Aalto University

School of Science

MS-E2177
Seminar on Case Studies in Operations Research

Final Report

Expert Judgements in Commercial Real Estate Loan Collateral Valuation

Group 1:
Niklas Kostiander (Project manager)
Elias Tukonen
Vu Nguyen
Deepanker Kanjarla
Sara Shinde

Aalto University
School of Science

June 8, 2026

Contents

1 Introduction

1.1 Background

When a customer applies for a bank loan, for example for a property or other asset, collateral is typically required to secure the credit and manage the bank's expected losses from loan issuance. Expected loss is viewed as the "expected cost of doing business" and it is the product of three factors: Probability of default (PD, probability of the obligor becoming insolvent), loss given default (LGD,percentage of exposure that is lost if obligor defaults) and exposure at default (EAD, unpaid loan amount if obligor becomes insolvent).

In real estate loans, most typically, the purchased property itself is pledged as this collateral. To mitigate credit risk, and manage their expected losses banks often value the property below its market price. This is done, for example, because the pledged property might need to be sold under distressed circumstances. The amount of which this valuation discounting is done, is called the "haircut". This haircutting process greatly affects the LGD of a issued loan, since LGD shrinks when we increase the collateral value relative to the exposure.

Therefore, an important step of a banks loan issuance process is the determination of an appropriate haircut. In the real-estate loan issuance sector, one of the most prominent limitations of this haircut estimation, is the limited data on liquidation values of properties in the case of insolvency. If large amounts of data were available on liquidation values of different types of properties, the banks could statistically model the appropriate haircuts for each individual property.

Especially in larger commercial real estate loan issuance, since sufficient data to make these types of haircut conclusions will likely not exist, banks rely greatly on the evaluations of expert individuals. This report introduces an approach into solving this expert elicitation problem using elicitation methods and a tailored questionnaire. This report was written for our client OP Pohjola, which is the largest provider of financial services in Finland.

1.2 Objectives

The objective for our project was to develop a reproducible framework for the elicitation of expert information, that would provide OP Pohjola with useful tools for improving the transparency and reproducibility of expert judgment in loan collateral valuation.

We were tasked to develop a framework for elicitation which would be well defined and generalizable, so that the approach could be utilized in other areas of business, where expert elicitation is needed. This framework would have to include building a model, that would have the possibility of combining data modeling with the expert elicited predictions for out of sample data.

The model should present haircuts as single values with confidence intervals, and it would have the capability of comparing and changing the influence of expert judgements and available data to the final haircut. The model would also have to display how each risk driver influences the final haircut. The framework building process should be clearly documented and reported.

2 Elicitation Methods

This section presents a literature review of structured expert elicitation methods, aiming to identify a suitable framework for the client. The framework should meet the regulatory requirements of banking, and the decision making from the elicitation should be transparent and well documented. For this project, operationally lighter frameworks are preferred, and the amount of experts that are resourced for the elicitation is around 5. The overall framework, advantages and limitations for Cooke's method, the Delphi method, and the IDEA protocol are discussed.

2.1 Cooke's Method

2.1.1 Introduction

Cooke's method is a performance-based elicitation method. Experts' individual performance is evaluated against a set of seed questions. The experts do not know the answers to these seed questions, but the analysts do. There is an expectation that the experts provide accurate and informative distributional judgements to effectively capture these values. Each expert is assigned a weight based on their statistical accuracy and informativeness. These weights are used to combine judgements from different experts and get a final estimate of the target values the analysts are trying to estimate. [16]

2.1.2 Procedure that Cooke's method typically follows

First, the analysts start with identifying the target variables they wish to estimate. Seed variables are variables whose values are already known. They are defined such that they trigger the same judgement heuristics as the target variables. Ideally, experts are not aware which variables are seed variables and which variables are target variables.

The number of experts on a panel is typically in the range of 5-20. The number of seed questions asked is typically in the range of 8-20. One can ask up to a 100 questions, though most studies involve fewer than 40. The variables are measured on a continuous scale.

Each expert is assigned a calibration score based on how close the probability distribution provided by them is to the empirical distribution obtained from raw frequencies. The Kullback-Leibler (KL) divergence measure, which measures the difference between two probability distributions, is used for this purpose. An expert whose calibration score falls below a threshold α is excluded from the further process and their estimates are not used. The value of α needs to be chosen carefully and has a significant impact on the analyst's performance.[16]

2.1.3 Suitability of the Cooke's method to the project, comparison to other methods and variations of the Cooke's method

A review of 33 studies conducted using Cooke's method shows that 7 of the studies had no statistically accurate experts. These studies had between 4 and 21 experts. Since our panel size will be on the smaller side (about 5), it is possible that we might end up in a situation where none of the experts meet our threshold of statistical accuracy. Studies show that linear weighting is generally the best way to combine expert judgements.[3] [2]

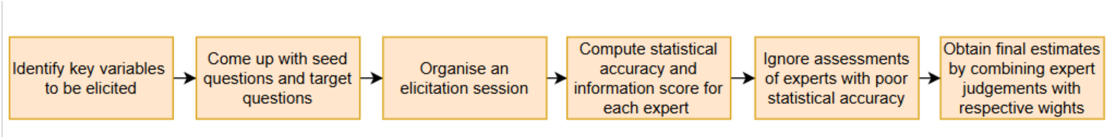

Figure 1: The Cooke's method.

2.2 The Delphi Method

The Delphi method is an expert elicitation and decision making technique developed in the 1950s by the RAND corporation. As described by the original article [4], "Its object is to obtain the most reliable consensus of opinion of a group of experts. It attempts to achieve this by a series of intensive questionnaires interspersed with controlled opinion feedback."

As an elicitation method, all classical "Delphi" approaches have the characteristics of anonymity, iteration, controlled feedback, and statistical aggregation of group responses [18]. Anonymity between expert panelists is kept to allow for freedom of expression, since in a non-anonymous group setting, dominant individuals or the majority often steer conversation. Controlled feedback between iterations of question rounds often consists of presenting the experts with statistical summaries of the answers of their peers. Occasionally, justifications on decisions are also provided. The end results are often derived from the mean or median of the panelists final round estimates.

A systematical review on Delphi methods ([18]) studied 27 different papers that either assessed the decision making process or applicability (ability to predict) of the Delphi method. Most of these papers assessed the result of elicitation technique by comparing the accuracy of the prediction of the final Delphi result, and the result of average/mean of the answers in the first round (i.e before iterative steps of the Delphi process). While most studies — 11 out of 15 that measured and compared outcomes — found the final Delphi result to be more accurate, this suggests it can be a reasonable approach in certain situations.

Since the Delphi method is fundamentally designed to reach the most reliable consensus, the very process by which that consensus is formed warrants scrutiny. Even under the anonymity the method provides, experts' responses in subsequent rounds may be unduly influenced by others' estimates — an influence that may not always be justified. When a Delphi process lacks a mechanism for panelists to justify their individual judgments, the gradual shift of responses toward the previous round's statistical average becomes difficult to interpret and validate.

2.2.1 Delphi Method as a Process

As described in the previous section, the Delphi method consists of iterations of rounds that involve different characteristics. Since all panelists work individually and anonymously, the method can be described in the eyes os a single panelists as a series of inputs and outputs. Figure 2 illustrates an example of a classical Delphi method process with 5 panelists.

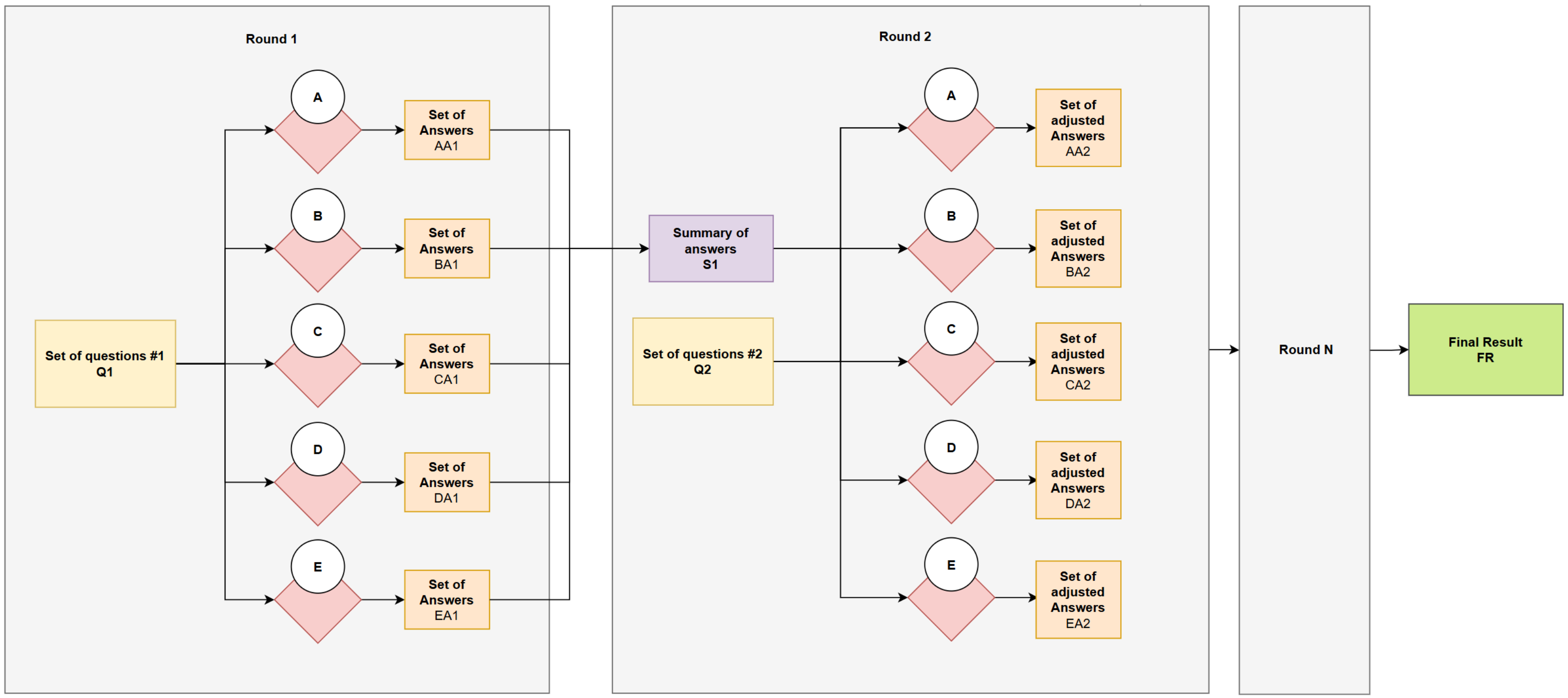

Figure 2: Structure of the Delphi Method process.

2.2.2 Delphi Method in Collateral Value Estimation

The Delphi method typically consists of several iteration rounds. A more recent study by Hsu and Sandford [12] describe that the most usual number of elicitation rounds is 4. Several rounds is needed to assure the convergence towards a consensus. The client requirements of fast execution, and an operationally lighter framework could be seen as not to fit the classical Delphi method. Therefore, we consider alternative approaches to the Delphi method in order to better fit our clients needs. The IDEA protocol is examined as a lighter version of the Delphi method.

2.3 IDEA Protocol

The IDEA protocol (Investigate, Discuss, Estimate, Aggregate) was developed to improve the accuracy of expert judgment by formalizing the interaction between participants. Unlike the classical Delphi method, which relies on an iterative consensus, the IDEA protocol uses one structured discussion for the experts to clarify their reasoning behind their estimates. The protocol does not try to reach a consensus but rather give the experts a possibility to benefit from the views of other experts. The experts still have the freedom to have their own opinion on the final estimate. [11].

While the group dynamic of a group of experts may pose a bias on elicitation, it is thought that facilitated interactions generate better results than relying solely on anonymous and independent expert judgments [10]. Since there are different biases in working alone and working in a team, the IDEA protocol provides a balance between the two, first the experts work alone, and after that they discuss their reasoning in a group.

2.3.1 Procedural steps

The protocol follows a strictly defined four-step process designed to minimize cognitive biases while maximizing the shared expertize of the experts, the overall procedure is shown in Figure 3.

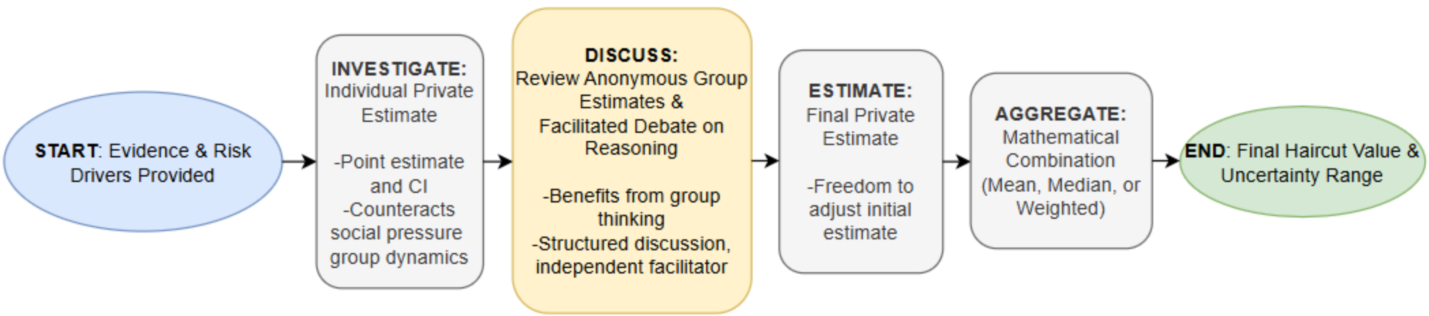

Figure 3: IDEA procedure for haircut estimation.

1. **Investigate:** Experts are provided with data to estimate the variable(s).

They perform a private, individual assessment to provide a first-round estimate. Notes on reasoning can be saved. This "pre-discussion" phase ensures that each expert's unique perspective is captured and also makes the experts more ready for the discussion.

2. **Discuss:** Following the initial estimates, the experts are presented with a anonymized summary of the group's results. The facilitator (preferably someone who is not part of the experts) then moderates a discussion. The goal is not to force a consensus, but to clarify reasoning and share the thought process behind estimates.

3. **Estimate:** After the discussion, experts provide a second and final private estimate. They are free to change their original value based on the new information shared or maintain their estimate if they wish.

4. **Aggregate:** The final individual estimates are mathematically aggregated (usually using a weighted or a simple mean) to produce the final value and a confidence interval.

2.3.2 Suitability for CRE haircut estimation

The IDEA protocol is particularly suited for the banking context for several reasons, the main one over the Delphi method is how quickly the IDEA protocol can be carried out. While the Delphi method may require four or more rounds over several sessions, the IDEA protocol can be executed in a single facilitated workshop. This advantage is especially important as the client has expressed their preference for lighter and more agile frameworks over time-consuming ones.

The IDEA protocol reduces bias by combining individual, and private, estimation and structured debate. It gives experts the freedom to decide for themselves while gaining information from other experts. This combination of private estimates and a group discussion can be considered the core of the IDEA protocol. When making initial estimates, adjustment from an anchored value has been found less accurate if the anchored value is self-generated by the expert [6].

The discussion phase of the method can be recorded, and provided to the authorities and supervisors as the basis of the final decision. This is needed particularly in the banking sector where regulation is stringent and decisions require transparency.

2.4 Chosen elicitation method

Out of the three discussed elicitation methods, the IDEA protocol is our chosen elicitation method for the project. This was to meet the client preference for an operationally lighter framework that could work with a small panel size. The Delphi method, while effective, requires several rounds of elicitation and discussion. A unique strength of the Cooke's method is that it assesses the statistical accuracy of an expert's predictions before assigning a weight to their estimate.

However, the Cooke's method was found not to be suitable for small panels like ours (panel size 5). We therefore chose the IDEA protocol for our project. Unlike the Delphi method, it only involves a single round of discussion, and two rounds of elicitation.

3 Risk drivers for collateral value and haircut estimation

In addition to selecting an elicitation framework, the project also requires identifying the main drivers that may explain why the realized selling price of collateral differs from its initially estimated market value. In this report, these risk drivers are treated as potential inputs to the presented haircut model for commercial real estate collateral. The aim at this stage is not to construct the final model, but to document the most relevant variables from the literature and show how they could later be measured in practice. Based on literature, the most relevant drivers can be grouped into macroeconomic conditions, market conditions, property-specific characteristics, tenancy-related characteristics, location-related characteristics, and environment-related characteristics [13, 8, 5, 1, 7]. The risk drivers are summarized in Table 1.

Table 1: Risk Drivers and how to measure them.

Risk Driver	**How to Measure**
Property Type	Office, Logistics, Retail, etc.
Building Age & Condition	Years since construction, years since last major renovation
Occupancy & Tenant Quality	Occupancy rate (%), vacancy rate (%), and the Weighted Average Lease Term (WALT) in years.
Location & Accessibility	City/sub-market categorical dummies, physical distance to the Central Business District (CBD) in km, and distance to public transport hubs.
Climate Risk (Physical)	Geographic hazard metrics tailored to Finland (e.g., flood, precipitation, heat, annual expected regional damage).
Climate Risk (Transition)	Official Energy Performance Certificate (EPC) ratings (A-G), ESG scores.
Macroeconomic Conditions	Regional/national unemployment (%), interest rates, GDP growth, and industry-wide default rates.
Market Liquidity	Quarterly transaction volumes in the sub-market, average time on market (months), and local sub-market vacancy rates.

3.1 Macroeconomic conditions

Macroeconomic conditions matter because collateral values do not depend only on the property itself, but also on the wider economic environment at the time of sale. For this reason, variables such as interest rates, unemployment, and GDP growth are natural candidates for the model. A weaker macroeconomic environment may reduce liquidity, increase required yields, and lead to realized prices that fall below earlier appraised values. Furthermore, the industry-wide default rates (percentage of default loans in a specific sector) are also relevant to the income of the property; it has been shown to have a strong relationship with LGD [17]. High default rates indicate a heavy discount in price during a liquidation sale due to a saturated industry-specific market situation.

3.2 Market conditions

Closely related to this, market liquidity is also an important driver to determine recovery value of the collateral. In the commercial real estate context, this can be measured using transaction volume, time on market, and local vacancy [1]. These variables capture how active the market is, how easily comparable assets can be sold, and whether demand in the surrounding market is weakening.

3.3 Property-specific characteristics

Property-specific characteristics are another key group of drivers. Property type and subtype matter because different commercial real estate assets do not face the same resale risk or buyer depth. A generic logistics property is typically easier to sell than a highly specialized asset, and the literature links limited re-deployability and a narrower buyer pool to larger discounts in distressed sales [5].

Building age and years since renovation can also be included, because they capture physical condition, technical quality, and possible obsolescence [1, 14]. These conditions directly affect liquidation values since buyers need to account for repair costs, leading to a decrease in value in a forced sale. The older a property is, the higher possibility of technical issues it bears. The reason for introducing the time since the last major renovation in addition to year of construction is that it can improve the quality of the building and lessen the need for immediate repair costs for the next buyer.

3.4 Tenancy-related characteristics

In addition, tenancy-related characteristics are important for income-producing real estate. Occupancy rate and weighted average lease term (WALT) provide information about current cash-flow stability and leasing risk. These variables are relevant because realized collateral value depends not only on the asset as a physical object, but also on the strength and stability of the income that the property generates [8, 1, 15]. In a forced sale, a high occupancy rate indicates good sustainability during the transition of the property to a new owner. For example, office buildings with high tenancy help reduce the maintenance costs of the building that the new buyer has to cover. This makes the property more preferable and liquid. Similarly, a long-term WALT ensures stability of the property in the long run and lessens the risks of vacancy. Properties with short WALT in distressed sale might be considered as already vacant in the eyes of buyers, which leads to a decrease in price.

3.5 Location and accessibility characteristics

Finally, location and accessibility should be included, since commercial real estate values are strongly shaped by where the property is located and how accessible it is. In a liquidation sale, it is extremely difficult to sell a property in an isolated region, far from public transport; even if there are buyers, the actual price would be heavily discounted given these poor conditions. In a simplified hypothetical model, this can be measured using city or sub-market indicator variables together with distance to the central business district and distance to public transport [1, 15].

3.6 Climate risk metrics

It is also important to consider climate risk in collateral valuation, since the risk of natural catastrophes or poor energy performance can affect the market value and make potential buyers less inclined to buy the property. Given a forced sale situation, they can make a property with high climate damage risk unmarketable or heavily discounted in price, influencing the banks to set a high haircut in the collateral valuation process.

Climate risks can be divided into transition risks and physical risks. Transition risks refer to the drop in collateral's value that can arise from the world's shift to a lower-carbon and more sustainable economy [7]. Environmental regulations and policies can pose a threat on the value of a property if it does not follow the standardized sustainable quality; the financial value of owning it would be decreased due to future upgrading costs to meet the policies. This can be measured with standardized ESG (Environmental, Social, and Governance) score and energy performance certificates (EPC) to estimate the overall environmental profile of the asset. Physical risks represent the impact of climate changes and extreme weather hazards, which can directly damage the quality and thus the price of the property. The metrics are based on common hazards in Finland, including flood, precipitation and heat metrics along with regional annual expected damage from these hazards.

3.7 Selected risk drivers

Among the found risk drivers, we have selected the four primary risk drivers for the analysis and further demo testing in this project: location, year of construction, property type, and occupancy rate. These four risk drivers possess high predictive power, availability and reliability. They correspond to primary factors of a property valuation: location provides information of the average price and demand; year of construction shows potential maintenance costs; property type indicates the buyer pool; and occupancy rate represents the current financial performance of the property. For each of the risk drivers, values are grouped into different categories to improve the efficiency of the expert elicitation process and ensure more consistent and standardized estimates, rather than relying on separate, specific inputs. The categories are defined as follows in Table 2.

Table 2: Categories of different risk drivers.

Risk driver	**Category**
Property Type	Retail Space Office Hotel Industrial Warehouse
Year of Construction	>2015 1990–2015 1970–1990 1950–1970 <1950
Location	Capital Region Big cities (not capital) Other parts of Finland
Occupancy rate	>80% 70–80% 50–70% <50%

The categories of property type are the 4 most common commercial property types; they are defined based on the difference in their liquidity profiles and value stability. A property with a more specialized (less versatile) purpose might be considered more unstable. The categories of construction year (the year of completed construction) are defined based on the time difference in architectural structure of the properties as well as maintenance risk. The categories of location are defined according to Statistics Finland's regional housing price classification. The categories of occupancy rate (proportion of total usable and leasable area in m^2 are currently in use) are decided based on the income stability of the property; lower occupancy rate indicates higher carrying costs.

4 Elicitation Process

4.1 Dimensionality

After finding the most important risk drivers, they need to be tested with a questionnaire. With 4 risk drivers, each having 4 categories, there is $4^4 = 256$ combinations to fully determine all of the interactions between the variables. It is not reasonable to have an expert answer hundreds of questions and not suffer from fatigue. Also, if new variables or categories are added, the amount of combinations grows exponentially making the model sensitive to future changes. Pre-processing is needed to cut down the amount of questions needed. [9].

One way to cut down the amount of questions is to use heuristics. If we know the most important interactions, e.g. if the occupancy rate is really low for an estate, the haircut is $x\%$ regardless of its location, the questionnaire does not need to have questions with low vacancy rates for all of the locations. These scenarios are statistically redundant and should not be in the questionnaire.

Another way of lowering the amount of questions is the assumption of independence of variables. If the variables are assumed to be independent, their interactions do not change the value estimated which would result in a rather simple questionnaire. If in reality there are correlations between the variables and the model assumes them to be independent, it would reduce the predictive power of the model, that is, being accurate at predicting the liquidation value of a property. With the assumption of independence, the amount of questions decreases to $m \cdot n$, where m is the amount of variables and n the amount of categories. This is a significant decrease compared to the exponential growth of questions without the independence assumption.

To move from fully independent variables, one could ask the experts to point out the most correlated groups of variables and test the interactions of these variable groups. A middle ground between fully independent and using all combinations should be found, either with the help of experts or using heuristics from literature. To find out the effect of different risk drivers on the liquidation value, we propose the use of a questionnaire with the following structure for the experts.

4.2 The questionnaire

The questionnaire consists of two parts: the first part is constructed for the estimation of liquidation value assuming independence between variables, and the second part is for discussion of possible dependencies between these variables. The overall process of the elicitation follows the IDEA protocol seen in Figure 3.

4.2.1 First part of the questionnaire

The questionnaire sets a standardized format for consistency over different responses, in which it introduces a CRE property with a market value of 1,000,000€ and asks for estimates of the liquidation values in different scenarios. Liquidation value is the price the bank expects to receive from the sale of the property when there is insufficient time to sell. This could be due to the bankruptcy of the client, for example. The liquidation value is typically lower than the fair market value. With this fixed property, we aim to isolate the impact of each risk driver (assuming independence) on the liquidation value of the property. For each question tab, only one variable is modified while the others remain "neutral" (unchanged, unrelated). The four selected risk drivers are location, year of construction, property type, and occupancy rate.

In order to reduce chances of fatigue for responders, the experts have two ways to answer the questionnaire. Either by providing numerical estimates for each category in each variable, or to answer based on predefined risk change levels relative to other categories. It might be easier for the expert to determine how much the liquidation value increases compared to the lowest liquidation value rather than giving it a numerical value. Table 3 displays the predefined risk levels.

Table 3: Predefined risk levels.

Level	**Description**
0	No change in liquidation value
1	Small increase in liquidation value
2	Moderate increase in liquidation value
3	Significant increase in liquidation value

If the expert finds it difficult to assess numerical values for all of the categories, or finds it easier to use relative changes, they are asked to rank the categories from the value with highest risk (lowest liquidation value) to the the value with the lowest risk (best liquidation value). After that, for each category, they should select a predefined risk change that represents how significant the increase in liquidation value is relative to the category that is ranked immediately above it (the next worse category). Finally, the liquidation values for the best and the worst categories are asked. This is done to calculate the numerical liquidation values for the intermediate categories, using multiple risk levels allows for nonlinearities between the categories to be determined. The process of answer aggregation is explained in Section 5.

The input of this part of questionnaire is a set of standardized liquidation values and ranking of risk drivers' categories provided by experts. The output is an aggregated result to establish distributions of liquidation value among each risk driver. With haircuts for different (independent) variables, the answers can be then aggregated to a model that predicts the liquidation value for a new property (see Section 5).

4.2.2 Second part of the questionnaire

The second part in the questionnaire is introduced to address further variable dependencies that the first part fails to capture due to assumptions of independence. It provides a table with 6 rows corresponding to 6 different pairwise combinations of variables for the experts to provide opinions on their interactions. They are asked to identify whether if there exists a dependency/correlation between two variables (Yes, Maybe, No) and provide their reasoning for it along with descriptions of the dependency or specific categories combinations that they find significant to the liquidation value. They also need to identify the direction of the impact this relationship has on the liquidation value (whether it leads to an increase/decrease). An example of the answer is described in Table 4.

Table 4: Example of the answers for variable dependencies.

Variable Pair	**Nature of Dependency (How?)**	**Impact on Liquidation Value (Direction?)**
Property Type $\times$ Location	Hotels and industrial warehouses get varying liquidation values depending on location, while apartment buildings are fairly consistent across regions.	Hotels in the Capital region have higher liquidation values due to tourism/business demand. Industrial warehouses in the Rest of Finland have lower values due to poor logistics access.

4.3 Discussion Phase and Logistics

In the discussion, the questionnaire responses are kept strictly anonymous to prevent group bias and professional hierarchy. The discussion needs to have a neutral facilitator who is responsible for regulating the discussion process so as to analyze and explore the reasons/assumptions behind the divergence of responses. This facilitator must have sufficient knowledge of collateral valuation and statistical analysis since they must present the answers to the experts, as well as a clear understanding of this framework to manage the flow of the discussion.

The discussion begins with the starting phase, where the facilitator needs to ensure that all required tasks are clearly and fully explained to the experts. After this, copies of all responses will be printed and distributed to all experts. For a clearer observation, a summary of results will be presented by the facilitator in the form of visualizations, focusing on the distribution of the answers for each change in a variable. A correlation heatmap can also be presented to provide an overview of the responses in the second part of the questionnaire. During this process, the facilitator would demonstrate the summary of results and include the experts' provided reasons of some notable answers if there is a significant divergence or any major agreements or disagreements spotted in the responses. Finally, the facilitator will end the group discussion and give a brief description of the next task in the framework (the second round of questions).

In the second round, the first part of the same questionnaire is sent to the experts again, allowing them to either keep their original answers or revise them if they wish. After the second round answer to the first part of the questionnaire, the facilitator aggregates the answers into a final comprehensive model form.

The entire process of the framework implementation and its logistics is pictured in the below figure 4.

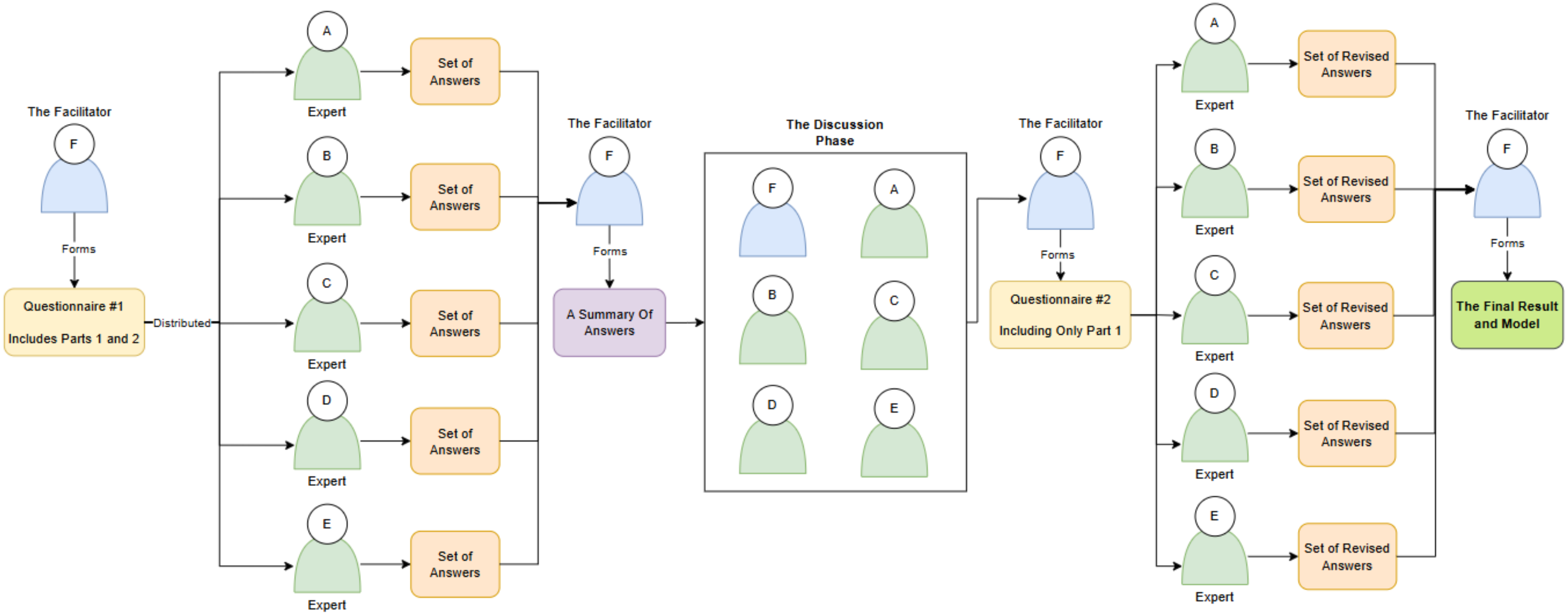

Figure 4: Structure of the Elicitation Process and Logistics.

5 Aggregation and Model

5.1 From Rankings to Numerical Haircuts

The first part of the questionnaire asks experts to do two things for each risk driver: rank its categories from highest to lowest liquidation value, and assign a qualitative step label (0, 1, 2, or 3) to each successive step in that ranking. This yields an ordered, semi-quantitative picture of how each category compares to the others within a single driver. To convert this into actual numerical haircuts, a third input is required: the cardinal liquidation value estimates that each expert provides for the same scenarios.

5.1.1 Example aggregation

If the expert gives answers with liquidation values, they are used as is. If, however, the rankings table is used, the following formula is used to calculate the missing liquidation values (V_k):

$$V_k = V_{min} + \left(\frac{\sum_{i=1}^{k} c_i}{\sum_{i=1}^{n} c_i} \right) \times (V_{max} - V_{min}),$$

where c_i is the change in liquidation values in each intermediate category, V_{min} and V_{max} represent the lowest and highest liquidation values, respectively. The sum in the numerator is the cumulative sum up to intermediate category k. Using this formula to a numerical example with an expert answer seen in Table 5 we get that the liquidation value for Big cities is:

$$V_1 = 0.6 + \frac{3}{3+1} \times (0.9 - 0.6) = 0.825.$$

Table 5: Example answer from an expert

Category	**Liquidation value change**	**Liquidation value (in M€)**
Other parts of Finland	-	0.6
Big cities (not capital)	3	-
Capital Region	1	0.9

5.1.2 Haircuts for the expert panel

When all of the numerical haircuts for each expert is determined, the overall haircut for each category ($h_{f,c}$) is calculated as the mean of expert answers for that category. The median might be preferred when the responses contain clear outliers.

5.2 Step 2: Applying the model

To estimate the haircut for an unknown property, each of its attributes is matched to the corresponding category in the lookup table, and the four category-level haircuts h_f are retrieved. Under the independence assumption, the final haircut estimate is a weighted combination of the four factor-level haircuts:

$$\hat{H} = \sum_{f=1}^{4} w_f \, h_f, \qquad \sum_{f=1}^{4} w_f = 1. \tag{1}$$

The weights w_f reflect the relative importance of each risk factor in explaining variation in liquidation values. In principle, these weights can be elicited directly from the experts alongside the liquidation value estimates, or derived from the relative spread of haircuts across categories within each driver. A driver whose categories produce a wide range of haircuts contributes more to the overall prediction, and assigning it a higher weight is consistent with that contribution.

For the purposes of this project, equal weighting ($w_f = 0.25$ for all f) is used as the baseline. This is the most transparent and conservative choice: it makes no assumption about which risk factor matters more, and it ensures that the model output is fully traceable to the elicited category-level values. Where OP wishes to depart from equal weighting in a future iteration, the weights can be treated as a further elicitation target or estimated from historical transaction data once sufficient observations are available.

5.3 Numerical example

Consider a property with the following attributes: located in Helsinki (Capital Region), property type Hotel, occupancy rate of 70%, built in 1999 (category 1990–2015). Table 6 shows the matched category-level haircuts.

Table 6: Haircut components for the example property.

Risk Factor	**Matched Category**	**Haircut (%)**
Location	Capital Region	9
Property Type	Hotel	25
Occupancy Rate	70–80%	18
Year of Construction	1990–2015	14
Weighted Average (equal weights)		**16.5**

Applying equation (1) with equal weights:

$$\hat{H} = \frac{9 + 25 + 18 + 14}{4} = 16.5\%. \tag{2}$$

The estimated liquidation value of the property is therefore

$$\hat{V}_{\text{liq}} = 1{,}000{,}000 \times (1 - 0.165) = 835{,}000. \tag{3}$$

5.4 Limitations of the independence assumption

The independence assumption simplifies computation and ensures model transparency, but it has a known limitation: it will systematically underestimate haircuts for properties that are simultaneously risky across multiple dimensions. For instance, a pre-1950 industrial warehouse in a rural area with low occupancy may carry compounding risks that a simple average does not fully capture. This is a recognized trade-off in expert elicitation models of this type, and the facilitated group discussion that follows the anonymous elicitation phase provides an opportunity to identify and discuss such cases

6 Results

6.1 Haircut estimates

Table 7 shows the aggregated answers of four real estate experts working for the client.[1] The table shows the estimated haircut, the t-distribution 95% confidence intervals, the standard deviation (SD) and the standard error (SE) for each elicited category. For the confidence intervals, a $t(3)$-distribution is used since we got answers from 4 experts and the degrees of freedom is $n-1$. In each variable, the values are sorted by descending haircut value. The findings are as follows:

The data suggests that the age has a clear relationship with the haircut value. Modern buildings (built after the year 2015) have the lowest haircut of 11%, whereas the oldest buildings ($<$1950) have a haircut of 27%. It is interesting that the worst haircut is given for buildings built in the year 1950-1970; these types of buildings may be considered old, but not old enough to have historical value. It should be noted that the age seems to have the largest deviation between categories; the haircut ranges from 0.11% to 0.34%, and the standard deviations for each category are moderate.

For property type, the results suggest that retail spaces are considered to perform well under liquidation. This is seen by the lowest haircut in the variable, a value of 18%. Other property types seem to perform almost equally with respect to each other, ranging between 33% and 35%. It should be noted that the standard deviation for industrial warehouses is the largest in the whole dataset, two experts considered them to perform really well, whereas two considered them to perform the worst. These types of irregularities are to be discussed in the discussion phase.

For location, bigger cities are favored. Big cities outside of the Capital region seem to perform the best, a haircut value of 19% is estimated for them. There is more

[1] The answers are from a pilot test and are not fully representative of the client's total expert pool.

deviation in the experts' opinion on the Capital region and other parts of Finland as the standard deviation is significantly larger compared to big cities.

Occupancy rate has a large deviation between categories ranging between 14% and 36%. The relationship between the haircut and the occupancy rate seems to be linear, an increase of around 7% between each category. This result is in line with the literature review, properties with existing tenants are more favorable to be sold under time pressure.

Table 7: Estimated haircut values by category.

Risk driver	**Category**	**Haircut**	**95% CI**	**SD**	**SE**
Property Type	Retail Space	0.18	[-0.01, 0.37]	0.103	0.059
	Office	0.33	[0.15, 0.50]	0.096	0.055
	Hotel	0.33	[0.16, 0.50]	0.094	0.054
	Industrial Warehouse	0.35	[-0.13, 0.83]	0.261	0.151
Year of Construction	>2015	0.11	[-0.05, 0.26]	0.084	0.049
	1990–2015	0.19	[-0.01, 0.39]	0.109	0.063
	1970–1990	0.26	[0.06, 0.46]	0.110	0.063
	<1950	0.27	[0.09, 0.45]	0.099	0.057
	1950–1970	0.34	[0.07, 0.62]	0.149	0.086
Location	Big cities (not capital)	0.19	[0.00, 0.39]	0.108	0.062
	Capital Region	0.20	[-0.07, 0.46]	0.146	0.085
	Other parts of Finland	0.25	[-0.08, 0.58]	0.178	0.103
Occupancy rate	>80%	0.14	[-0.10, 0.38]	0.131	0.076
	70–80%	0.21	[-0.05, 0.46]	0.137	0.079
	50–70%	0.29	[0.05, 0.53]	0.131	0.076
	<50%	0.36	[0.11, 0.62]	0.138	0.079

6.2 Simulated data

To test the model, we use a simulated dataset. This data was provided by the client but it should be emphasized that this is a completely simulated dataset, and **does not** reflect the real processes or data of OP-Pohjola. The only purpose of the dataset is to test the model in the project context that has been laid out by the project team.

The simulated dataset consists of 188 properties with varying market values, risk drivers, and realized liquidation values. The mean haircut percentage is 19.7% and the median is 18.1%. The minimum and maximum haircuts in the dataset are 0% and 58.9%. A haircut of 0% means that the property got liquidated with its fair market price. The market values of the properties vary from 163 000€ to 2 575 000€.

6.2.1 Mean haircut model

To begin the testing, we use the mean haircut model explained in Section 5. Figure 5 shows the cumulative distributions for the simulated haircuts and the predicted haircut. It is seen that the model presented by the project only predicts values in the interval $[0.155, 0.393]$ and thus cannot capture smaller or larger haircuts which are present in the simulated data. Figure 6 shows the relationship between the simulated liquidation value and the predicted liquidation value. In the data, especially for larger estates with larger liquidation values the model undervalues the property. The predicted liquidation value consistently falls under the perfect prediction line shown in red. For the whole dataset, the mean absolute error for the haircut percentage is 13.5%.

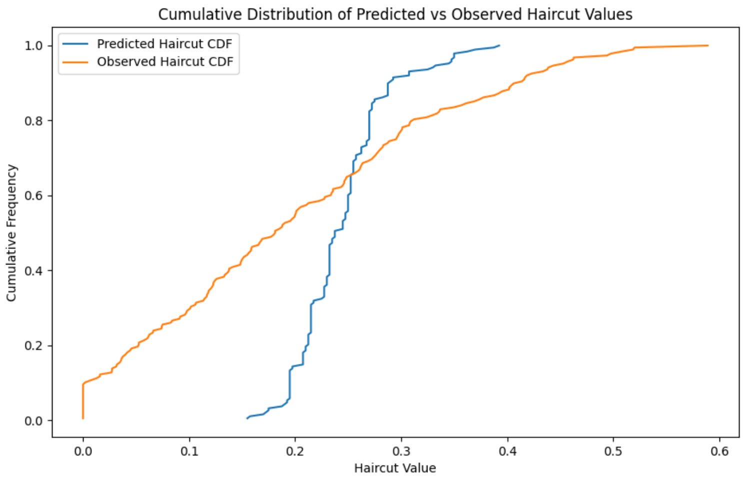

Figure 5: CDF of simulated haircut values and predicted haircut values.

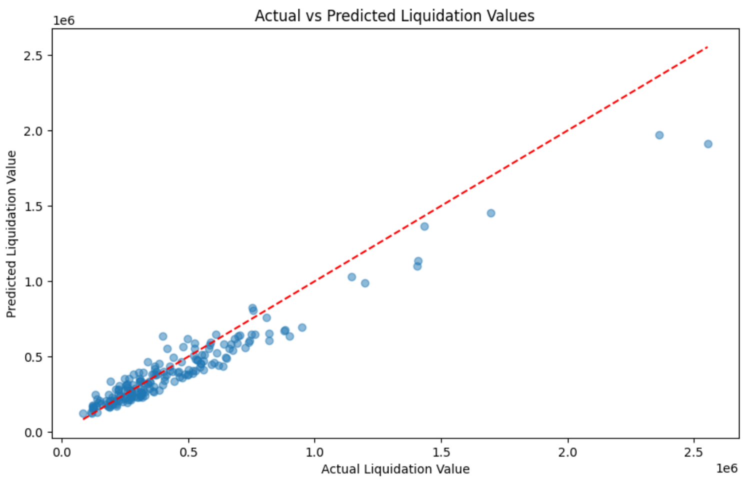

Figure 6: Simulated versus predicted haircut, perfect prediction as a red line.

The model assumes "neutral" properties, as the expert elicitation framework intentionally excluded extreme outlier scenarios where there are multiple risk drivers realized. Consequently, the model's predictive range is constrained to the interval $[0.155, 0.393]$. To evaluate the model's performance within its intended scope, a secondary test was conducted by filtering the simulated dataset to include only observed (simulated) haircuts within this range. This reduced the sample size to 81 properties. Such filtering is justified because the underlying questionnaire specifically instructed respondents to consider "average" property profiles. Therefore, the model is incapable of capturing the outliers, such as the liquidation value being equal to the market value, which are present in the full simulated dataset.

The results for the filtered dataset are seen in Figures 7 and 8. It can be seen from the figures that the filtered dataset and the predicted haircut have comparable distributions. The predicted liquidation value is more accurate, a mean absolute error in the haircut of 6.4%.

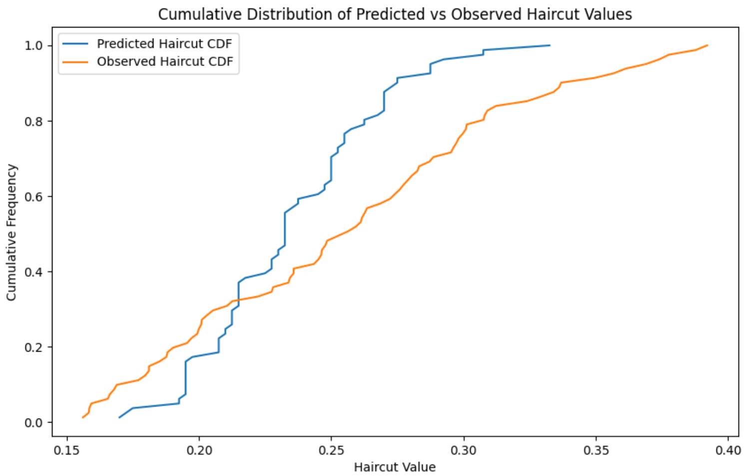

Figure 7: CDF of filtered (n=81) simulated haircut values and predicted haircut values.

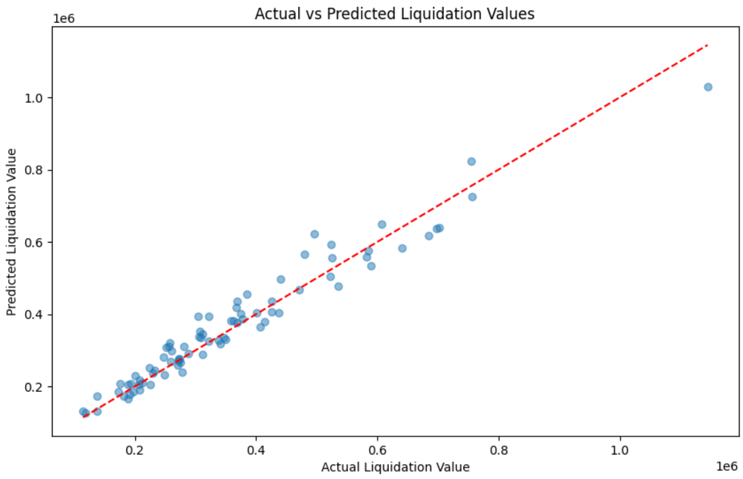

Figure 8: Filtered dataset liquidation values versus predicted, perfect prediction as a red line.

6.2.2 Predicting with individual variables

To continue the testing, we assess the predictive performance of each risk driver individually. This is achieved by generating predictions based solely on a single variable's values and computing the Mean Absolute Error (MAE) across their categories. For example, the simulated data is filtered to only include properties in the capital region, then the haircut value for the capital region in Table 7 is used to predict the haircut and the error is calculated. The results can be seen in Table 8. The MAE is greater than 0.1 for all categories, the lowest one being for Occupancy rates of 70–80% and the highest for very low occupancy rates (<50%).

Table 8: Mean Absolute Error (MAE) by category, whole simulated dataset.

Risk driver	**Category**	**MAE**
Property Type	Retail Space	0.130
	Office	0.186
	Hotel	0.166
	Industrial Warehouse	0.162
Year of Construction	> 2015	0.159
	1990–2015	0.130
	1970–1989	0.128
	< 1950	0.130
	1950–1970	0.214
Location	Big cities (not capital)	0.127
	Capital Region	0.126
	Other parts of Finland	0.116
Occupancy Rate	$> 80\%$	0.121
	70–80%	0.113
	50–70%	0.148
	$< 50\%$	0.226

Using the mean haircut model the MAE is around 0.13, some of the predictions made by the individual variables outperform the model with all of the variables. For example, the mean haircut model is outperformed by only predicting based on the location of the property.

A similar analysis is done for the filtered dataset, the results are shown in Table 9. Again, it can be said that only predicting haircuts based on the location seems to perform the best, outperforming the model with all the variables for the Capital Region and Other parts of Finland. Furthermore, only the category *Location = Other parts of Finland* and *Occupancy rate = 70–80%* get a more accurate prediction than predicting the median of the dataset.

Table 9: Mean Absolute Error (MAE) by category, filtered dataset.

Risk driver	**Category**	**MAE**
Property Type	Retail Space	0.063
	Office	0.079
	Hotel	0.099
	Industrial Warehouse	0.119
Year of Construction	> 2015	0.138
	1990–2015	0.046
	1970–1989	0.048
	< 1950	0.057
	1950–1970	0.110
Location	Big cities (not capital)	0.076
	Capital Region	0.048
	Other parts of Finland	0.037
Occupancy Rate	> 80%	0.116
	70–80%	0.040
	50–70%	0.065
	< 50%	0.128

7 Conclusion

The questionnaire was used for this project was brief. To obtain greater accuracy, the questionnaire used would have to be more detailed, and filling it out would require more time and effort on part of the experts. The questionnaire is built on the assumption of independence. While this helps simplify the questionnaire, the independence assumption is unrealistic. Under the independence assumption, the model will systematically underestimate haircuts for properties that are simultaneously distressed across multiple dimensions.

In the mean haircut model, some of the predictions made by the individual variables outperform the model with all of the variables. For example, the mean haircut model is outperformed by only predicting based on the location of the property.

This report proposes a framework to approach expert elicitation problem in CRE collateral valuation. This was accomplished by doing a literature review to research suitable elicitation methods, creating a questionnaire to get inputs from experts, building the model, and finally, testing the model on simulated data.

We think that the IDEA protocol is suitable for expert elicitation in CRE collateral valuation. It is easier to implement than the Cooke's method and less time-intensive than the Delphi method.

References

[1] Björn von Ahlefeldt-Dehn et al. *Commercial Real Estate Pricing Dynamics: A Machine Learning Analysis*. Tech. rep. Real Estate Research Institute, 2023. URL: https://www.reri.org/research/files/2023funded_commercial-real-estate-pricing-dynamics.pdf.

[2] Abigail R Colson and Roger M Cooke. "Expert elicitation: using the classical model to validate experts' judgments". In: *Review of Environmental Economics and Policy* (2018).

[3] Roger M. Cooke. "Validation in the Classical Model". In: *Elicitation: The Science and Art of Structuring Judgement*. Ed. by Luis C. Dias, Alec Morton, and John Quigley. Cham: Springer International Publishing, 2018, pp. 37–59. ISBN: 978-3-319-65052-4. DOI: 10.1007/978-3-319-65052-4_3. URL: https://doi.org/10.1007/978-3-319-65052-4_3.

[4] Norman Dalkey and Olaf Helmer. *An Experimental Application of the Delphi Method to the Use of Experts*. RAND Memorandum RM-727-PR. Prepared for United States Air Force Project RAND. Santa Monica, CA: RAND Corporation, July 1962.

[5] Irem Demirci, Umit G. Gurun, and Erkan Yönder. *Real Estate Holdings of Public Firms and Collateral Discount*. Tech. rep. Real Estate Research Institute, 2016. URL: https://www.reri.org/research/article_pdf/2015_demirci-gurun-yonder_collateralpricingandcre.pdf.

[6] Nicholas Epley and Thomas Gilovich. "Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors". In: *Psychological science* 12.5 (2001), pp. 391–396.

[7] European Central Bank. *Guide on climate-related and environmental risks: Supervisory expectations relating to risk management and disclosure*. Tech. rep. European Central Bank, Nov. 2020, pp. 35–37. URL: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.202011finalguideonclimate-relatedandenvironmentalrisks~58213f6564.en.pdf.

[8] European Mortgage Federation - European Covered Bond Council. *Study on the Valuation of Property for Lending Purposes*. Tech. rep. European Mortgage Federation - European Covered Bond Council, 2017. URL: https://hypo.org/sites/default/files/2025-08/EMF-ECBC-Study-2017-FINAL.pdf.

[9] Richard F Gunst and Robert L Mason. "Fractional factorial design". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 1.2 (2009), pp. 234–244.

[10] AM Hanea et al. "Classical meets modern in the IDEA protocol for structured expert judgement". In: *Journal of Risk Research* 21.4 (2018), pp. 417–433.

[11] Victoria Hemming et al. "A practical guide to structured expert elicitation using the IDEA protocol". In: *Methods in Ecology and Evolution* 9.1 (2018), pp. 169–180.

[12] Chia-Chien Hsu and Brian A. Sandford. “The Delphi Technique: Making Sense of Consensus”. In: *Practical Assessment, Research, and Evaluation* 12.10 (2007), pp. 1–8.

[13] International Valuation Standards Council. *IVS 104: Bases of Value*. Tech. rep. International Valuation Standards Council, 2016. URL: https://www.ivsc.org/wp-content/uploads/2021/10/IVS104BasesofValue.pdf.

[14] Peter F. Korpacz. “Appraising Retail Properties”. In: *Shopping Centers and Other Retail Properties: Investment, Development, Financing, and Management*. Ed. by John R. White and Kevin D. Gray. New York: John Wiley & Sons, 1996. Chap. 20, p. 426. ISBN: 0-471-04002-9. URL: https://books.google.fi/books?id=oMU0RE32DmUC&pg=PA426.

[15] Mariya Letdin et al. “An Analysis of the Determinants of Retail Capitalization Rates”. In: (Mar. 2022). URL: https://www.reri.org/research/files/determinants_of_retail_caprates.pdf.

[16] John Quigley et al. “Elicitation in the Classical Model”. In: *Elicitation: The Science and Art of Structuring Judgement*. Ed. by Luis C. Dias, Alec Morton, and John Quigley. Cham: Springer International Publishing, 2018, pp. 15–36. ISBN: 978-3-319-65052-4. DOI: 10.1007/978-3-319-65052-4_2. URL: https://doi.org/10.1007/978-3-319-65052-4_2.

[17] Emily Johnston Ross and Lynn Shibut. *What Drives Loss Given Default? Evidence From Commercial Real Estate Loans at Failed Banks*. Working Paper 2015-03. Federal Deposit Insurance Corporation (FDIC), 2015, pp. 1–3. URL: https://fdicdev.prod.acquia-sites.com/analysis/cfr/working-papers/2015/2015-03.pdf.

[18] Gene Rowe and George Wright. “The Delphi technique as a forecasting tool: issues and analysis”. In: *International Journal of Forecasting* 15.4 (1999), pp. 353–375. ISSN: 0169-2070. DOI: https://doi.org/10.1016/S0169-2070(99)00018-7. URL: https://www.sciencedirect.com/science/article/pii/S0169207099000187.

8 Self Assessment

How closely did the implementation follow the initial project plan?

The implementation of this project followed the roadmap in the initial project plan, but all of the objectives were not met. While we successfully implemented a framework for eliciting the liquidation value of CRE, we did not reach all initial objectives due to falling behind schedule.

In what regard was the project successful?

The project was successful in delivering a generalizable framework for expert elicitation tailored to the client's needs. By selecting the IDEA protocol and designing a questionnaire based technique, we provided the client with a framework that improves the transparency and generalizability of expert judgment. Our framework is generalizable and transparent in each step which were in the core of the initial project plan.

We consider the literature review successful. It tackled the two main problems of the topic, elicitation method and real estate value estimation variables.

The project also improved the soft and technical skills of the team members. We got experience in giving presentations and working on a long-term project. We improved our understanding of the problems with dimensionality, as well as how to combat human biases with elicitation methods.

In what ways was the project less successful? While the framework proposed works as intended, the assumptions in it are restricting. Our model assumes that each variable determining the liquidation value is independent from each other, an assumption that cannot be justified in reality. This simplification was due to us unable to resolve the dimensionality problem with eliciting interactions of variables. Balancing between expert questionnaire fatigue and model assumptions was a difficult task that we could not solve. This results in the model not being able to capture complex interactions between variables.

What could have been done better? There are areas of improvement in our team dynamics and the course structure:

- **Project management and team roles:** Our team struggled with organization. A significant amount of time was used simply allocating and giving out tasks rather than focusing on the problem. The team lacked a strong project manager which lead to inefficiencies in the workflow as not everyone knew what they were supposed to do during the project. We should have given group members more long-term autonomy make the workflow of individuals more streamlined rather than consisting of individual tasks.

- **Course structure:** Before the project topic preferences were given, preferences for being a project manager could have been gathered. This way, no team is left without a suitable project manager. This also combats the problems of having multiple leader candidates in one team and no candidates in another.

- **Team spirit:** A lot of the project's success is dependent on the team spirit. We could have improved our team dynamics by holding more meetings in person rather than online. It would have been wise to have more meetings in the beginning to get the project going as quickly as possible.
- **Definition of success:** Defining success of a project would have been a good thing, from both the team's side as well as the client. In the first meeting(s) with the client, we should have asked what would a successful project look like, write that down and come up with a plan to reach that. This was done to some degree, but when progressing with the project the definition of success deviated a little too much. The team got new ideas, the client responded to our ideas and the original plan was left to little attention.

Overall, this project gave the team experience in working with a real client and gave us insight on building transparent models with expert judgment.