

Modelling Banking-hall Yield for Property Investment

Malvern Tipping

Department of Engineering and Built Environment,
Anglia Ruskin University, UK, and

Roger Newton

Statistical Consultant,
Chelmsford, UK

Abstract

Purpose – The study seeks to build a predictive model for the investment yield of British banking-halls.

Design/methodology/approach – Empirical data of similar lots sold at previous auctions are subjected to statistical analyses utilizing a cross-sectional research design. The independent variables analysed are taken from a previous study using the same cases. Models are built using logistic regression and ANCOVA.

Findings – Logistic regression generally generates better models than ANCOVA. A division of Britain on a north/south divide produces the best results. Rent is as good as lot size and price in modelling, but has greater utility, because it is known prior to auction.

Research limitations/implications – Cases analyzed were restricted to lots let entirely as banking-halls. Other lots comprising premises only partially used as banking-halls might produce different results. Freehold was the only tenure tested.

Practical implications – The study provides a form of predictive modelling for investors and their advisors using rent which is known in advance of any sale.

Originality/value – The study makes an original contribution to the field, because it builds a predictive model for investment yields for this class of property. Further research may indicate if similar predictive models can be built for other classes of investment property.

Keywords: Banking-hall; investment; portfolio; predictive framework; rent; yield; index.

Article classification: Research paper.

1. Introduction

1.1 Purpose

This paper investigates the predictor variables and statistical models with most utility for forecasting yield from UK banking premises auctioned for sale-and-leaseback. The paper has its origins in a doctoral thesis (Tipping 2011) but contains additional material.

Investors' appetite for investment property is encapsulated by the yield which it produces. The initial yield is the norm for valuing property income (Mackmin, 2009). The ability to make useful forecasts of yield in advance of auction has clear utility in investment decisions and portfolio construction. In addition, the ability to identify factors affecting yield is valuable for academic research.

1.2 Sale-and-leaseback

Banking-halls were originally made available as property investments through being sold by the retail banking companies on a sale-and-leaseback basis. This process was started by Barclays Bank in 1992. The very largest lots were generally sold off the market on a private treaty basis, often bundled into small portfolios. The remaining lots were in the main sold individually at auction. Although several of the banks have now divested themselves of most of their freeholds, these properties continue to be traded between investors at auction.

Tipping and Bullard (2007) give the following explanation of sale-and leaseback:

“The normative sale-and-leaseback transaction is one in which the owner of a property sells that property to a third party and simultaneously takes a lease on that property from the third party (Adams and Clarke, 1996). In other words, the original owner sells the property to an investor, who immediately becomes his landlord.”

2. Literature Review

With a scarcity of extant literature on factors influencing the investment yields of banking-halls, it is necessary to look further afield. Banking-halls fall into a sub-division of retail property, which is one of the three main types within commercial property.

Ambrose and Nourse (1993) suggest that property investment yields are influenced by the following attributes: property type; region; lot size as defined by price; time and possibly location at parochial level.

Since the findings of Ambrose and Nourse, it has been established that property should be treated as having different categories rather than being treated as a single asset class for the purposes of analysis. Chen *et al.* (2004), and MacGregor and Schwann (2003) demonstrate how investment yields are influenced by property type. Isaac and Steely (2000) argue for the application of the *efficient portfolio* (Markowitz, 1952) to property investment. Fraser (2004: 115) argues that through diversification even within a specific market sector, stock market investors can spread their risk by investing in several companies within that sector. Applying this construct to property, it follows that property occupiers operating within a specific sector may be perceived as having differing risk and

that that will in turn impact upon the yields of narrowly defined types of real estate. There is no reason to suggest that this applies any less to retail bank premises.

Hutchinson *et al.* (2011) argue that since specific types of tenant have a tendency to locate in particular premises in certain locations and are prone to enter into like lease forms, these factors will be collectively represented by investors' perceptions of tenants' strength of covenant. Moreover, it is suggested by Hutchinson *et al.* that the effects of tenants' strength of covenant will be compounded in times of economic downturn.

Although the literature supports the view that location influences property yields, there is a general lack of consistency about how location is defined and measured. Evans (1995) builds different levels into location. Ambrose and Nourse adopted five regions across the United States in their study. These might be termed mega regions when compared with the twelve regions in the United Kingdom as defined by the Government Office of the Regions (GOR). Cheng and Roulac (2007) identify region as being a major influence in the United States. However, a problem arises in how region is defined or measured in a different country. The raw data with respect to United Kingdom banking-halls use the category of region on the basis of the GOR definition. Byrne *et al.* (2013) suggest that an emphasized yield compaction in the Greater London area arises from investors' irrational appetite for investment in the capital.

Within the United Kingdom, the work of Hoesli *et al.* (1997) used regions loosely based upon the GOR definition in their study into influences on property yields. However, they found that although rents were influenced by region, there was not a clear effect upon yields based upon the GOR definition. They attributed this absence in the main to macro-economic influences. On the other hand, they did identify significant effects on yield when they defined location by dividing the country into three super-regions. Essentially, such a definition was based upon Greater London as a region and then two more regions founded on a north-south divide.

Guy and Henneberry (2000), Lee and Byrne (1998) also argue that commercial property investment yields across the United Kingdom are influenced by a north-south divide. The evidence in support of a north-south divide is not universal.

Enever and Isaac (2002) do not support a significant north-south divide on the basis of rent and growth. Rather, they attribute differences in rent and growth to location at a more parochial level. Their attributions are not directly with respect to yield.

Whitmore (2009) and Baum *et al.* (2006: 76) suggest that lot size, as defined by a property's capitalization, has an effect upon yield. Output from some of the research departments of the leading property consultancies supports this view. However, any influence attributable to lot size should be tempered by interactions with other factors (Whitmore).

The findings of Leone (2011) show that indices reflecting the United Kingdom's macro-economic cycle directly influence the performance of property held through the medium of real estate investment trusts (REITs). Property investment yields and other economic indicators fluctuate over time. The investment property yield cycle is highlighted by Krystalogianni and Tsolacos (2004). Numerous economic indicators are linked to the economic cycle which is spread over time. Dunse *et al.* (2007) show how time influences the yields of offices in the United Kingdom. Such cyclical fluctuations are market risk and differ from specific risk arising from the differing attributes of individual properties (Fraser, 2004).

The stock market's treatment of individual companies is a reflection of the possibility of those within any given sector performing differently. Such differentiation is specific risk. Baum *et al.* (2006: 246) state that the differing security of tenants should be considered during valuation. Since the capital value is a function of the income and the yield attributable to the risk, it follows that the specific tenant within a category of property may influence yield.

2.1 Discussion

From the body of the literature, likely factors influencing yield can be identified. These are property type, location, lot size, time and the security of the tenant. Location is subject to several levels and it is more difficult to identify the most appropriate level. However, there is a suggestion that super-regions or GOR regions might be the most appropriate levels.

Although Enever and Isaac prefer a more parochial level of location as being an influence on rent and growth, this does not necessarily exclude other levels from having an influence upon yield. Dunse *et al.* show how time affects offices. There is no reason to suggest that such an effect does not extend to retail property and therefore retail bank premises. Since specific risk necessitates that tenants' security is considered as affecting yield, such an approach should be applied to individual banking companies according to the relative strengths of their balance sheets.

Since the literature does not suggest a linear effect, interactions between two or more variables should not come as a surprise.

3. Methodology

3.1 Predictor Variables

Questionnaires, based on the results of semi-structured interviews, were used to identify those factors most likely as having a significant effect on yield. The participants in the interviews and respondents to the questionnaires were leading professional practitioners dealing with the sale by auction of retail banking premises as investments.

The following factors were identified for analysis:

- Twelve Government Office of the Regions (GOR) geographic regions. This factor was ranked first, having been the only factor to have been identified by all the respondents as being significant
- Property value. 91.7 per cent of respondents in the original study had identified price in the form of lot size group as being significant. However, actual price is not known until the fall of the hammer at the auction. In contrast, rent, which is highly correlated to price, is known beforehand. Therefore, this paper examines the utility of rent in addition to price and lot size.
- Banking company was also identified by 91.7 per cent of respondents as being significant
- Economic conditions. These vary over time. Such an economic cycle can be represented by either fitting an economic variable to different time zones or by using an appropriate index. Three potentially useful indices were identified: the IPD UK Retail Property Index (Investment Property Databank, 2008), which is a commercial index that must be purchased, and the DSH7 and DSG9 indices from the Office for National Statistics (ONS Business Investment Time Series) which relate to investment in real estate renting and business.

Tenure had been ranked as the next significant factor by the respondents in the original study. Since the terms of reference for the original study had specified that the study was to be limited to freehold properties, other tenures were excluded from that study. The paucity of cases with respect to other tenures also means that tenures cannot easily be analysed within the context of this paper.

3.2 Data Collection and Validation

Data were collected on lots sold by the banks on a sale-and-leaseback basis and such lots subsequently sold on by property investors (Tipping and Bullard, 2007: 208-210). All these lots were sold at public auction in London. The original full dataset comprised 1,012 cases based on hammer prices from July 1997 to July 2006.

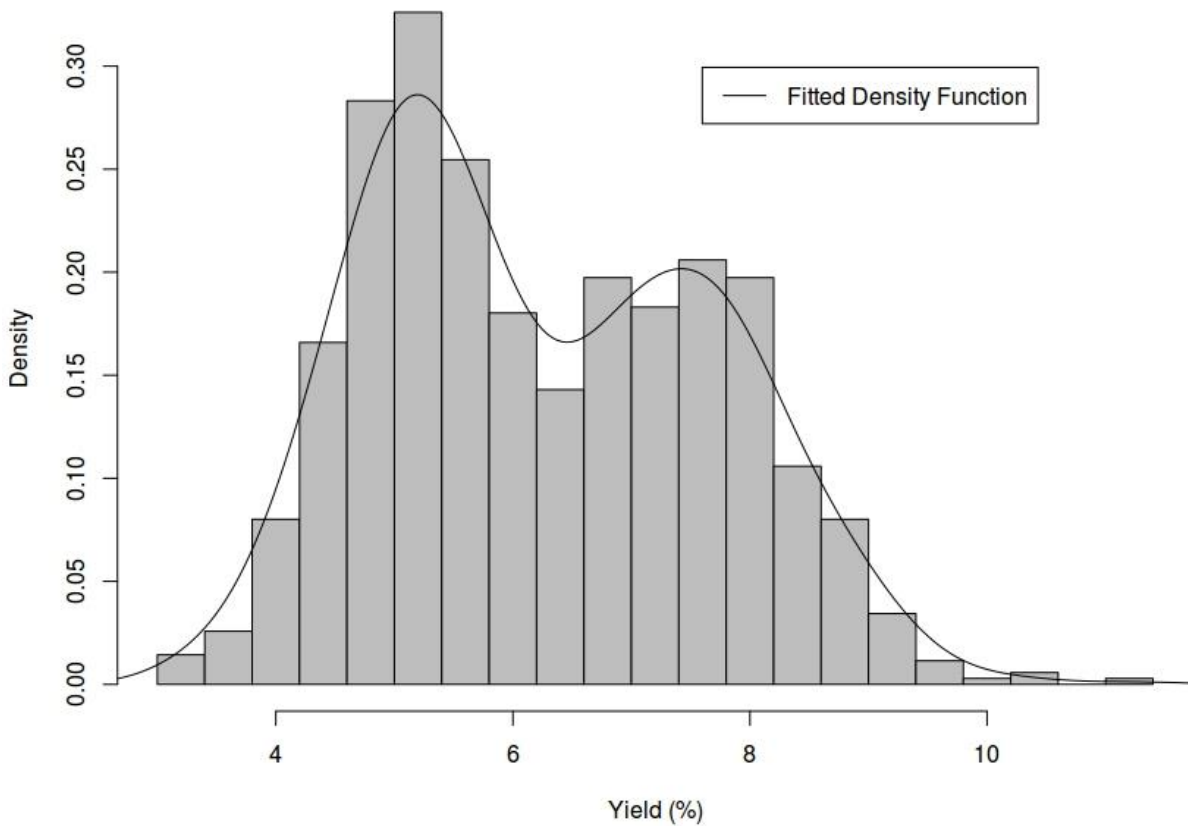
The London-based property auction houses use an eleven region geographical classification whose breakdown approximately corresponds with the GOR regions with the exception that the auction houses merge Yorkshire and Humberside into the North-east of England region. The eleven regions include Northern Ireland but, since no banking premises from this region were included in the study, a ten region auction house based geographical classification was used.

The initial data set was subjected to extensive quality control inspection. In particular the data accuracy of outliers was checked, some data entry errors were corrected and cases with missing data were removed. In addition, since the study was to be limited to freehold properties only, cases with other tenures were removed. Banks contributing particularly few cases (less than five) to the data set were also deleted. This left a Verified Freehold Dataset comprising 874 cases and involving 12 banks.

3.3 Exploratory Data Analysis

Exploratory analysis was carried out to elucidate the structure of the data set and to investigate the nature of the variables as identified in section 3.1 in preparation for the selection of appropriate forms of model to fit.

Figure 1 - Histogram of Yield for the Verified Freehold Dataset



A histogram of yield across the Verified Freehold Dataset (see Figure 1) showed a bimodal distribution. Yield could therefore be treated as either a continuous numeric variable or as a categorical variable with values divided between a low yield group (yield < 6.34%) and a high yield group (yield \geq 6.34%).

The data set involved 12 banks, 10 geographical regions and at least 3 levels of lot size (corresponding to the stamp duty levels) with two levels of yield (taking yield as a categorical variable). If economic activity was evaluated only on a yearly basis the eight years represented in the data (1997 to 2006 but excluding 1998 and 1999) required 8 values. A cross-tabulation of the data would then have 5760 ($= 12 \times 10 \times 3 \times 8 \times 2$) combinations of levels (cells). It is widely held (Agresti, 1996: 28, 34 and 190-194) that an average of at least 5 cases per cell is required for reliable model fitting. Since the data set contained only 874 cases the data was clearly too sparse and unbalanced for analysis in its existing form. The usual solution to this problem is to reduce the dimensionality of the data, typically by combining levels of categorical variables to produce fewer new levels. This, naturally, reduces the level of detail in the modelling.

In this case the number of geographic regions was reduced from ten to four. The new geographic regions were called “provinces”. Since no uniquely appropriate way of combining the original ten regions was apparent two different combinations were used and labelled Provinces A and Provinces B. The relationship between the original ten auction house regions and the breakdown into Provinces A and Provinces B are shown in Tables 1 and 2.

Table 1 *Regions combined as Provinces A*

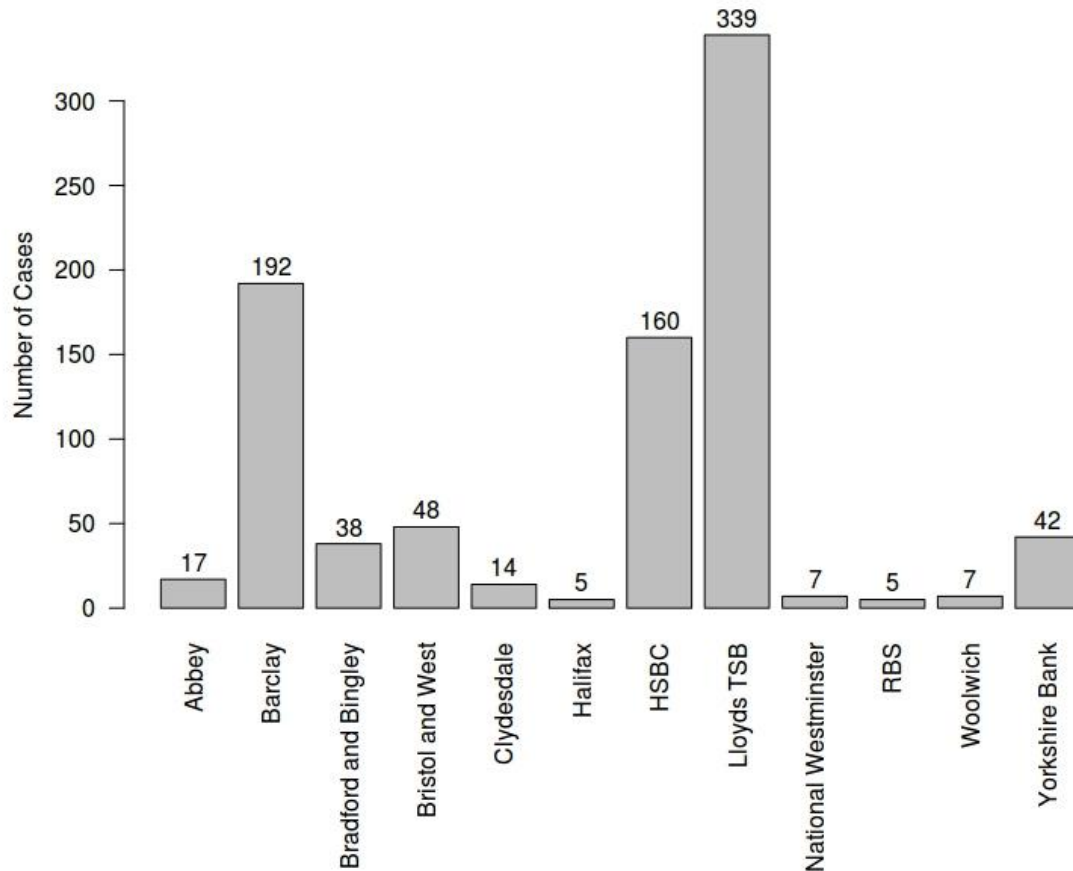
Regions	Provinces A
London-M25 South-East South-West	South
East Anglia West Midlands East Midlands	Midlands
North-East North-West	North
Wales Scotland	Celtic

Table 2 *Regions combined as Provinces B*

Regions	Provinces B
London-M25 South-East	London & South-East
South-West Wales	Wales & South-West
East Anglia West Midlands East Midlands	Midlands
North-East North-West Scotland	North Britain

Inspection of the number of cases per banking company (see Figure 2) showed that three banks (Barclay, HSBC and Lloyds TSB) provided 79% of the cases in the Verified Freehold Dataset. The study would therefore proceed using only the data for those three banks. This provided a Three Banks Dataset containing 691 cases.

Figure 2 - Number of Cases per Banking Company in the Verified Freehold Dataset



The number of levels of the lotsize variable was reduced from three to two by combining the lowest original two levels leaving a final two level breakdown with low lotsize (price \leq £500,000) and high lotsize (price $>$ £500,000).

In models where economic conditions were to be represented by fitted variables representing different time periods it was decided to use only two periods: early (calendar years 1997 to 2002 inclusive) and late (calendar years 2003 to 2006 inclusive).

Retaining two levels for yield the cross tabulation cell count became $3 \times 4 \times 2 \times 2 \times 2 = 96$. The 691 entries in the Three Banks Dataset therefore provided enough cases to carry out reasonable analyses with the data structured in this way.

3.4 Final Selection of Variates

Based on the previous discussion the variables chosen for inclusion in the model fitting were as follows.

The dependent variable: **yield** taken as either a continuous variable or as a categorical variable with the values divided between low and high yield groups.

The predictor variables:

(1) **Geographic Area** as represented by four ‘provinces’ derived from the original auction house classification. Two different sets of provinces were used denoted Provinces A and Provinces B.

(2) **Property Value** represented by either rent, price or lotsize. Rent and price are continuous variables while lotsize is categorical.

(3) **Economic Conditions** as represented by the IPD index, the ONS DSH7 and DSG9 indices or by fitted values over early and late time periods.

(4) **Banking Company**. A categorical variable taking one of three values: LTSB, Barclays and HSBC.

3.5 Model Selection

The types of statistical model which can sensibly be fitted to the data are determined by the characteristics of the dependent and descriptor variables and, in particular, by whether the dependent variable is numeric or categorical.

Since the dependent variable, yield, could be taken as either continuous or categorical, two forms of analysis could be undertaken: Logistic Regression with yield categorical and Analysis of Covariance (ANCOVA) with yield continuous. Logistic Regression calculates the probability that a particular case will fall into the high yield group. ANCOVA predicts a numerical yield value.

With four ways of representing economic conditions, three ways of representing property values and two different geographical breakdowns there were 24 possible model structures to be examined using Logistic Regression and ANCOVA. A total of forty eight models in all.

The dataset used was the Three Banks Dataset comprising 691 cases. Note that every model included banking company as a variate.

The overall analysis was structured hierarchically. At the top level the work was divided into Logistic and ANCOVA analyses. Within each of those the study was performed separately for Provinces A and Provinces B. Within each province the utility of each representation of property values was examined. Finally, the effects of different representations of economic conditions were considered within each of the previous subdivisions. Results were evaluated from the bottom up i.e. best economic indicator within each property value type then best indicator-property value combination within each geographic classification etc.

3.6 Model Fitting and Validation

The 48 models were fitted using the ‘R’ Statistical Software Package (R Core Team, 2013). Two approaches were used in model development. The statistical significance of variables was initially evaluated using the Akaike Information Criterion (Akaike, 1974). Models were simplified by the

stepwise elimination of terms from an initial saturated model containing all possible terms. Thereafter the statistical significances of remaining terms were examined using F-tests and terms not significant at at least the 10% level were removed. The results of removing some additional statistically significant terms were examined to see if this had a practical effect on model utility as indicated by visibly poorer performance as seen in the ROC charts (see section 3.7).

Viewed as categorical data, the 691 case Three Banks Dataset was too small to be split into separate model-fitting and model-validation data sets. Two levels of property value, four geographic regions, three banks, two levels of economic activity and two levels of yield gives a total of 96 categories. With only 691 cases this gives an average of only slightly more than the minimum number of five cases per category (Agresti, 1996: 28, 34 and 190-194) needed for reliable analysis. Model validation was therefore undertaken using a cross-validation method. The data were split into 10 mutually exclusive groups. Each group in turn was used as the validation group. The model was fitted using the data from the other 9 groups and the fitted model was used to predict the results for the validation group. In this way the predicted values do not depend in any way on the actual values of the predicted group. All data presented in what follows are from the validated results.

3.7 Evaluation of Results

Three criteria were used in choosing between models.

- Performance: which model gave the most accurate predictions when comparing actual and predicted yields.
- Parsimony: for equal performance which model contained the fewer terms.
- Utility: with equivalent performance and parsimony the model which was the easiest to use and generalise was preferred.

For logistic models a useful and informative way of displaying the results is as a confusion matrix. Logistic regression gives a probability that each test case will fall into the high yield group. By setting a threshold on this probability each case can be assigned to the high yield group (if the calculated probability exceeds the threshold) or the low yield group. The actual and predicted numbers of cases assigned to each group can be inspected using a simple 2-by-2 confusion matrix. The example shown in Table 3 is for the Logistic Provinces A/Rent/IPD Index model with a threshold at 0.5 (50%).

Table 3 - Confusion Matrix for the Provinces A/Rent/IPD Index Model

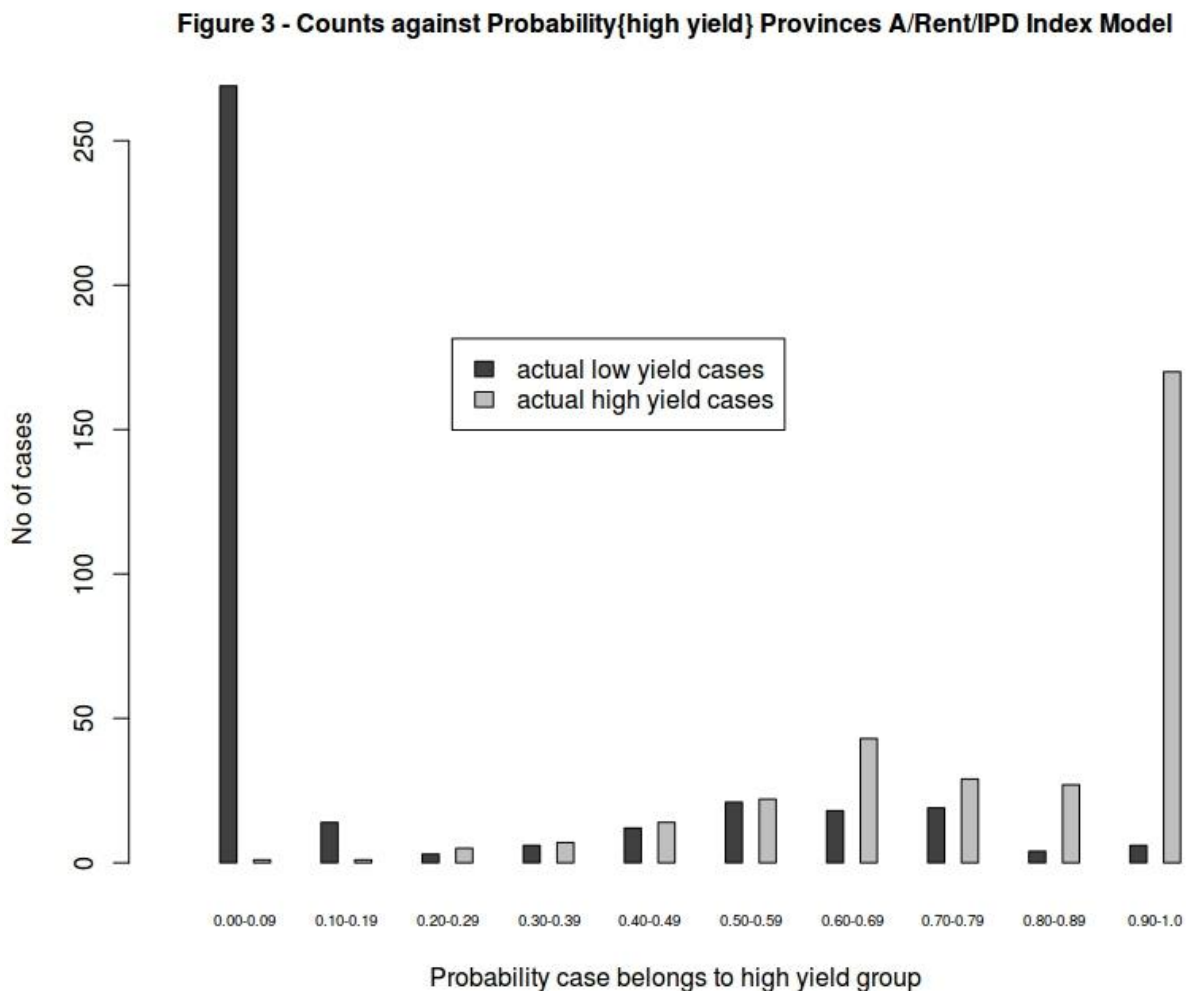
		Actual Yield Group	
		Low	High
Predicted Yield Group	High	68	291
	Low	304	28

In order to be useful for investment purposes, certain things are required of the results.

- Of those cases predicted as high yield the proportion truly having high yield must be large. This proportion is called the Positive Predicted Value (PPV).

- Similarly of those cases truly having high yield the proportion predicted as high yield should be large. This proportion is called the True Positive Rate (TPR).

A confusion matrix gives the PPV and TPR values for a given threshold but the matrix (and the corresponding PPV and TPR values) will change with the threshold value as is clear from the example bar chart of probability values shown in Figure 3.



A more informative way of examining the results from a logistic model is to plot the PPV values against the corresponding TPR values in a variation of the Receiver Operating Characteristic (ROC) chart. Figure 4 shows the ROC chart for the Logistic Provinces A/Rent/IPD Index model. The shaded area at the top is the critical region selected for this study. Values in this region represent usefully high PPV values (and corresponding thresholds) on which property investment decisions might be based. The further to the right in this area the higher the TPR value and the larger the population of useful predictions. The threshold value increases along the curve from lower right to upper left. The perfect result would be a value in the top right hand corner (PPV = 1.0, TPR = 1.0) which would mean that the predicted yield group membership perfectly matched the actual yield group membership.

ANCOVA results were generated in an analogous manner. ANCOVA produces predicted yield values which may be plotted against actual yield values. Setting a threshold on actual and predicted

yield divides the graph into four sections (see Figure 5). The numbers in each section give a matrix analogous to the confusion matrix and a pseudo-ROC chart (Figure 6) may be generated by varying the threshold level. Care is necessary in interpreting results in this form. Only a certain middle range of thresholds is useful in this situation. For example: too low a threshold will place all the results in the high-high category giving $PPV=TPR=1.0$. Similarly too high a threshold will put every case into the low-low category which is an equally uninformative result. The threshold range over which results may be sensibly interpreted is indicated by the phi coefficient which takes values in the range -1 to +1 (see Sheskin, 2003: 534). In this study results are presented only for thresholds which gave phi of at least 0.5. The two cases of extreme thresholds noted above would give phi values of zero.

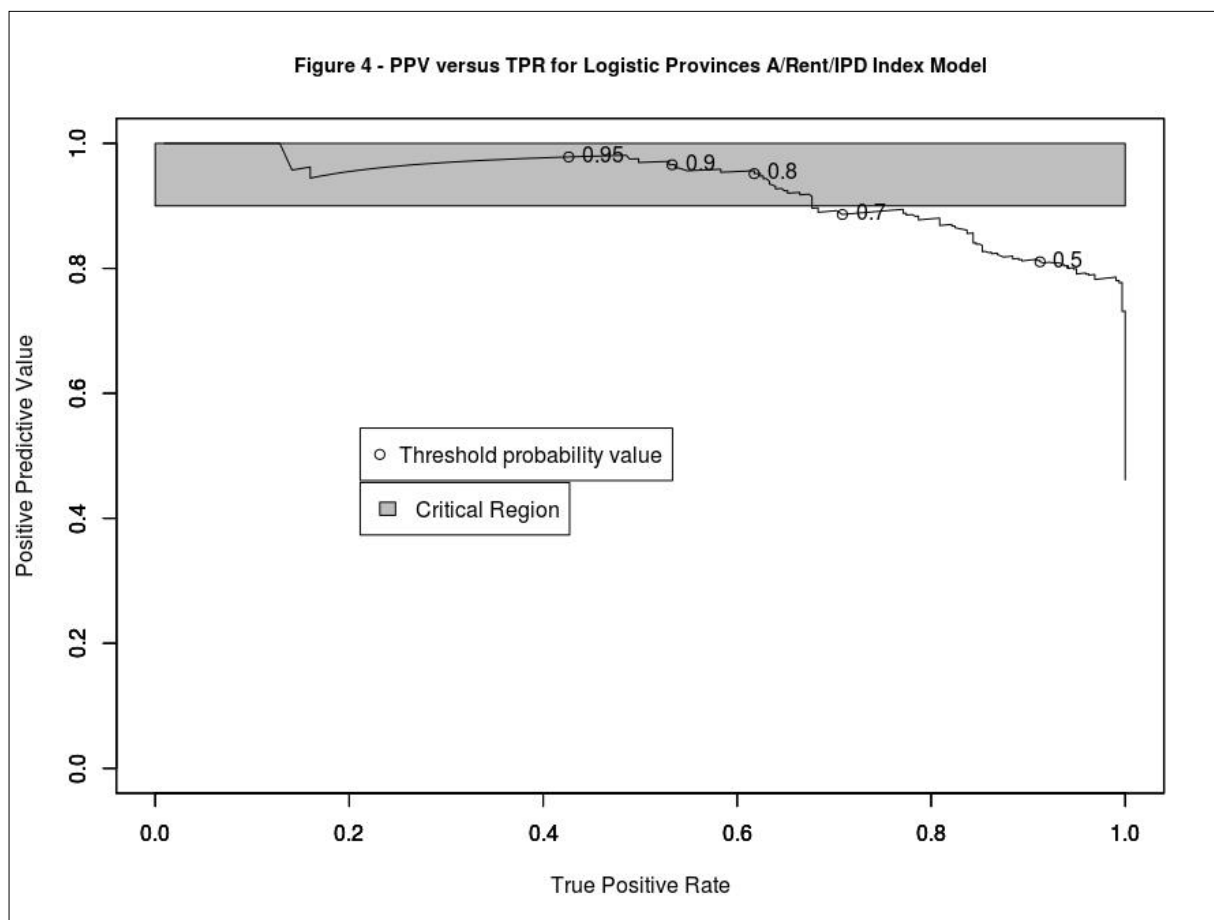


Figure 5 - Predicted Yield against Actual Yield - Provinces A/Rent/IPD Index Model

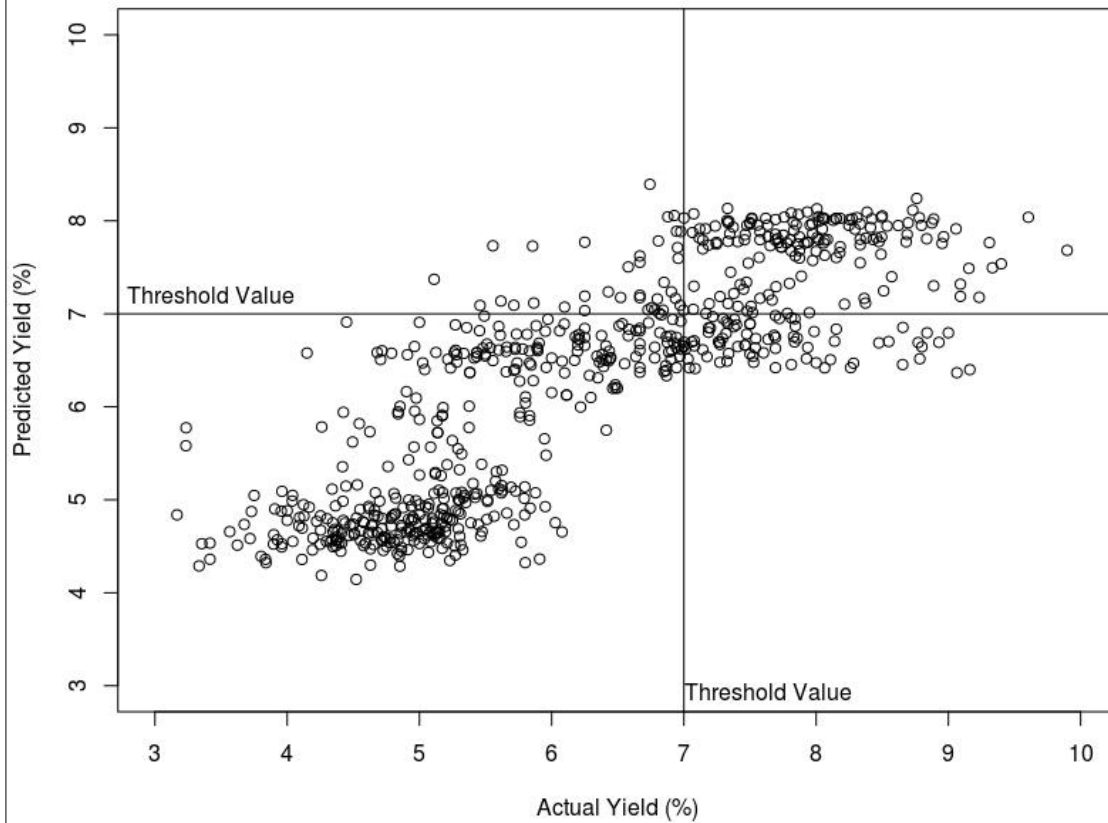
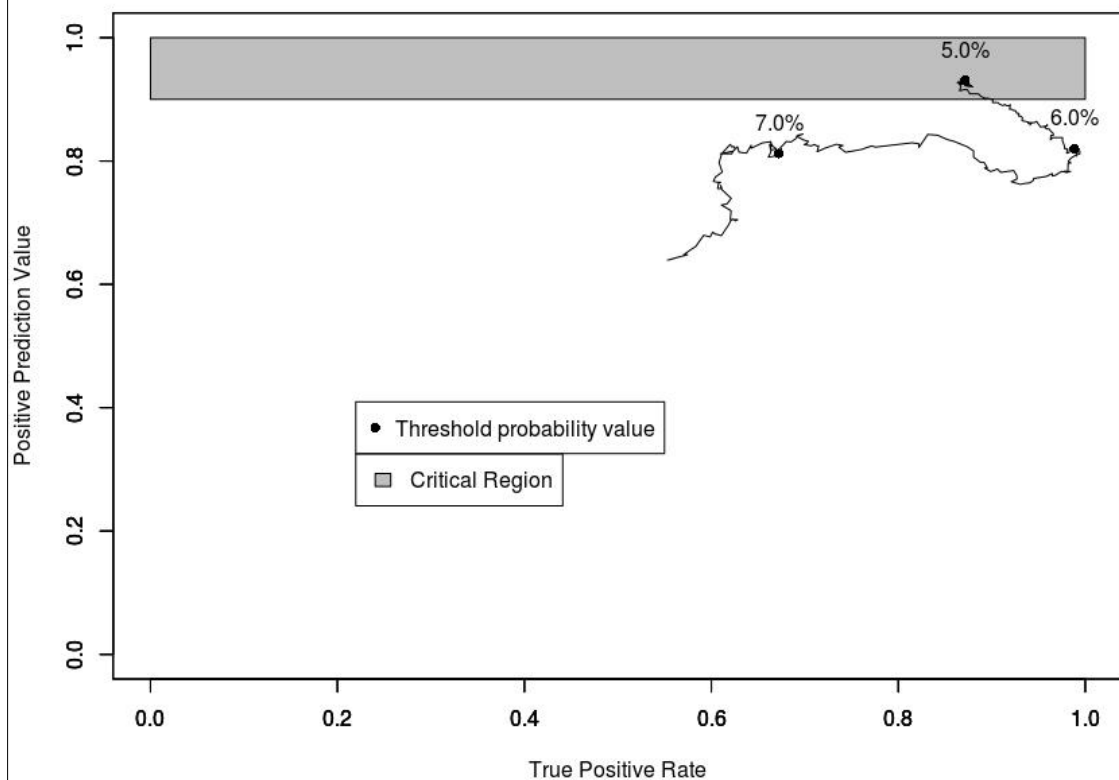


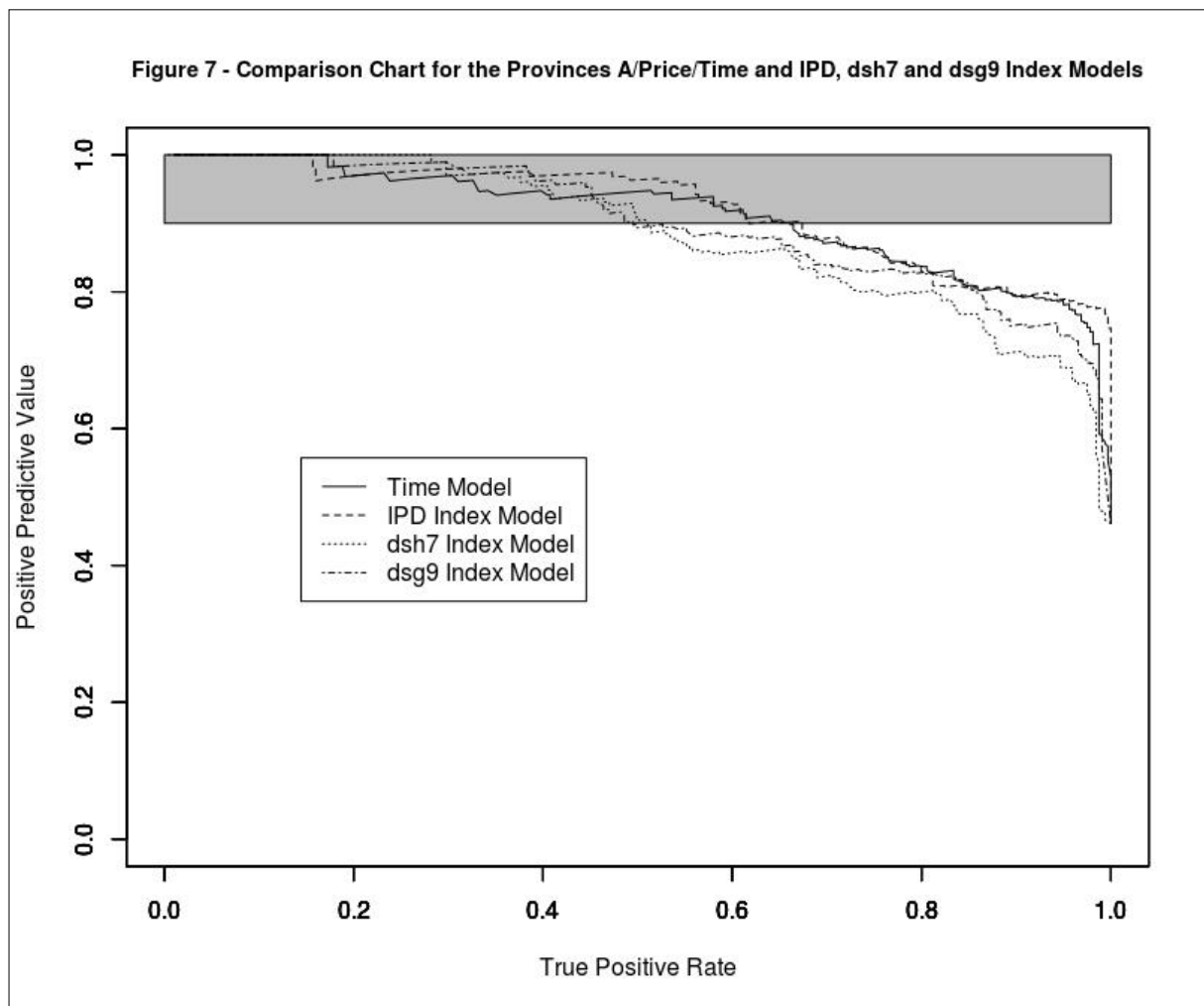
Figure 6 - PPV versus PTR for ANCOVA Provinces A/Rent/IPD Index Model



4. Results

4.1 Introduction

It was almost impossible to choose between models on the grounds of performance. One model was hardly ever uniformly better than another i.e. one ROC chart graph was always higher and to the right of the other. This is illustrated in Figure 7 which compares the ROC chart results for the time and IPD, DSH7 and DSG9 index sub-models for the Logistic Provinces A/Price model grouping.

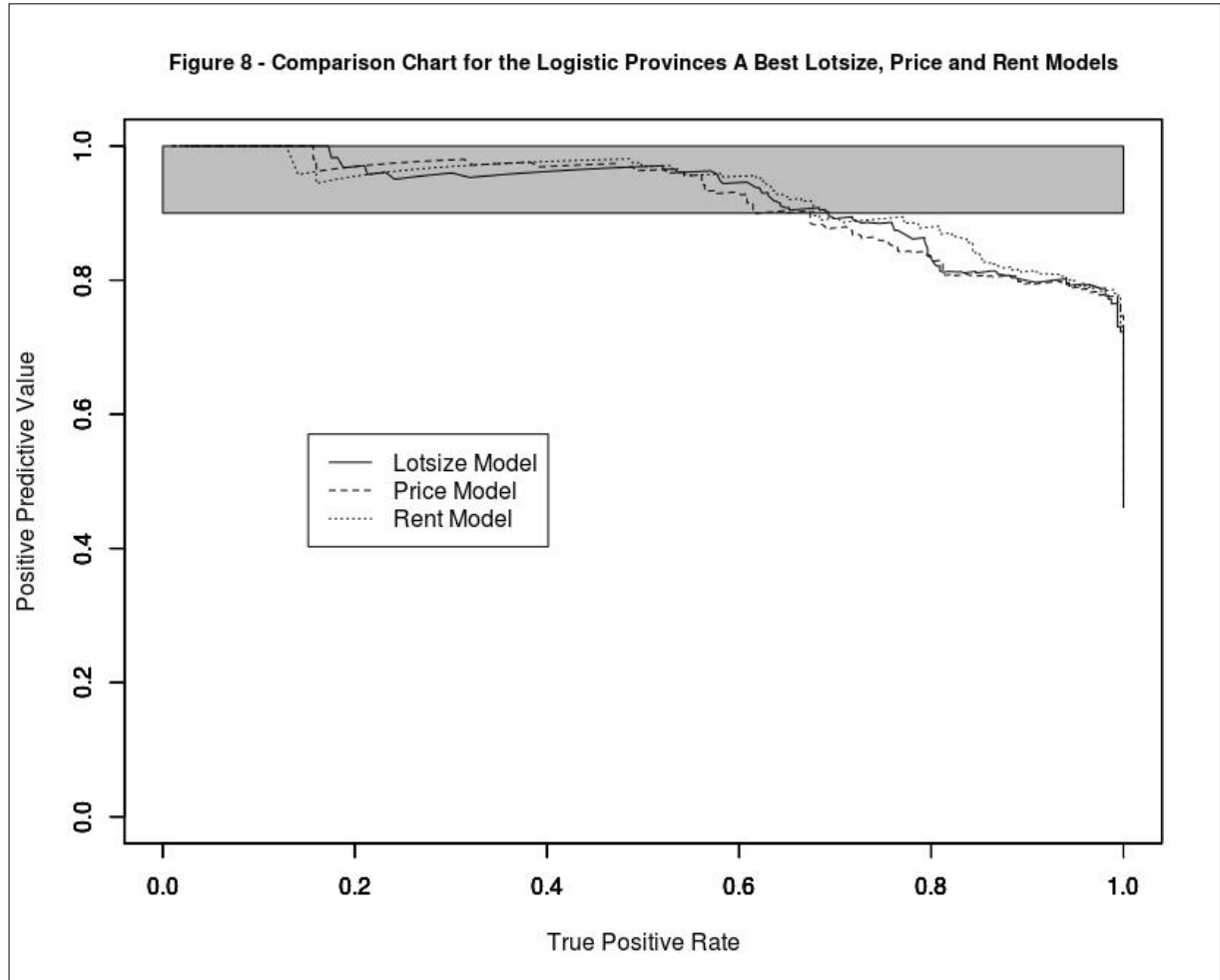


Parsimony, therefore, most frequently provided the basis for model selection. For the models in Figure 7 the IPD index model comprised only three terms (excluding the constant) while the time, DSH7 and DSG9 models respectively had five, six and eight terms. Almost without exception the best choice among the economic condition variables was the IPD index on the grounds of model parsimony. Even when not clearly the outright best choice, on this basis the IPD index was arguably no worse than any of the alternatives.

4.2 Logistic Analysis – Provinces A

Figure 8 shows the results for the best lotsize, price and rent models for the Logistic Provinces A analyses. Each of these models incorporates the IPD index to represent economic conditions. It is

again difficult to discriminate between the models on the basis of performance and all three models comprise three terms. The rent model was therefore preferred on the grounds of utility since the rent value is available before auction and the price and lotsize values only become available afterwards. The lotsize and price models are good explanatory models while the rent model is far more useful for prediction or forecasting.



For the Logistic Analysis using the Provinces A geographic classification, the best model was of the form:

$$\text{Yieldgroup} \sim \text{Provinces A} + \text{Rent} + \text{IPD index}$$

The banking company variate was eliminated completely from the model and only the linear terms in IPD index, rent and provinces A were retained. All interaction terms were rejected as not significant either statistically or practically.

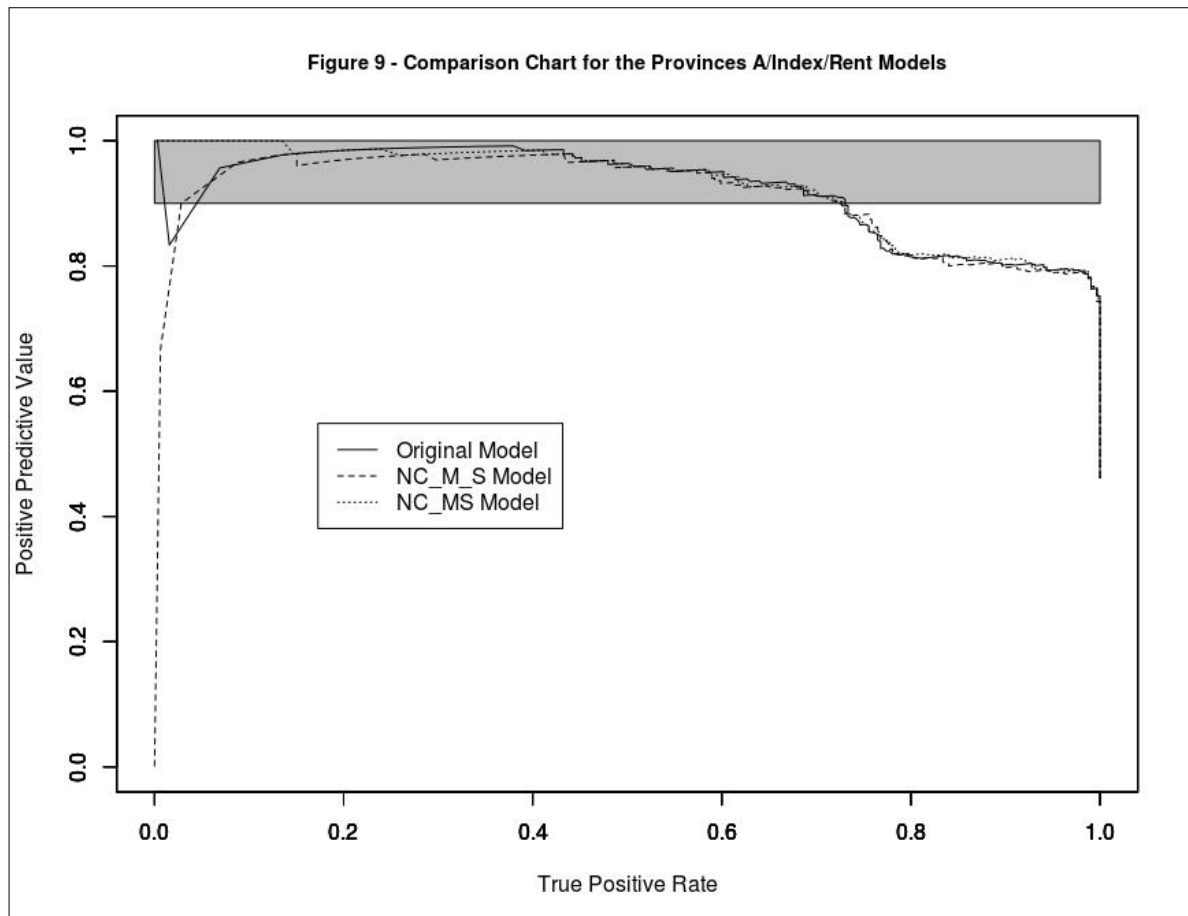
Examination of the coefficients for the models (see Table 4) suggested that there was little difference between some of the geographic provinces in the Provinces A classification.

Table 4 – Coefficients for the Logistic Provinces A/Rent/IPD Index Model

Coefficients	Estimate	Std. Error	z value	Pr(> z)	Signif. Level
(Intercept)	-2.450e+01	2.739e+00	-8.946	< 2e-16	0.1%
Index	3.859e+00	3.885e-01	9.934	< 2e-16	0.1%
ProvincesAMidlands	-2.467e+00	7.323e-01	-3.369	0.000756	0.1%
ProvincesANorth	-9.414e-01	8.008e-01	-1.176	0.239768	>10%
ProvincesASouth	-3.208e+00	7.158e-01	-4.482	7.38e-06	0.1%
Rent	2.668e-05	5.537e-06	4.818	1.45e-06	0.1%

Investigation showed differences between the combined North and Celtic provinces and between the combined South and Midlands provinces to be statistically significant at beyond the 0.1% level. The difference between the Midlands and South was significant at the 5% level while the difference between North and Celtic regions was not significant at the 10% level. Two alternative geographic breakdowns were therefore examined.

The North and Celtic provinces were combined to give the UK broken down into three super provinces (Northceltic, Midlands and South). The South and Midlands were also combined to give a two super province breakdown (Northceltic and Midlandssouth). Examination of the ROC chart (Figure 9) shows little difference in performance between the three models incorporating the different geographic descriptions. The simplest two area model was preferred by parsimony.



Overall, the best model from the 12 in the Logistic/Provinces A section of the analysis was of the form:

$$\text{Yieldgroup} \sim \text{SuperprovincesA} + \text{Rent} + \text{IPD index}$$

where SuperprovincesA was a categorical variable whose super provinces were the combined South and Midlands provinces of Provinces A and the combined North and Celtic provinces of Provinces A.

Table 5 shows the model coefficients.

Table 5 - Coefficients for the Logistic Superprovinces A Best Model

Coefficients	Estimate	Std. Error	z value	Pr(> z)	Signif. Level
(Intercept)	-2.478e+01	2.641e+00	-9.384	< 2e-16	0.1%
Index	3.815e+00	3.806e-01	10.024	< 2e-16	0.1%
SuperprovincesA-SouthMidlands	-2.337e+00	4.087e-01	-5.719	1.07e-08	0.1%
Rent	2.520e-05	5.354e-06	4.706	2.52e-06	0.1%

4.3 Logistic Analyses - Provinces B

The results for the Logistic Analyses using the Provinces B geographic breakdown were generally similar to those for Provinces A. Overall, the preferred model again included the IPD index and rent variates and was of the same form:

$$\text{Yieldgroup} \sim \text{Provinces B} + \text{Rent} + \text{IPD index}$$

as the equivalent Provinces A model.

Further investigation of the Provinces B geographic classification showed that combining the London, Midlands and Wales provinces into a single super province did not reduce the practical utility of the model and so the final best Logistic Provinces B model was of the form:

$$\text{Yieldgroup} \sim \text{SuperprovincesB} + \text{Rent} + \text{IPD index}$$

where SuperprovincesB is a categorical variate whose super provinces were the original Northern province of Provinces B and the combined London-Midlands-Wales provinces of Provinces B.

The coefficients are shown in Table 6.

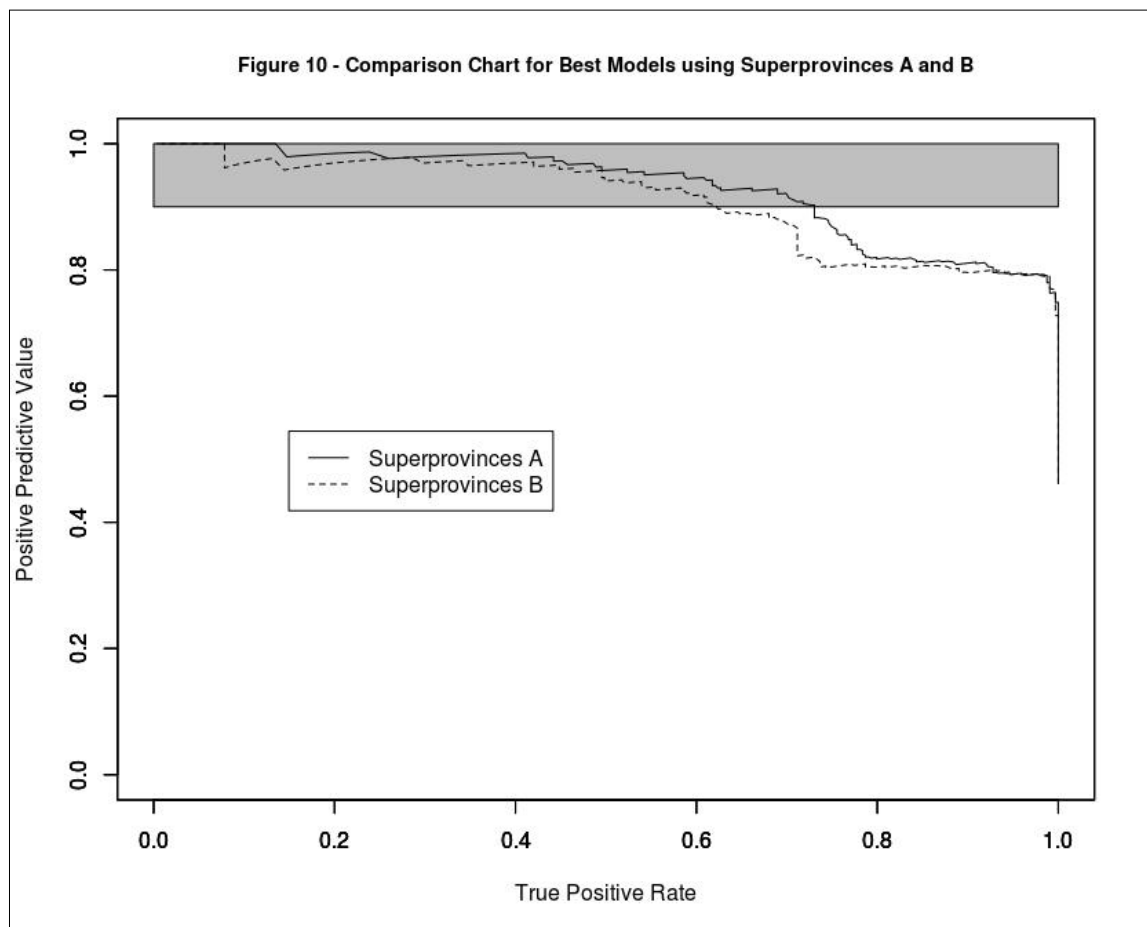
Table 6 - Coefficients for the Logistic Superprovinces B Best Model

Coefficients

	Estimate	Std. Error	z value	Pr(> z)	Signif.Level
(Intercept)	-2.467e+01	2.611e+00	-9.450	< 2e-16	0.1%
Index	3.821e+00	3.801e-01	10.051	<2e-16	0.1%
SuperprovincesB	-2.233e+00	4.784e-01	-4.667	3.05e-06	0.1%
South Britain					
Rent	2.293e-05	5.283e-06	4.341	1.42e-05	0.1%

4.4 Logistic Analyses - Provinces A and Provinces B Comparison

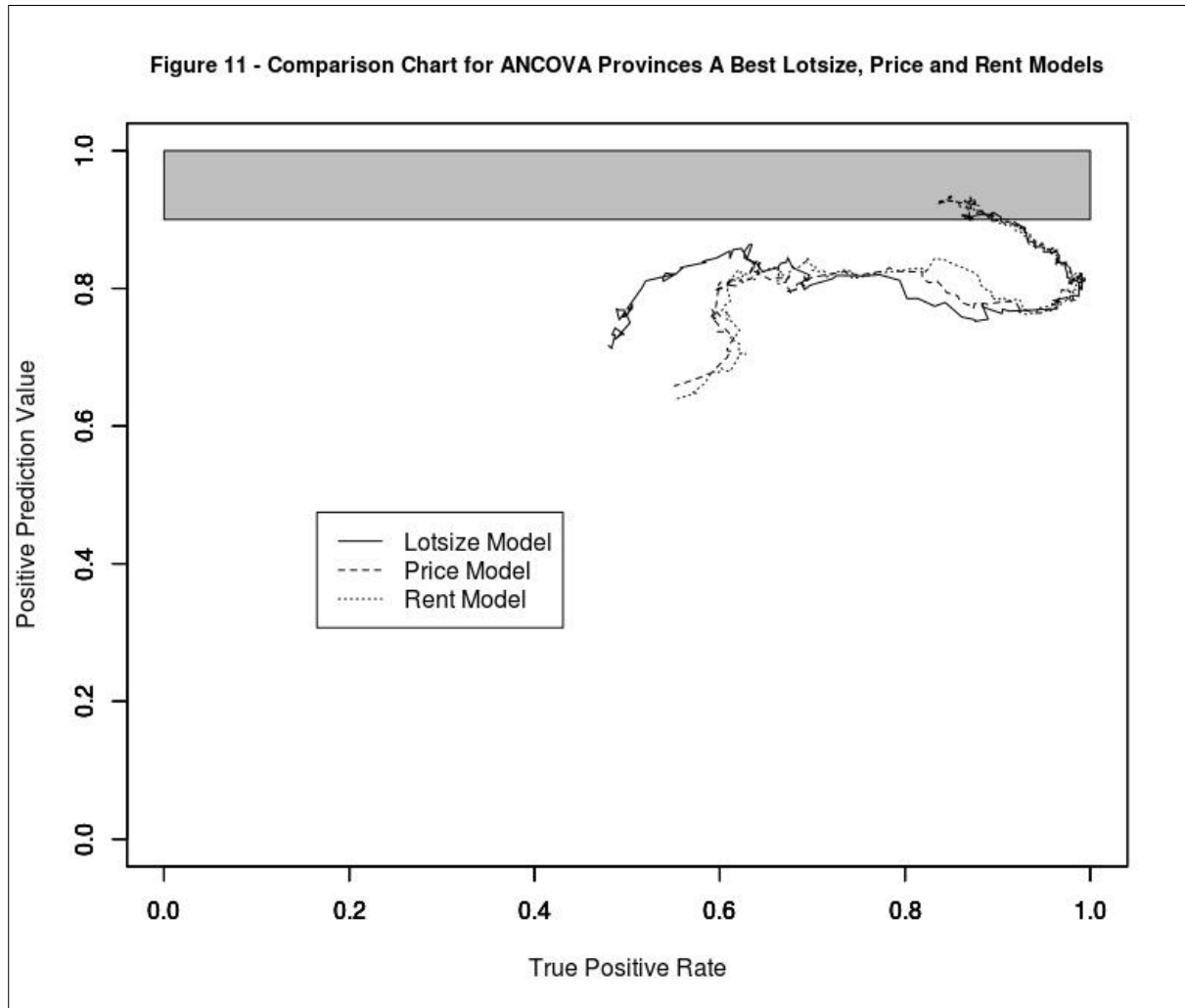
Figure 10 shows the ROC chart results for the overall best logistic models for Superprovinces A and Superprovinces B. Unusually for this study, the results from Superprovinces A are seen to be almost uniformly better (and always no worse) than Superprovinces B. Although the differences are generally quite small, the Superprovinces A geographic breakdown is preferred.



The Superprovinces A and Superprovinces B geographic structures clearly indicate a North - South divide in property investment returns with northern locations giving the better yields.

4.5 ANCOVA Provinces A and Provinces B

The ANCOVA analyses for both Provinces A and Provinces B gave very poor results. None of the models produced predictions which were judged to be particularly valuable for sound investment decisions making. Figure 11 illustrates this point for the best lotsize, price and rent models for Provinces A. (All three models used the IPD index to represent economic conditions.) The graphs lie almost entirely outside the critical region.



The results for Provinces B were similar.

Overall, the best Provinces A model was again preferred on performance grounds. This model has the form:

$$\text{Yield} \sim \text{Provinces A} + \text{Bank} + \text{Rent} + \text{IPD index} + \text{Bank:Provinces A}.$$

The Analysis of Variance Table is given at Table 7 and the coefficients are given in Table 8.

Table 7 - Analysis of Co-variance Table Best ANCOVA Model (Provinces A)

Variable

	Deg. of Freedom	Sum of Squares	Mean Square	F value	Pr(>F)	Signif. Level
Index	1	921.57	921.57	1482.1892	< 2.2e-16	0.1%
Bank	2	7.91	3.96	6.3640	0.0018274	1%
ProvincesA	3	39.33	13.11	21.0868	4.549e-13	0.1%
Rent	1	32.78	32.78	52.7237	1.058e-12	0.1%
Bank:ProvincesA	6	15.97	2.66	4.2814	0.0003021	0.1%
Residuals	677	420.94	0.62			

Residual standard error: 0.7885 on 677 degrees of freedom

Multiple R-squared: 0.7074

Adjusted R-squared: 0.7018

F-statistic: 125.9 on 13 and 677 DF,

p-value: < 2.2e-16

Table 8 - Coefficients for the ANCOVA Provinces A/Bank/Rent/IPD Index Model

Coefficients

	Estimate	Std. Error	t value	Pr(> t)	Signif. Level
(Intercept)	-5.268e-01	4.143e-01	-1.272	0.2039	>10%
Index	1.054e+00	4.084e-02	25.810	< 2e-16	0.1%
BankHSBC	1.745e-01	3.929e-01	0.444	0.6571	>10%
BankLTSB	5.570e-01	3.810e-01	1.462	0.1443	>10%
ProvincesAMidlands	-3.051e-01	3.731e-01	-0.818	0.4138	>10%
ProvincesANorth	-2.783e-01	3.795e-01	-0.733	0.4635	>10%
ProvincesASouth	-3.979e-01	3.604e-01	-1.104	0.2700	>10%
Rent	7.225e-06	1.000e-06	7.222	1.38e-12	0.1%
BankHSBC:	3.301e-01	4.398e-01	0.751	0.4531	>10%
ProvincesAMidlands					
BankLTSB:	-5.586e-01	4.004e-01	-1.395	0.1634	>10%
ProvincesAMidlands					
BankHSBC:	2.328e-01	4.270e-01	0.545	0.5858	>10%
ProvincesANorth					
BankLTSB:	-1.843e-01	4.140e-01	-0.445	0.6563	>10%
ProvincesANorth					
BankHSBC:	1.472e-01	3.983e-01	0.370	0.7118	>10%
ProvincesASouth					
BankLTSB:	-7.531e-01	3.848e-01	-1.957	0.0507	10%
ProvincesASouth					

4.6 Predictive Validation

The models developed in this study were validated using the cross-validation technique described in section 3.6 above. As an additional study the predictive validity of the logistic and ANCOVA models was further examined by applying the equations to a new additional set of data comprising 110 bank premises auctioned in 2006 - 2007.

Yields overall had fallen to such an extent by this period that only two of 110 cases produced values which fell into the high yield group (yield $\geq 6.34\%$).

Both the logistic and ANCOVA models could still be used as a basis for investment decisions, however, by selecting those cases with the highest predicted probabilities of being in the high yield group (for the logistic model) and selecting for the highest predicted yields in the ANCOVA model.

Figure 12 shows a histogram of the distribution of the yields for the validation sample and, superimposed, the distribution of the top 11 investment prospects as selected using the best logistic model. Clearly the selected investment prospects are better, on average, than a random sample from the original population. The mean yield for the population is 4.82% whilst that for the selected investments is 5.44%. The difference is statistically significant at the 1% level.

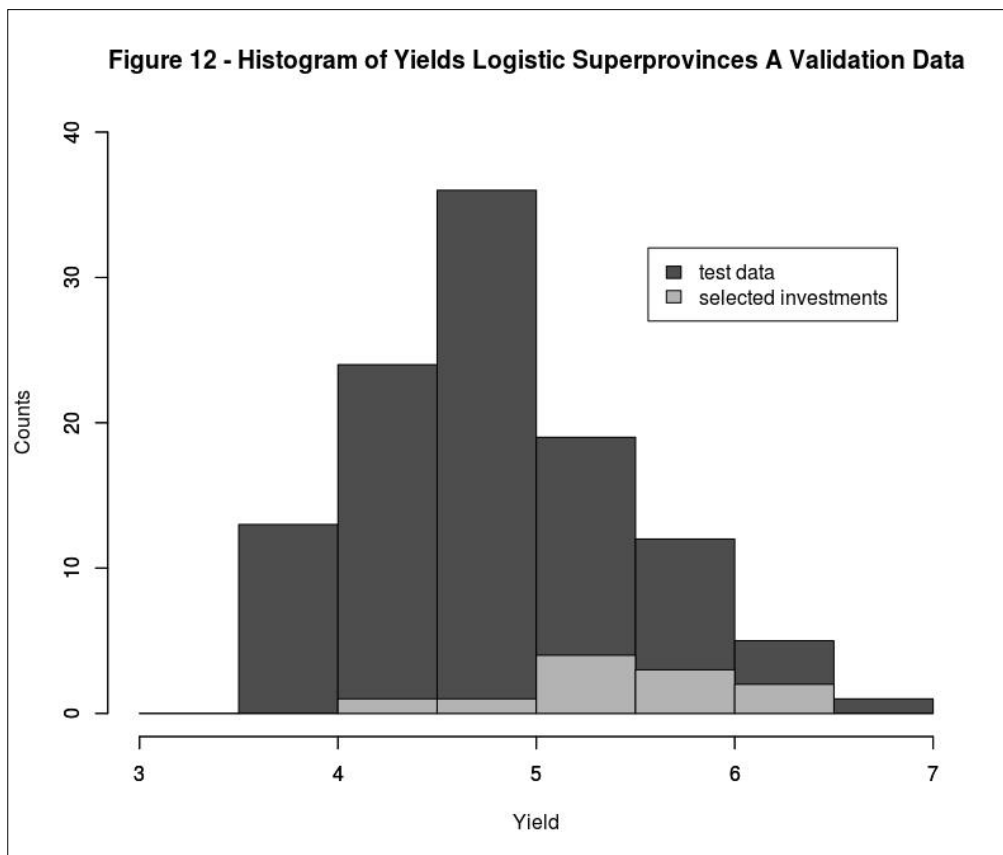
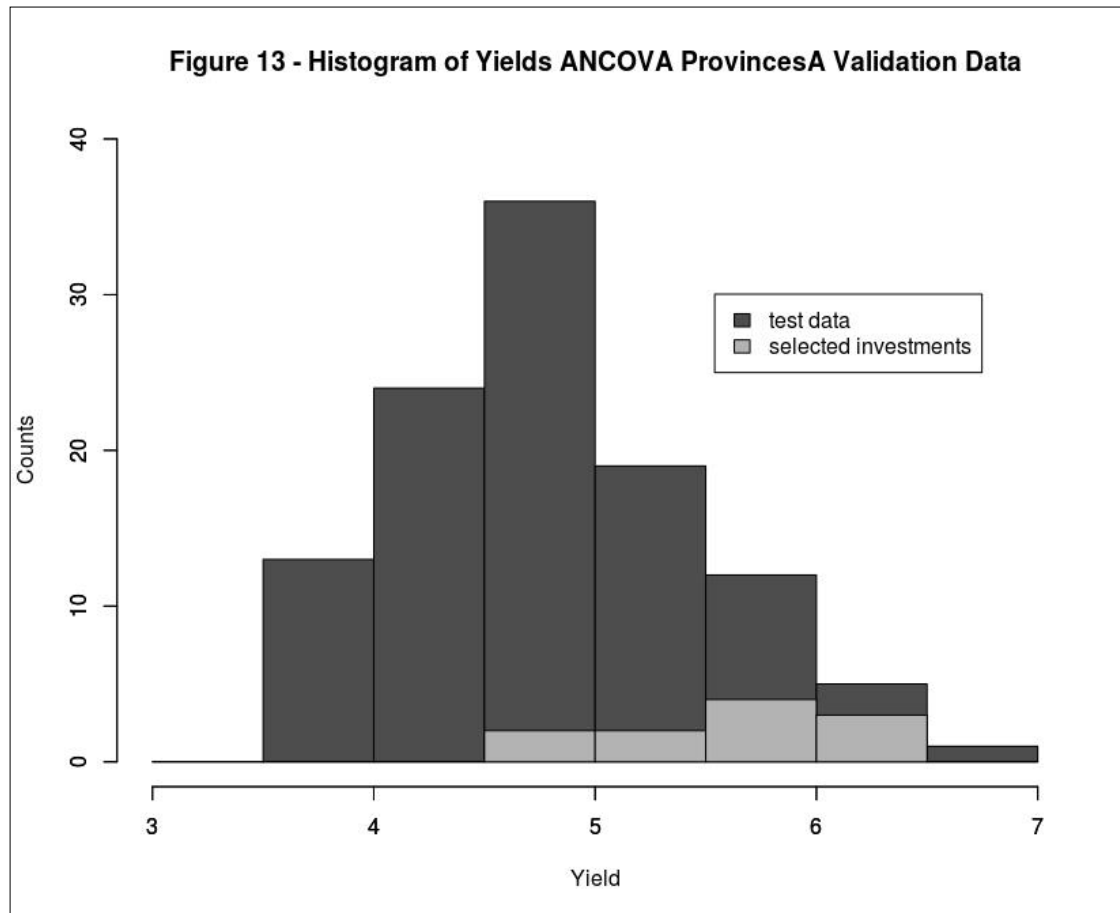


Figure 13 shows the equivalent results for the best ANCOVA model. Again the selected investment prospects are an improvement on the original population. The population mean yield is 4.82% and the sample mean is 5.64%. The difference is statistically significant at the 0.1% level.



5.0 Conclusions

The analysis showed that it is feasible to usefully forecast yields produced by banking-hall premises based on geographic location, rent and economic conditions and, possibly, banking company.

Better short-term forecasts were obtained using logistic models (with yield as a categorical variable) than with ANCOVA models (with yield as a continuous variable). The ANCOVA results, however, generalised better to a very different economic climate where the original breakdown of yield into high and low yield groups had lost its utility.

The Provinces A geographic breakdown generally gave better results than Provinces B. The super provinces classification revealed a clear North/South divide with yields being higher from properties in the northern areas. The Provinces A combination had originally grouped Wales with Scotland into a Celtic grouping, whereas the Provinces B combination had initially grouped Wales with the south-west of England. It was found that the most useful model was derived from combining the Provinces A North and Celtic regions and placing Wales in a super province comprising Wales, Scotland and the north of England. No loss in utility was incurred by simplifying the geographic breakdown to only two super provinces.

Economic conditions were equally well represented by the economic index variables and by the fitted values for two different time periods. The specially tailored IPD index generally gave simpler models than either of the two more general ONS indices tested (DSH7 and DSG9). The IPD index must, however, be purchased while the ONS indices were available at no cost to the user.

The use of property rent to represent property value gave results generally as good as those obtained using price or lot size for this purpose. Rent has greater utility than the other variables since it is available before auction rather than afterwards. Banking-halls let at higher rents were found to attract higher yields than those let at lower rents.

Banking company was found to have no significance in the best logistic models identified in the study. It did appear along with a bank/provinces interaction term in the better ANCOVA models.

The best models found were:

Logistic:	$\text{yieldgroup} \sim \text{Super Provinces A} + \text{Rent} + \text{IPD Index},$
ANCOVA:	$\text{yield} \sim \text{Provinces A} + \text{Bank} + \text{Rent} + \text{IPD Index} + \text{Provinces A:Bank}.$

6.0 Discussion

6.1 Statistical and Practical Significance

There is a difference between statistical significance and practical significance. A relationship which is statistically significant may give predictions so uncertain that they are useless in a practical situation. With enough data statistical procedures can detect very tiny correlations between variables. The influence of a predictor may be real but so small as to have no practical effect. The models presented here are particularly simple in form. Studies producing statistical models frequently retain all the terms which are *statistically* significant, at some level reflecting the analyst's aversion to risk. In the present work terms were retained in the models because of their *practical* significance for investment decision making. While formal statistical methods were used to guide the selection of terms in the models, statistically significant terms which, upon investigation, did not contribute anything obviously valuable to the decision making process were deleted.

6.2 The effects of a changing economic climate

The data used for Tipping's 2011 study related to the years immediately preceding the 2008 Credit Crunch. At the commencement of the Credit Crunch and the resultant economic downturn, the British retail banks ceased placing their operational premises into sale-and-leaseback packages. Also, during the depths of the economic downturn, fewer banking premises were traded by investors. The resultant sparseness of data is likely to make a multivariate further analysis at a meaningful level impossible. It is known that the economic cycle has an effect on yield. However, in the light of the changed economic environment, it remains to be seen if any interactions between the independent variables have changed or remain unaltered post-2008.

6.3 Applications to other classes of property investment

Potentially, the models developed in the banking-hall study may be capable of application to other property classes. However, caution needs to be exercised, because it is possible that effects, especially interactions between the independent variables, may be different for other types of property.

6.4 Further work/future studies

Portfolios of other clearly defined classes of property have recently been sold at auction on sale-and-leaseback and more are likely to come to the market. For example, it is suggested that the current trend for the sale-and-leaseback of betting shops is likely to continue (Allsop, 2014). Also, it was announced in 2014 that the Co-operative Group was considering a large-scale sale-and-leaseback programme of its operational estate as one means of plugging a gap in its finances (Dixon, 2014). However, in order to replicate the statistical methods used in the present study, it is necessary that future studies use datasets comprising sufficient cases to adequately populate a contingency table. Such prerequisite is likely to limit the scope for further work. Therefore, any such study should be focused on those types of property which can be identified as having had the greatest number of transactions across variables and categories over a period.

Since banking-hall investments had been shown to be sought after by private investors, it was expected that increased competition amongst this type of purchaser for more affordable investments would have resulted in lower yields for small lot sizes relative to larger lot sizes. Qualitative research suggests that lot size impacts upon the yields of retail bank premises, but without identifying which lot size would produce the highest yield. Notwithstanding that yield is a function of rent, effects attributable to lot size mean that there will be some yield shift along the range of rents. The literature is not clear whether yield reduces or increases as rent rises. Although the expectation is that competition between smaller investors will drive down the yields of banking-halls let at lower rents, the findings of Hutchinson *et al.* could be used to suggest the opposite. They found that in some sectors ‘a wall of money’ from institutional investors was driving down the yields of better properties during the later years of our study. The implication is that since such institutional investors might be expected to be more likely to buy premises let at higher rents, this would drive down the yields on those lots. However, our study found that banking-halls with higher rents generated higher yields. Although it is true that the normative view is that larger lot sizes are more secure and let to tenants with a stronger covenant, something else is clearly occurring with retail bank premises. Certainly, high street retail bank premises have proven very popular with private investors, who have seen banking-halls as a secure investment.

The supposition is that bank premises with lower rents have attracted a greater number of small, private investors due to their affordability. This has driven prices up and yields down for lower rent banking-halls in comparison with higher rent ones. Such an effect is contrary to the assumptions that may be drawn from Hutchinson *et al.* for other types of property. This is an area which merits further research investigating specific property types.

References:

Adams, A.T. and Clarke, R. (1996), “Stock market reaction to sale and leaseback announcements in the UK”, *Journal of Property Research*, 13, pp.31-46.

- Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, New York: John Wiley & Sons, Inc.
- Akaike, H. (1974), "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, 19(6), pp.716-721.
- Allsop. (2014), "Allsop Commercial Auctions: 1013 annual Review", London, <http://www.auction.co.uk/commercial/docs/AnnualReview2013PDF.pdf>. Accessed 20 Feb 2014.
- Ambrose, B.W. and Nourse, H.O. (1993), "Factors influencing capitalization rates", *Journal of Real Estate Research*, 8(2), pp.221-238, American Real Estate Society.
- Baum, A., Mackmin, D. And Nunnington, N. (2006), *The income approach to property valuation*, London: Reed Business Information (EG Books).
- Bryne, P., Jackson, P. and Lee, S. (2013) "Bias or rationality? The case of UK commercial real estate investment", *Journal of European Real Estate Research*, vol. 6, no 1, pp. 6-33, Emerald.
- Chen, J., Hudson-Wilson, S. and Nordby, H. (2004), "Real estate pricing: spreads & sensibilities: why real estate pricing is rational". *Journal of Real Estate Portfolio Management*, 10(1), pp. 1-22.
- Cheng, P. and Roulac, S.E. (2007), "Measuring the effectiveness of geographical diversification", *Journal of Real Estate Portfolio Management*, 13(1), pp. 29.
- Dixon, A. (2014), "Co-op mulls £100m portolio sale", *Estataes Gazette*, 1359, p.3, Business Source Premier, EBSCOhost.
- Dorling, D. (2011), "The North-South Divide - Where is the line", Available at: <http://www.sasi.group.shef.ac.uk/maps/nsdivide/index.html>. Accessed: 14 June, 2013.
- Dunse, N., Jones. C., White, M., Trevillion, E. and Wang, L. (2007), "Modelling Urban Commercial Property Yields: Exogenous and Endogenous Influences", *Journal of Property Research*, 24(4), pp.335-20.
- Evans, A.W. (1995), "The property market: ninety per cent efficient?" *Urban Studies*, 32(1), pp.5-30.
- Enever, N. and Isaac, D. (2002), *The valuation of property investments*, London: Estates Gazette Ltd.
- Fraser, W.D. (2004) *Cash-flow Appraisal for Property Investment*, Basingstoke: Palgrave.
- Guy, S. and Henneberry, J. (2000), "Understanding urban development processes: integrating the economic and the social in property research", *Urban Studies*, 37(13), pp.2399.
- Hoesli, M., Lizieri, C. and MacGregor, B. (1997), "The spatial dimensions of the investment performance of UK commercial property", *Urban Studies*, 34(9), pp.1475.
- Hutchinson, E., Adair, A. and Findlay, N. (2011), "The impact of covenant strength on property pricing", *Journal of Property Research*, vol. 28, no. 2, June 2011, pp. 167-188 Routledge.
- Investment Property Databank (2008), *IPD United Kingdom retail property index*, London: Investment Property Databank.
- Isaac, D. and Steley, T. (2000), *Property Valuation Techniques*, Basingstoke: Palgrave.
- Krystalogianni, A. and Tsolacos, S. (2004), "Regime switching in yield structures and real estate investment", *Journal of Property Research*, 21(4), p.279.
- Lee, S. and Byrne, P. (1998), "Diversification by sector, region or function?" *Journal of Property Valuation and Investment*, pp.38-19.
- Leone, V. (2011), "From Property Companies to Real Estate Investment Trusts: The Impact of Economic and Property Factors on UK Listed Property Returns", *Economic Issues*, vol. 16, part 1.
- Macgregor, B. and Schwann, G. (2003), "Common features in UK commercial real estate returns", *Journal of Property Research*, 20(1), pp.23-48.

- Mackmin, D. (2009), "Put the bases to the test", *Estates Gazette*, 26 September, pp.106-8, London: Reed Business Information.
- Markowitz, H. (1952), "Portfolio selection", *Journal of Finance*, pp.77-91.
- Office of National Statistics. Business Investment Time Series Identifiers DSG9 and DSH7.
- R Core Team (2013), "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Sheskin, D.J. (2003), *Handbook of Parametric and Nonparametric Statistical Procedures, Third Edition*, Chapman & Hall/CRC, Boca Raton.
- Tipping, M. and Bullard, R.K. (2007), "Sale-and-leaseback as a British real estate model", *Journal of Corporate Real Estate*, 9(4), pp.205-17.
- Tipping, M. (2011), "British Banking-halls as a Property Investment", Doctoral thesis, Anglia Ruskin University, UK.
- Tipping, M. and Lam, T. (2010), *Factors most influencing United Kingdom retail property yields: a theoretical perspective*, COBRA Conference, Paris, 2nd-3rd September, 2010, London: Royal Institution of Chartered Surveyors.
- Whitmore, J. (2009), "Prime yields plateau ahead of predicted year-end pick-up", *Property Week*, 8 May 2009, London: CMP Information.

Biographical notes:

Malvern Tipping is a Fellow of the Royal Institution of Chartered Surveyors with thirty-five years' experience. He holds a B.Sc. in Urban Land Administration from Portsmouth University and a Professional Doctorate (D.Prof) in the Built Environment from Anglia Ruskin University both in the UK. Currently, a director of property investment and development companies and a Visiting Research Fellow of Anglia Ruskin University, his research interests include valuation and sale-and-leaseback.

Roger Newton has over 40 years experience as a statistician working in Nationalised Industry, the Civil Service and the Aerospace Industry. He holds a B.A. in Mathematics and Physics from Keele University and a M.Sc. in Statistics from Imperial College, London. He is a Fellow of the Royal Statistical Society and a Chartered Statistician.