Multiple Regression Analysis – A Case Study

The first step in a case study analysis involves research into the subject property and a determination of the key factors that impact that property. Then, in an effort to determine any effect on value, case studies are developed from other properties that are similarly situated with respect to the subject property and its value drivers. In order for the analysis to be reliable, case studies must make "apples to apples" comparisons, that means that case studies being used must have similar property, market, and value drivers (highest and best use) as the subject property.

Typically, the case-study method is considered a special case of the Sales Comparisons approach to value. Statistical studies, including simple and multiple regression analysis, are considered a refined version of sales comparison analysis that generally uses more data and allows for the statistical testing of the results.

The purpose of this MRA case-study analysis is to demonstrate how this statistical technique can be used to isolate the impact of various value drivers and also to illustrate how MRA outcomes can be used to "test" the reasonable of the variables that form the basis of market value.

Following is a brief description of the physical and functional characteristics of the case-study subject property.

Current Use: The property is improved with a one-story, 8,967± SqFt, concrete tilt-up ("CTU") warehouse building and parking that was erected in 1974.
Construction Type: Class "C"– Concrete Tilt-up.
Year Built: 1974. The actual age is 40 years.
Effective Age:20years
TotalEconomic Life:45years.
Remaining Economic Life: 25 years.

1

Multiple Regression Analysis

Description of Improvements: This is a concrete tilt-up warehouse/retail building with office space and these general specifications:

Vertical Clearance:	20’ Min. / 24’ Max.
Office Space:	796± SqFt (22.39% of GBA)
Parking:	12 Spaces
Parking Ratio:	1.509/1,000 SqFt
Sprinklers:	The property is fully sprinklered.
Electrical Power	
Main:	2000 Amps, 480 VAC, 3 Phase, 3 Wire
Subpanel:	400 Amps, 120/240 VAC, 3 Phase, 3 Wire
Loading:	0 “Dock High” Spots 3 Ground Level Doors 0 Rail Spurs

Functionality: Interviews with local brokers and market participants suggested that a lack of dock high loading spots, common for distribution warehouses, might be a functional detriment to an industrial property. Manufacturing properties built in past decades, such as the subject, often have only ground level doors. While our analysis of market leases did identify a general market preference for dock high doors, it appeared that the subject is leased at market rates.

Multiple regression analysis (MRA) is a statistical method that correlates the behavior or variation of a number of factors, or independent variables, in order to ascertain their individual and combined impact upon a single factor, called the dependent variable.

“The multiple regression equation describes the average relationship between these variables, and this relationship is used to predict or control the dependent variable.2” In statistical terms, this is known as a hedonic pricing model.

The principal drawback to multiple regression analysis is that it is a very data- hungry technique. Large data sets must be available for the analysis to be reliable. Smaller data sets run the risk that a few observations can significantly affect the outcome of the regression model. This can lead to a lack of multivariate normality, which is one of the underlying assumptions of a multiple regression analysis. Some researchers indicate that “a good rule of thumb is to use at least 10, preferably more, comparable sales per independent variable included in the MRA [Multiple Regression Analysis] model. 3” Typically, appraisers do not have that much data in a given market necessary to reach a reliable conclusion. In the case of the subject case-study property, it is located in a very active submarket, historically speaking. Vacancy rates have traditionally been very low and the market has remained relatively active over the years. Thus, we were able to compile sufficient data to overcome this hurdle.

The other drawback is that multiple regression analysis is statistical in nature, meaning that it relies on hard data characteristics (i.e. building size, year built, clear height, etc.) and thus, cannot truly take into account all of the “soft” characteristics of a property (i.e. individual location, effective age, etc.). Therefore, sufficient data must be not only known about each comparable, but must be mathematically quantifiable. In the case-study data set, we identified seven critical variables in our analyses, all of which are mathematically quantifiable. These variables included:

Rented SqFt (X1)
Percent Office Space (X2)
Year Built (X3)
Clear Height (X4)
Truck-High Doors (X5)
Ground Level Doors (X6)
Date of Lease (X7)

After defining the variables and the data set, we ran the regression through the *Data Analysis ToolPack* in Microsoft Excel. According to the program, “The Regression analysis tool performs linear regression analysis by using the ‘least squares’ method to fit a line through a set of observations.” The “least squares” method is one of the more common ways of determining the regression equation “by solving a system of simultaneous linear equations in which the unknowns are the constants of the regression equation.4”

A first regression was run with the data from 167 leases. Then, 20 were eliminated, because their residuals (the difference between predicted and actual values) exceeded one standard error. A second analysis was run with the remaining 147 leases. The results are below.

Regression Statistics	
Multiple R	0.963238794
R Square	0.927828975
Adjusted R Square	0.924194463
Standard Error	845.2831017
Observations	147

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	1276803988	182400569.7	255.282954	4.85766E-76
Residual	139	99315989.57	714503.5221		
Total	146	1376119978			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-25199.92159	9041.39028	-2.787173301	0.006061244	-43076.35789	-7323.485294
Rented SqFt	0.299058402	0.010727598	27.87747948	8.86251E-59	0.277848032	0.320268771
% Office	1922.329272	859.5535812	2.236427506	0.026916907	222.8389989	3621.819545
Year Blt Clr	6.205948575	4.584375236	1.353717411	0.178023625	-2.858176447	15.2700736
Ht TD GL	46.36805005	25.80616359	1.796781993	0.074541676	-4.655324137	97.39142424
Lease Date	280.7750856	76.40493815	3.674829041	0.000338913	129.7089368	431.8412344
	-140.4947672	72.23526439	-1.944960933	0.053798454	-283.3167296	2.327195116
	0.354482196	0.090385688	3.921884132	0.000137489	0.175773619	0.533190774

The resulting regression equation is:

Monthly Rent =

- 25199.92159
+ 0.299058402xSqFtinPremises
+ 1922.329272xPercentOffice
+ 6.205948575xYearBuilt
+ 46.36805005xFeetofClearHeight
+ 280.7750856xNumberofTruck-HighDoors
- 140.4947672xNumberofGroundLevelDoors+

0.354482196 x Date of Lease

Using the characteristics of the case-study property in the equation yields a rent indication of $11,432 per month, or $0.367/SqFt/Mo., unadjusted for any upgrades.

Reasonableness:

Most of the other statistics in the above regression summary are tests indicating whether the model and the variables are statistically significant or not. The most important is the "Adjusted R-Square." The "R-Square" statistic tells us how much of the change in the dependent variable (in our case, total rent) is accounted for by the changes in the independent variables (the property characteristics). The R-Square must be between zero and one, with a higher number indicating more explanatory power by the independent variables.

The Adjusted R-Square simply accounts for the degrees

of freedom (the number of observations less the number of variables) in the model. In this case, the two numbers are very close to one (R-Square of 0.9278 and Adjusted R-Square of 0.9242). Together, these statistics indicate that over 92% of the change in total rent can be explained by changes in the seven chosen variables. In addition, the model is not over-specified with useless variables that might skew the results. The "Standard Error" statistic gives

us the average error
in the predicated dependent variable. This essentially is a measure of the accuracy of the estimate provided by the regression model. If the standard error is large, relative to the prediction, then estimates of the dependent variable based on this model will not be very reliable. One standard error is equivalent to about a 66% confidence level. Therefore, we could predict that the total rent for the subject case-study property has a 66% chance of being $11,432 ± $845. For our purposes, we want a better significance level. A 90% confidence interval can be obtained by multiplying the standard error by 1.65. Therefore, we have a 90% chance, based on the original data set, that the subject case-study property will rent for somewhere between:

$11,432–(845X1.65)= $10,037/Mo.

And

$11,432+(845X1.65)= $12,826/Mo.

The "F" Statistic tells us whether or not we can reject the assumption (the "null hypothesis") that the independent variables (jointly) are insignificant. In other words, how significant is the regression analysis as a whole? Our F Statistic is 255.282954. At a 95% confidence level, with

147 degrees of freedom, the critical F statistic is 2.164. The fact that the F Statistic is so much larger than the critical number, a byproduct of a large and consistent data set, underscores the significance and reliability of this model. Our model's F probability is 4.857 X 10-76, practically zero. Therefore, we can conclude that the coefficients indicated by the regression model did not occur purely through chance.

We also calculated the Durbin -Watson statistic for this regression model. This is a test used to find the presence of autocorrelation in the residuals, which, if present, could invalidate the standard error measurements. The test returns a value between zero and four. A value of exactly two indicates no autocorrelation at all. A value of zero indicates very high positive autocorrelation, while a value of four indicates very high negative autocorrelation. The statistic of the regression model is 2.375, well within the bounds of acceptability.

The other relevant statistics in the table are the individual "t-stats" for each independent variable. These are simply the ratios of the coefficients to their corresponding standard errors. A higher t-stat indicates that the variable is more statistically significant to the model. A low t- stat indicates the opposite. A general rule of thumb is that a t-stat above two signifies a significant variable. In the case of our model, most of the variables were very significant. The t-stat for the "Rented SqFt" variable was 27.8775, indicating an extremely high correlation between size and lease rate. This relates to our common sense understanding that larger spaces rent for more money. The date the comparable was leased also has a relatively high t-stat of 3.9219. Again, this makes intuitive sense, as the date the lease was signed plays a large role in the rental rate.

It appears that the least relevant variable was the year that the building was built. This can be explained, because, while the year built is intended to be a proxy for condition (i.e. newer buildings are typically in better condition than older buildings), it doesn't take renovations and quality of upkeep into account. An older building that has been very well-maintained can certainly be in just as good, and perhaps better, condition than a newer building with lower quality construction and less maintenance.

The most interesting relationship that can be taken from the t-stats is the relationship between truck-high and ground level doors. With a t-stat of 3.6748, truck high doors must play a key role in the determination of a rental rate. Again, this makes perfect sense, as a building with more truck-high doors, all else being equal, should rent for more than a building without. However; the ground level doors variable has a *negative* coefficient, indicating that for every ground level door added to a building, the rental rate actually goes *down.* At first, this does not make sense. Why would adding something that theoretically adds functionality to a building decrease the amount that the market will pay for it?

The answer lies in the typical user in the market. The outsourcing boom of the 90s and the import boom coinciding with the rise of China and India have resulted in the typical user in the market shifting from manufacturing to warehousing. Conversations with market participants revealed that the typical user now is importing and exporting goods to and from the Ports of Los Angeles and Long Beach. They need warehouse space to store their goods before distribution. Since many of those goods are moved via trucking, truck-high doors are very useful for the typical warehouse user. Therefore, the market is not necessarily discriminating *against* ground level doors. They are, in effect, cheering *for more* truck high doors.

The fact remains that the subject case-study lacks any truck high doors. It has three ground level doors, one facing a primary arterial and two in the rear yard facing a secondary feeder street.

Nevertheless, the subject case-study suggests that the subject case-study property appears to be renting at a market rate, as indicated by the regression analysis. The current tenant makes good use of the subject case-study facilities as they currently exist and do not particularly need any truck-high doors.