10 12 23</td>
<td>7 7 9 23</td>
<td>1 1 1 4</td>
<td>1 1 1 4</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>The Federal Bank</td>
<td>1 2 7 10</td>
<td>3 3 4 10</td>
<td>1 1 2 1</td>
<td>1 1 2 1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The J&amp;K Bank</td>
<td>2 1 3 6</td>
<td>2 2 2 6</td>
<td>0 0 0 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Karnataka bank</td>
<td>1 1 1 3</td>
<td>1 1 1 3</td>
<td>0 0 0 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ICICI Bank</td>
<td>13 16 28 57</td>
<td>17 18 22 57</td>
<td>3 3 4 10</td>
<td>3 3 4 10</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>HDFC Bank</td>
<td>66 51 51 168</td>
<td>50 52 65 168</td>
<td>8 9 11 28</td>
<td>8 9 11 28</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Kotak</td>
<td>1 3 3 7</td>
<td>2 2 3 7</td>
<td>0 0 0 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Yes Bank</td>
<td>5 17 5 27</td>
<td>8 8 11 27</td>
<td>1 1 2 5</td>
<td>1 1 2 5</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Axis Bank</td>
<td>28 24 39 91</td>
<td>27 28 35 91</td>
<td>5 5 6 15</td>
<td>5 5 6 15</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>ING Vyasya</td>
<td>2 1 5 8</td>
<td>2 2 3 8</td>
<td>0 0 1 1</td>
<td>1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Indusind Bank</td>
<td>2 4 6 12</td>
<td>4 4 5 12</td>
<td>1 1 1 2</td>
<td>1 1 1 2</td>
<td></td>
</tr>
<tr>
<td><strong>Column Total (Cj)</strong></td>
<td><strong>359 371 465</strong></td>
<td><strong>G=1195</strong></td>
<td><strong>359 371 465 1195</strong></td>
<td><strong>60 62 78 200</strong></td>
<td><strong>65 67 81 213</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The Source of Column (2) to (4) is [http://bankifsccode.com/](http://bankifsccode.com/) and remaining columns are own calculations.
It is evident from the table that 20 major commercial banks (10 from each sector and 5 of each sector follow one of two models) have been selected who are offering bancassurance services with significant share. The size of the sample is taken 10 times the number of banks i.e., 200 bank branches for survey (Table 3.2 represents the number of bank branches selected from each city for study).

However, selection of the number of bank branches to be surveyed from each target cities and bank was challenging task and some scientific method was needed to overcome the same challenge. To confront the said challenge, the maximum entropy principle has been used. As per this principle, following steps are used to get the number of branches of each bank to be surveyed from each of three Cities:

1) Calculate the expected frequency of each cell in a bank and cities cross-tabs using the formula:

\[ E_{ij} = \frac{RowTotal(R_i) \times ColumnTotal(C_j)}{GrandTotal(G)} \]

2) Calculate proportion of each expected frequency \( E_{ij} \) to grand total \( G \) i.e., \( P_{ij} = \frac{E_{ij}}{G} \);

3) Multiply each \( P_{ij} \) with desired grand total \( DG \) i.e., 200 banks.

All these steps can be explained using the distribution of each bank branches in three sampled cities given in Table 3.2. Column (7) to (10) of the table represents expected frequencies computed following step (1). The step (2) has not been shown in Table 3.2 for formatting reasons, whereas, the calculation of step (3) have been given in (11) to (14) columns. The distribution of branches given in these columns is for 200 target sample size. However, some cell frequencies are observed to be zero i.e., none of the branch of that \( i^{th} \)bank and \( j^{th} \)cities will be selected for the sample. Thus, these types of cells have been adjusted with one instead of taking zero frequency. Hence the overall adjusted sample size is 213 bank branches. In sum, the study is based upon the survey of customers’ and bank employees of these 213 bank branches.

3.2 Sample Characteristics

Two types of questionnaires have been designed to know about the perspective of customers and bankers for bancassurance; one for customers and other for bankers. The data on customers’ response have been collected from 213 bank branches. From each branch, 3 customers had been chosen so as to get a response of 639 customers (see Questionnaire in
Annexure-I. The 39 customers didn’t return back the questionnaires, whereas, 49 customer questionnaires found to be incomplete while feeding the data. Thus, 49 questionnaires discarded in anticipation of data redundancy error. Thus, the number of customer responses recorded for the analysis is 551. However, none of the banker questionnaires is incomplete because the interview method has been used to collect the requisite information. Thus, 184 responses of specified bank employees selling insurance have been used for analyzing the bankers’ perspective on bancassurance. The characteristics of both types of responses have been discussed as follows:

### 3.2.1 Customers’ Response

The demographic characteristics examined using frequency tables represents that the responses of 217 female respondents and 334 male respondents have been used. The frequency count of model category represents that among 551 customer respondents 285 belong to banks following JV model of bancassurance while 266 belong to agency category banks. Out of total count, 333 customers have purchased life insurance, whereas, 218 have purchased general insurance. Furthermore, the insurance products are sold by the banks under two categories; standalone and tied-up categories. The tied-up category products are sold by the bank jointly with other bank products e.g., saving account holders are given accidental insurance by SBI. Another category i.e., standalone consist of the insurance products sold individually without the combination of any bank product. The frequency count of each category of insurance in Table 3.3 delineates that 285 respondents have purchased only standalone product, whereas, 244 have purchased only tied-up insurance. However, 22 customers have purchased both standalone and tied-up insurance products.

<table>
<thead>
<tr>
<th>Type of Insurance Product</th>
<th>Tied-up Insurance</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Standalone Insurance</td>
<td>0</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>244</td>
<td>307</td>
</tr>
<tr>
<td>Total</td>
<td>285</td>
<td>266</td>
</tr>
</tbody>
</table>

**Source:** Author’s Calculations

Further, Table 3.4 provides model wise demographic distribution of the respondents. The four attributes Gender, Age, Education, and Income have been tabulated against two
models of bancassurance namely, Joint Venture (JV) and Agency. On Gender base, it may be observed that sample is dominated by the male respondents and the proportion of female respondents is lesser than the male ones. In age distribution, the highest proportion is of 26-40 age group people in both of the models. The education of maximum respondent is graduate and a huge percentage is of post-graduates too. These groups of people are assumed to have a significant level of awareness regarding the insurance product they are purchasing from bank. Furthermore, the sample is dominated by the people living in income groups 10000-30000 and 30000-50000. These categories of customers have sufficient funds to save for and invest in diverse financial assets like bonds, insurance etc.

| Tie-up Type | Gender | Age 18-25 | Age 26-40 | Age 41-50 | Age 51+ | Total | Education Under-Graduate | Graduates | Post-Graduates | Others | Total | Income <10000 | Income 10000-30000 | Income 30000-50000 | Income >50000 | Total |
|-------------|--------|-----------|-----------|-----------|---------|-------|------------------------|-----------|-----------------|--------|-------|--------------|----------------|----------------|----------------|---------|-------|
| JV          | Female | 104       | 181       | 285       | 90      | 128   | 39                     | 39        | 26              | 5      | 8     | 15           | 285            |                |                | 35      | 269   |
| Agency      | Female | 110       | 156       | 266       | 74      | 116   | 36                     | 40        | 206             | 55     | 112   | 99           | 206            |                |                | 47      | 304   |
| Total       | Female | 214       | 337       | 551       | 164     | 244   | 75                     | 68        | 551             | 106    | 233   | 167          | 551            |                |                | 47      | 563   |

Source: Author’s Elaborations

3.2.2 Bankers’ Response

The analysis of bankers’ perspective is based upon the data collected from the executives of same 213 bank branches through questionnaire (see Questionnaire in Annexure-II). The questionnaire was distributed to 213 bank branches in three sampled cities namely, Ludhiana, Jalandhar and Amritsar in the year 2012. However, only 184 executives could respond adequately to all queries in the questionnaire. The 29 executives could not return the questionnaire back because of their busy schedule. Thus, the analysis of bankers’ perspective is based on the response of 184 respondents.

3.3 Methodology Used

To achieve the set objectives and test the set hypotheses, different mathematical and econometrics techniques such as Data Envelopment Analysis (DEA), Ordered Probit Analysis (Oprobit), and Structure Equations Modeling (SEM) have been used. The DEA has been used to evaluate the performance of banks without and with bancassurance income, whereas, the Probit model has been applied to test the factors affecting i) level of awareness regarding bancassurance amongst bank customers; and ii) factors affecting overall satisfaction from bancassurance among bank customers. However, the SEM has been utilized to model i)
factors affecting the choice of product among bank customers; ii) the drivers of cross buying intention amongst the bank customers; and iii) the sources of satisfaction from bancassurance. All of these techniques have been explained as follows.

3.3.1 Data Envelopment Analysis and Efficiency Evaluation

Data envelopment analysis (DEA) is a data-oriented approach for evaluating the performance of a set of peer entities called decision making units (DMUs) whose performance is characterized by multiple measures/indicators (Zhu, 2003). The technique has been explained in brief as follows.

3.3.1.1 Technical and Scale Efficiency in DEA: A Theoretical Exposition

The technique of DEA calibrates the level of technical efficiency on the basis of estimated discrete piecewise frontier (or so called *best-practice* or *efficient frontier* or *envelopment surface*) made up by a set of Pareto-efficient DMUs. In all instances, these Pareto-efficient DMUs located on the *efficient frontier*, compared to the others, minimize the use of productive resources given the outputs (input-oriented measure), or maximize the outputs given the inputs (output-oriented measure) and called the *best-practice performers* or *reference units* or *peer units* within the sample of DMUs. These Pareto-efficient DMUs have a benchmark efficiency score of 1 that no individual DMU’s score can surpass. Further, this *efficient frontier* provides a yardstick against which to measure the relative efficiency of all other DMUs that do not lie on the frontier. The DMUs which do not lie on *efficient frontier* are deemed to be relatively inefficient (i.e., Pareto non-optimal DMUs) and receives a score between 0 and 1. The efficiency score of each DMU can be interpreted as the radial distance to *efficient frontier*. In short, DEA forms a non-parametric surface frontier (more formally a piece-wise-linear convex isoquant) over the data points to determine the efficiency of each DMU relative to this frontier.

A graphical conceptualization of the DEA technique in the output-oriented framework is represented in Figure 3.1 where the *efficient frontier* drawn refers to an elementary case with two outputs (\(y_1\) and \(y_2\)) and one input (\(x\)) under the assumption of constant returns-to-scale (CRS). Suppose there are three DMUs say \(A\), \(B\) and \(C\). Further assume that DMUs \(A\) and \(B\) are efficient. Thus, they are located on the boundary of *best-practice frontier*. Further, DMU \(C\) employs the same quantities of the inputs as used by DMUs \(A\)and\(B\) but produces less of both of the outputs.
Hence, DMU C is not efficient as DMUs A and B. If we measure the deviation of DMU C from the best-practice frontier in a radial way, its relative technical efficiency is given by $0C/0K$, which can also be thought of as a ratio of actual (observed) output (at C) to maximum potential output (at K). This measure is referred as the Farrell’s output-oriented measure of technical efficiency.

The above conceptualization of the concept of technical efficiency is based upon the assumption of constant returns-to-scale (CRS). In DEA literature, the measure of technical efficiency corresponding to CRS assumption is generally referred as overall technical efficiency (OTE) which captures the efficiency due to both managerial and scale effects. The CRS assumption is only appropriate if all DMUs are operating at an optimal scale. When DMUs are not operating at optimal scale (i.e., variable returns-to-scale (VRS) prevails), the overall technical efficiency (OTE) can be decomposed into pure technical efficiency (PTE) and scale efficiency (SE). The PTE is a measure of managerial efficiency i.e., the capability of the management to convert the inputs into outputs. However, the SE measure indicates whether the DMU in question is operating at optimal scale size or not (Kumar and Arora, 2008).

Figure 3.2 illustrates the decomposition of OTE into PTE and SE in envelopment surface in single-input and single-output space. As shown, the envelopment surfaces may be either linear as in CRS case, or convex as in case with VRS. The CRS and VRS cases are
detailed: the CRS surface is the straight line $0BN$ and the VRS surface is $GABCE$. For ease of exposition the interior (or inefficient) DMU is represented by point $D$. Now the technical efficiency of any interior point (such as $D$) is intuitively given by the distance between envelope and itself.

Using an output-orientation, the technical efficiency at point $D$ would be given by $PD/PN$ in the CRS case, $PD/PM$ in the VRS case, and the scale efficiency would be $PM/PN$. Finally for the DMU on the envelopment surface, such as denoted by $B$, the technical efficiency measure for both VRS and CRS would be identical as DMU $B$ is found to be operating at CRS as well as VRS frontier.

![Figure 3.2: Decomposition of Overall Technical Efficiency into Pure-Technical and Scale Efficiencies](image)

**Source:** Kumar and Arora (2007)

### 3.3.1.2 CCR and BCC Models

Several different mathematical programming models have been proposed in the literature [see, Charnes et al. (1994) and Cooper, Seiford and Tone (2007)]. Essentially, each of these models seeks to establish which of $n$ DMUs determine the best-practice frontier or efficient frontier. The geometry of this surface is prescribed by the specific DEA model employed. In the present study, we utilized CCR model, named after Charnes, Cooper and Rhodes (1978) and BCC model, named after Banker, Charnes and Cooper (1984) to obtain efficiency measures under CRS and VRS assumptions, respectively.
To illustrate the CCR and BCC models, consider \( n \) DMUs, \( j=1,...,n \). The units are homogeneous with the same types of inputs and outputs. Assume there are \( m \) inputs and \( k \) outputs. Let \( x_j \) and \( y_j \) denote, respectively, the input and output vectors for the \( j \)-th DMU. Thus, \( x_j \) is a \((m \times 1)\) column vector and \( y_j \) is a \((k \times 1)\) column vector. Moreover, \( X=(x_1,x_2,...,x_n) \) is the \( m \times n \) input matrix and \( Y=(y_1,y_2,...,y_n) \) is the \( k \times n \) output matrix.

The CCR model assigns weights to each input and output, and then assesses the efficiency of a given DMU by the ratio of the aggregate weighted output to the aggregate weighted input. The weights assigned must be non-negative. Also, they must restrict each DMU from receiving a ratio (of the weighted output to the weighted input) that is greater than 1. Mathematically, when evaluating the efficiency of the DMU “\( o \)” , we solve for the following problem:

\[
\begin{align*}
\text{Maximize} & \quad \frac{u^T y_o}{v^T x_o} \\
\text{Subject to:} & \quad \frac{u^T y_j}{v^T x_j} \leq 1, \\
& \quad j = 1,...,n, \\
& \quad u, v \geq 0.
\end{align*}
\]

where \( u \) is the \((k \times 1)\) vector of output weights and \( v \) is the \((m \times 1)\) vector of input weights. “\( T \)” denotes the matrix transpose operator. Thus, \( u \) and \( v \) are chosen to maximize the efficiency measure of the DMU “\( o \)” subject to the constraints that the efficiency levels of all units must be less than or equal to 1.

The above problem has an infinite number of solutions. To generate a unique solution, the following constraint is imposed: \( u^T y_o = 1 \). The maximization problem then becomes:

\[
\begin{align*}
\text{Minimize} & \quad v^T x_o \\
\text{Subject to:} & \quad u^T y_o = 1, \\
& \quad u^T y_j - v^T x_j \leq 0, \\
& \quad j = 1,...,n, \\
& \quad u, v \geq 0.
\end{align*}
\]

The duality problem to (3.2) can be written as follows:
Maximize $\phi_o$ \quad (3.3)

Subject to: $\phi_o y_o \leq \lambda^T Y$,

$x_o \geq \lambda^T X$ ,

$\lambda \geq 0$.

where $\lambda$ is a $(n \times 1)$ column vector and $\phi_o$ is a scalar. In other words, we search for all linear combinations of input vectors in current practices that can be provided by the input vector of the “o” unit. We then compute the maximal proportional output vector that can be produced by these linear combinations. Let $\phi_o^*$ denote the solution to (3.3). Obviously $\phi_o^* \geq 1$. If $\phi_o^* = 1$, then the DMU “o” is (CCR) technically efficient, otherwise, $\phi_o^* > 1$ and “o” is (CCR) inefficient. Later, we also denote $1/\phi_o^*$ by $TE_{CRS}$, the efficiency score measured by the CCR method. Note that the linear programming problem must be solved $n$ times, once for each DMU in the sample.

Underlying the CCR method is the assumption of constant returns-to-scale (CRS). The CRS assumption is only appropriate when DMUs are operating at an optimal scale. Imperfect competition, constraints on finance, etc., may cause a DMU to be not operating at optimal scale [Coelli, Rao and Battesse (1999)]. The BCC model modifies the CCR model by allowing variable returns-to-scale (VRS). This is done by simply adding the convexity constraint $e^T \lambda = 1$ into problem (3.3), where $e$ is a $(n \times 1)$ column vector of ones. We denote $\phi_o^{**}$ be the solution to the new problem and $1/\phi_o^{**}$ by $TE_{VRS}$, the efficiency score measured by BCC model. Clearly, $TE_{CRS} \leq TE_{VRS}$. Note that the BCC method measures purely the technical efficiency whereas CCR method measures both pure technical efficiency and scale efficiency. By using $TE_{CRS}$ and $TE_{VRS}$ measures, we derive a measure of scale efficiency i.e., $SE = TE_{CRS} / TE_{VRS}$. However scale inefficiency can be due to the existence of either sub-optimal scale size (i.e., increasing returns-to-scale (IRS)) or supra-optimal scale size (i.e., decreasing returns-to-scale (DRS)). The nature of scale inefficiencies for a particular DMU can be determined by executing an additional DEA program with the assumption of non-increasing returns-to-scale (NIRS) imposed and following methodology given in Figure 3.3.

By adding the restriction $e^T \lambda \leq 1$ in DEA model (3.3) the $TE$ scores assuming NIRS can be calculated. The calculation of technical efficiency assuming NIRS facilitates the
identification of the nature of returns-to-scale. Let the measure of $TE$ assuming NIRS be denoted by $TE_{NIRS}$. The existence of increasing or decreasing returns-to-scale can be identified by seeing whether the $TE_{NIRS}$ is equal to the $TE_{VRS}$.

![Figure 3.3: Determination of Returns-to-scale](source: Kumar and Arora (2007))

### 3.3.2 Probit and Ordered Probit Models

In case the dependent variable is dichotomous (i.e., categorical binary variable), the use of the method of Ordinary Least Square [i.e., Linear Probability Model (LPM)] fails to provide efficient estimates of the parameters of the following regression model:

$$y_i = \beta 'x + u$$  \hspace{1cm} (3.4)

where,

$$y_i = \begin{cases} 1: & \text{if event occurs} \\ 0: & \text{otherwise} \end{cases}$$

Where, $\hat{y}_i$ is equal to the probability ($p_i$) of occurring the favorable event. However, the use of OLS will provide $\hat{y}_i$ that may range outside the bounds of probability (0, 1). Moreover, the use of OLS will produce less efficient estimates of $\hat{\beta}$ because of the presence of Heteroscedasticity. Thus, the estimation of the parameters of model (3.4) requires the use of either Probit or Logit model using the method of Maximum Likelihood. The following density function is used to get the probability estimates of favorable event for each cross-sectional unit. the given values of $\hat{\beta}$ and $\sigma$:
Using the density function, marginal effect of each variable can be computed using following formula:

\[
\text{Marginal Effect of } X_k = \frac{\partial f(Z)}{\partial Z} \times \frac{\partial Z}{\partial X_k}
\]  

(3.6)

The marginal effect represents the change in probability of favorable event because of a unit change in the value of independent variable \(X_k\). In equation (3.5), the variable \(Z\) has to be computed using the values of all independent variables for each cross-sectional unit, point estimates of vector \(\beta\) and \(\sigma\). Thus, the foremost requirement is estimation of parameters \(\beta\) and \(\sigma\). The estimates of these parameters have to be estimated using the method of maximum likelihood (ML). The following likelihood function can be formulated for a dichotomous dependent variable:

\[
L = \prod_{y_i=1} f \left( \frac{1 - \beta' x}{\sigma} \right) \prod_{y_i=0} f \left( \frac{-\beta' x}{\sigma} \right)
\]

\[
= \prod_{y_i=1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(1 - \beta' x)^2}{2}} \prod_{y_i=0} \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\beta' x)^2}{2}}
\]

(3.7)

Maximizing the likelihood function \(L\) to obtain the extreme values of \(\beta\) and \(\sigma\) will provide the ML estimates of the model (3.4). Using these estimates the variable \(Z\) can be computed and so the marginal effects can be obtained. The same procedure can be extended for \(m\) categories dependent variable model (i.e., Polychotomous dependent variable model). These \(m\) categories may be i) independent of each other, ii) ordered response form, and iii) sequential response form.

In analyzing the factors affecting awareness regarding bancassurance among bank customers and factors affecting overall satisfaction from bancassurance among bank customers, the ordered polychotomous dependent variables have been used. In our case the dependent variable is polychotomous and continuous in nature i.e., representing preference at liker scale. Thus, ordinary logit/Probit models are not applicable. Here the use of ordered Probit model has been suggested instead of ordinary Multilogit/probit models. Suppose the dependent variable is continuous with following categories:
Given that the dependent variable representing $j$ points preference scaling, the simple categorical dependent variable model is not applicable. Thus, the use of the ordered response/choice dependent variable is suggested by the researchers. Let $B$ be the vector consisting of all explanatory variables affecting the extent of awareness regarding bancassurance among bank customers and $\beta$ be the vector of all slope parameters to be estimated. Then, the likelihood function may be written as:

$$L(\beta, \mu_0, \mu_1, \mu_2) = P(y_0 = 0) \times P(y_1 = 1) \times P(y_2 = 2) \times P(y_3 = 3)$$  \hspace{1cm} (3.8)

Where,

$$P(y_0 = 0) = P(y_i^* \leq \mu_0) = P(\beta'x + \varepsilon_i \leq \mu_0)$$
$$= P(\varepsilon_i \leq \mu_0 - \beta'x)$$
$$= \Phi(\mu_0 - \beta'x)$$

$$P(y_1 = 1) = P(\mu_0 \leq y_i^* \leq \mu_1) = P(\mu_0 \leq \beta'x + \varepsilon_i \leq \mu_1)$$
$$= P(\mu_0 - \beta'x \leq \varepsilon_i \leq \mu_1 - \beta'x)$$
$$= \Phi(\mu_1 - \beta'x) - \Phi(\mu_0 - \beta'x)$$

$$P(y_2 = 2) = P(\mu_1 \leq y_i^* \leq \mu_2) = P(\mu_1 \leq \beta'x + \varepsilon_i \leq \mu_2)$$
$$= P(\mu_1 - \beta'x \leq \varepsilon_i \leq \mu_2 - \beta'x)$$
$$= \Phi(\mu_2 - \beta'x) - \Phi(\mu_1 - \beta'x)$$

and

$$P(y_3 = 3) = P(y_i^* \geq \mu_2) = P(\beta'x + \varepsilon_i \geq \mu_2)$$
$$= P(\varepsilon_i \geq \mu_2 - \beta'x)$$
$$= 1 - \Phi(\mu_2 - \beta'x)$$

Thus, the likelihood function may be written as:

$$L(\beta, \mu_0, \mu_1, \mu_2) = \Phi(\mu_0 - \beta'x) \times \left[ \Phi(\mu_1 - \beta'x) - \Phi(\mu_0 - \beta'x) \right]$$
$$\times \left[ \Phi(\mu_2 - \beta'x) - \Phi(\mu_1 - \beta'x) \right] \times \left[ 1 - \Phi(\mu_2 - \beta'x) \right]$$  \hspace{1cm} (3.9)
Where, $\Phi(.)$ represents cumulative distribution function (CDF) defined as follows:

$$
\Phi(\mu_0 - \beta'x) = \int_{-\infty}^{\mu_0 - \beta'x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu_0 - \beta'x)^2} d(\mu_0 - \beta'x); \ Pr \text{ obit Estimation}
$$

and, $\Phi(\mu_0 - \beta'x) = \int_{-\infty}^{\mu_0 - \beta'x} \frac{e^{-(\mu_0 - \beta'x)}}{\left(1 + e^{-(\mu_0 - \beta'x)}\right)^2} d(\mu_0 - \beta'x); \ Logit \text{ Estimation}

Maximizing (3.9) with respect to $\beta, \mu_0, \mu_i$, and $\mu_2$, the estimates of these parameters can be obtained. The point estimates of $\beta$ are slope estimates whereas, $\mu_0, \mu_i$ and $\mu_2$ are unknown threshold parameters representing threshold limits of $y_i^*$. When faced with a ranking problem, we develop a “sentiment” about how we feel concerning the alternative choices and the higher the sentiment, the more likely a higher-ranked alternative will be chosen(Hill et al., 2011). For example, in our case, higher the extent of awareness, the more likely a higher value alternative will be chosen. These sentiments (or extent of awareness) are unobservable and when they enter decisions are called latent variables denoted by $y_i^*$.

In categorical response models, simple point estimates of $\beta$ cannot be used for interpretation purpose. The use of marginal effects is generally preferred which may be obtained as:

$$
\frac{\partial P(y = 0)}{\partial X_k} = -\phi(\mu_0 - \beta'X) \times \beta_k
$$

$$
\frac{\partial P(y = 1)}{\partial X_k} = \left[\phi(\mu_0 - \beta'X) - \phi(\mu_i - \beta'X)\right] \times \beta_k
$$

$$
\frac{\partial P(y = 2)}{\partial X_k} = \left[\phi(\mu_i - \beta'X) - \phi(\mu_2 - \beta'X)\right] \times \beta_k
$$

$$
\frac{\partial P(y = 3)}{\partial X_k} = \phi(\mu_2 - \beta'X) \times \beta_k
$$

(3.10)

In these expressions $\phi(.)$ denotes the probability density function of a standard normal variate, and its values are always positive. These marginal effects represent change in the probability of being completely aware about bancassurance because of a unit change in independent variable. The direction of effect depends upon the sign of $\beta_k$; a positive value represent positive whereas, a negative value signifies adverse impact.
3.3.3 Structural Equations Modeling and Confirmatory Factor Analysis: Deriving Drivers of Cross-buying Intention, Sources of Motivation and Satisfaction

The Confirmatory Factor Analysis (CFA) is a structural equations modeling (SEM) based technique and the factor loading so obtained can be used to construct the weights of each component variable that are used to construct the latent variables used to explain the drivers of cross buying intentions, sources of motivation and satisfaction. Furthermore, the hypothesis of inter-factor relationships have been also been tested under same framework. Using CFA, a set of latent (Factors) variables is specified with deterministic components. The latent are unobserved variables that are supposed to be formed by the component variables. In simple word, the latent is determined by the combination of component variables and supposed to cause a significant variation in each of the component variable. Thus, the variation in component variable is bifurcated into the explained variation by the latent and unexplained error variance. Suppose, \( Y_1, Y_2, \ldots, Y_5 \) are highly correlated variables that can be represented by a single factor \( F \); then under CFA the latent \( F \) can be designed in the following way:

![Figure 3.4: Latent (Factor) Formation under CFA](image)

Source: Author’s Elaboration

In SEM, a latent variable can be represented in a circle, while a deterministic/observed variable is represented by a rectangle. The arrows represent the direction of causality. The variation in the component variables \( (Y_1, Y_2, \ldots, Y_5) \) is thus explained jointly by the latent \( F \) and the respective error/residuals terms that have been distributed normally with mean 0 and variance 1. The latent/factor \( F \) can be used as a combined proxy of component variables, if a
significant variation in each $Y_k$ is explained by the latent $F$. However, Figure 3.4 explains factor component relationship, whereas, inter-factor relationship can also be observed via assigning direction of causality from one latent to another. The nature of causality depends upon a-priori information. For example, suppose another latent $F_j$ is represented by three factors $X_1$, $X_2$, and $X_3$; and it is observed on a-priori grounds that $F_1$ affects $F$, then the relationship between $F_1$ and $F$ is given by the following Model 3.5. Using the Model given in Figure-3.5, two latent/factors $F_1$ and $F$ can be generated and the inference that $F_1$ significantly affects $F$ can also be tested.

In the same way, the drivers of cross-buying intentions, sources of motivation and satisfaction have been derived. For example, Figure-3.6 provides the structure of the model used to detect the drivers of cross-buying intentions. The circles represent the latent variables that are unknown e.g., the latent of *Overall Satisfaction* is formed by the combination of three variables MFd11, MFe11, MFF11. The said latent should explain a significant proportion of variation in each of the three variables. The circle $e12$, $e13$, and $e14$ are residuals that consist of the effect of all variables other than the latent *Overall Satisfaction* on each of the component variable. In simple, these variables explain the proportion of unexplained variation in each of the three dependent variables. The residuals are also unknown, formed by the model itself and so represented by circle too.

**Figure 3.5: Inter – Factor Relationship under CFA**

Source: Author’s Elaborations
Figure 3.6: Structural Equations Model (SEM) explaining Drivers of Cross-Buying

Source: Author’s Elaborations
Further, one of the component variables of a latent (such as Overall Satisfaction) has to be assigned unit regression weight. The loadings of all the component variables are then divided by the loading of that particular variable to obtain the regression weights of each component variable. The intra-factor and inter-factor relationships have been defined through single arrow pointers and have been estimated using the method of Maximum Likelihood Estimation (MLE). The model shows how component variables are affected by the latent factor and the error residuals. The standardized parameter estimates are known as factor loading, whereas, non-standardized parameter estimates represents the impact of a unit change in latent variable say Overall Satisfaction on its component variables under evaluation. Further, the standardized parameter estimate represents the correlation of factor with the respective component variable; higher the value higher the importance of variable in defining latent/factor and vice-versa. In the said model, eight latent variables have been constructed to define drivers of cross-buying intentions. The inter-factor relationships have been assigned following a-priori information in literature.

3.4 Summary and Conclusions

The present chapter offers a brief introduction to the methodology used to test the set hypotheses. In attaining the set objectives of study, different mathematical and econometric methods have been used. The linear programming based data envelopment analysis used to analyze the impact of bancassurance on the performance of banks. It is evident from the chapter that the production structure of the sampled banks has been defined with and without bancassurance income to ascertain the efficiency change due to bancassurance.

Further, the tools such as Probit model and Ordered Probit models have been discussed in detail that helps to realize the importance of dummy dependent models in business research. In the analysis of determinants of choice of insurance product, determinants of customer satisfaction, etc. the techniques of Probit and ordered Probit analysis have been utilized. A brief introduction to structural equations modeling (SEM) based confirmatory factor analysis (CFA) is an additional feature of the same chapter. The CFA has been utilized in determining the drivers of cross-buying intention, sources of motivation to purchase bancassurance amongst bank customers, and sources of customers’ satisfaction. The sources of motivation to initiate bancassurance at branch level among bankers of Punjab have also been analyzed using SEM based CFA.
Endnotes:

1 The cluster of bank is defined on the basis of sector (i.e., Public or Private) and model of bancassurance followed (i.e., JV or Agency).
2 A DMUs is regarded as the entity responsible for converting inputs into outputs and whose performances are to be evaluated. For the purpose of securing relative comparisons, a group of DMUs is used to evaluate each other with each DMU having a certain degree of managerial freedom in decision making. In the present study the DMUs are the sampled commercial banks.
3 Under this option we assume that outputs change in direct proportion to the change in inputs regardless of the size of the DMU. That is, CRS assumes that the DMU’s scale of operations doesn’t influence its efficiency.
4 The OTE is also known as global technical efficiency.
5 The PTE is also known as local/managerial technical efficiency.
6 Variable returns-to-scale assumes that changing inputs may not necessarily result in a proportional change in outputs. That is, as a DMU becomes larger, its efficiency would either fall or rise.
7 The convexity constraint $e^T \lambda = 1$, essentially ensures that an inefficient DMU is only “benchmark” against DMUs of a similar size.