Dear Reader:

Thank you for your interest in our Whitepaper, "A Statistical Study of Traffic Capture Rates for Petroleum Marketing Properties". Enclosed is your complimentary copy. We hope you will find it helpful and informative. This paper describes a model to predict traffic capture rates for convenience stores and gas stations.

For additional information about current valuation issues in the retail convenience channel, you may wish to visit our website at www.cstorevalue.com. On the "Video Insights" page you will find short, appraisal-related video clips on these important topics:

1. Development Process
2. Supply and Demand
3. Hypermarket Competition
4. What Should be in an Appraisal of a C-Store: A Lender's Perspective
5. A Business Appraiser's Perspective
6. Retail Site Analysis
7. Below-Cost Selling

At our website you will also find useful Whitepapers and current valuation metrics.

If we can be of any assistance to you, it would my pleasure to personally serve you. Please let us know and thank you again.

All the Best.

Robert E. Bainbridge MAI, SRA. MRICS
C-Store Valuations

# WHITEPAPER NO. 2

**A STATSITICAL STUDY OF TRAFFICE CAPTURE RATES**
**FOR PRETROLEUM MARKETING PROPERTIES**

By
Robert E. Bainbridge

# Investigating Fuel Sales in the Convenience Industry

### *Introduction*

With the current business model, the convenience industry depends upon the retail sale of motor fuel. Today, motor fuel sales account for about 50 percent of gross profit for a typical convenience store.[1] With such a strong dependence upon fuel sales and with the recent entry of competing retail channels into the fuel sales business, such as WalMart and Albertsons, this paper will investigate why consumers purchase motor fuel at specific locations.

Proprietary computer modeling has been developed by a number of industry firms that attempt to predict the fuel sales volume at a specific location. The predictor variables and models are closely guarded secrets and this information is not usually divulged to the public or competitors. Being able to accurately predict the volume of retail fuel sales for a proposed convenience store site is a very valuable skill because of the strong economic dependence convenience stores have on petroleum sales. A proposed store that fails to achieve projected fuel sales volume will not be able to survive under the current industry store design concepts.

The ideal predictive model is one that would predict fuel sales and do so accurately. Ideally, the model would use objective, easy to obtain information that was not subjective, biased or subject to interpretation. The model would be accurate and applicable under a variety of geographic and operating scenarios.

The minimum threshold level of accuracy desired varies from one user to another. For my purposes, I hope the model achieves at least 65% of explaining the variation in fuel sales among these retail outlets. To me this implies some statistical significance. At this level, it could be argued that more than half the variation is accounted for by the selected predictor variables.

In reality, the fuel sale volumes for individual retail outlets are not available from any public source. If this information was available from public sources, we could include hundreds, perhaps thousands of data relationships into our model. The lack of readily available fuel sales data makes it very challenging to develop a model because the response variable ( fuel volume) is difficult to obtain. In our case, this limits our investigation to about 35 retail

[1] *State of the Industry*, published by the National Association of Convenience Stores. 2001.

outlets where we have fuel sales information. Ideally, I wanted to have more data; 100 or more would have been preferable. This number of data is not possible at the present time. However, because of the practical usefulness of the study, and because this study is needed in my work, I will proceed with the data that is available. The selection of these 35 data does necessarily represent a random sample. The data is selected from our office appraisal files. But, beyond what is mentioned below, the data is believed to be representative of convenience stores using the same business model as the data here. The business model examined here is a 1990s convenience store concept with fuel sales and in-store merchandise. Fueling positions usually number from two to ten, and the store building usually varies from 1,000 to 4,000 square feet. Typically, these are national oil company retail fuel brands. The properties are usually in high traffic locations. Because of the business model profile of these data, the results of any statistical model developed in this study would not necessarily be applicable to other retail formats, such as service stations, card lock facilities or today's mass retailers, such as WalMart or Albertsons.

Measuring and recording the response variable can be problematic because of the extreme volatility in retail fuel prices over any given period. For example, in one twelve-month period, the retail price of regular unleaded gasoline varied by almost $1.00 per gallon in the Seattle area. To eliminate the volatility of retail fuel prices, the number of gallons dispensed has been relied upon as the measure of fuel sales. In the convenience industry, this is commonly referred to as "gallonage".

To limit our model to the most objective criteria possible, we will examine the relationship between fuel sales and:

*Predictor Variables*
•Traffic Volume (TRAFFIC)
•Fueling Positions (FP)
•Competitive Fueling Positions (CFP)
•Trade Area Population (POP)
•Canopy Design, Starting Gate (CANS)
•Canopy Design, 4-Square (CAN4)

Traffic volume is defined as the average daily traffic passing by the store as measured by the State Transportation Department. Our intuition would suggest that higher traffic volume sites will sell more gasoline than lower traffic volume sites.

Fueling positions refers to the store's number of operating fueling positions. A fueling position is the space allocated to a single customer vehicle. For example, a two-sided dispenser that can be accessed from two sides creates two fueling positions. The convenience industry believes that the larger the number of fueling positions, the better the store can accommodate peak volume periods, and thereby sell more gasoline.

Competitive fueling positions is the number of competing retail fuel positions within the subject's trade area. Trade area can not be dogmatically defined. The trade area includes all site s that compete directly with the subject. Presumably, the greater the number of competing sites, the less fuel sales volume the subject will be able to achieve, all other factors being equal.

Trade Area population is defined as the number of residents within a two-mile radius, the standard definition of the trading area, or competitive market of a convenience store. Presumably, higher density residential areas create a greater demand for motor fuel.

Conventional wisdom in the retail fuel industry believes that starting gate canopy designs are more efficient for moving customers through forecourt than four-square designs. Starting gate designs place all of the fueling positions parallel to each other. Four-square designs place two or more fueling positions within the same line of travel. This can cause a forward position to be blocked, or inaccessible, when the rear position is being used by a customer. Starting gate canopy designs are viewed as superior by the industry.

For this nominal scale data of canopy type, dummy variables are entered for CANS or CAN4. CANS designates a starting gate design, while CAN4 designates a four-square design.

***Methodology***

The predictor variables for each of the 35 retail fuel outlets have been recorded on data sheets that appear in the addenda. For identification purposes, each of the 35 outlets is sequentially numbered, the oil company brand name has been recorded along with the store's location as to the city and state.

<u>Factors Not Considered</u>

This model attempts to explain variation fuel sales volume, not to explain every reason as to why customers buy gasoline. I have not considered factors that are not distinguishable. In other words, if all the sample data is believed to share a particular attribute, no reason exists to include that attribute in the model. Keeping the predictor variables as low as possible reduces the likelihood of multi-collinearity and keeps the degrees of freedom as high as possible. Hopefully, these objectives produce a more reliable statistical model.

Below are the factors that are not included because of the homogeneity of the data with respect to these characteristics.

*Pricing*
None of the 35 retail outlets in this study are discount retailers. The retail pricing posture of these sample outlets are very similar and competitive. Although industry surveys consistently indicate that the retail price is the number one reason as to why customers buy gasoline at a particular location, these data are not differentiated by price.

*Proximity to Mass Merchandisers*
Beginning in the year 2000, WalMart entered the retail fuel sales business. After an initial trial period, they were convinced that they could profitably enter the gasoline market on a national scale. During the next couple of years, other mass retailers, such as Costco, Fred Meyer and Albertsons responded by entering the retail gasoline business also. Often, the pricing strategy of the mass retailer has been aggressive and even predatory in nature in an effort to gain market share. And, mass retailers have gained market share, accounting for about 15 percent of all retail motor fuel sales. This percentage is expected to grow. Today, convenience stores that are in the unfortunate position of being located close to a discount mass retailer are experiencing a dramatic drop in fuel volume.

However, none of our data samples are effected by mass retailers, because all the data in this study predates the year 2000. Prior to that time, no mass retailers had entered the retail fuel business.

*Branded Gasoline*
The brand of gasoline at one time played a large role in why the customer shopped at a particular location. However, the importance of a brand name has diminished greatly over the last 20 years. With few exceptions, brand and product loyalty does not exist in today’s retail landscape. Price has been inculcated into consumers as the criteria of good consumerism. In any event, all of these samples are recognized

national oil company brands. So, little differentiation exists among these data samples for this characteristic.

The most direct relationship between branding and fuel sales is probably the number of customers within a region that hold a credit card for that brand. Many dispensers today accept a wide variety of credit cards. So, even this factor is diminishing in importance.

*Merchandise Sales*
For a business that sells both, in-store sales may or may have a significant influence on gasoline sales. Here, all of the retail outlets offer merchandise. In this model, merchandise sales are not believed to be playing a significant role in explaining fuel sales volume.

If there is a relationship between merchandise and gasoline sales, it is a general observation that it is the fuel customer that makes the impulse decision to buy merchandise, rather than the merchandise customer that makes the impulse decision to buy gasoline. Fuel is the product leader, not merchandise.

The factors that are considered important in accounting for variance in fuel volume are used in this model. The predictor variables for each store have also been recorded on these data sheets.

Mini-Tab, a statistical analysis program, has been utilized to measure the relationship between the predictor variables and the response variable, which in this case is fuel sales volume. Some factors may prove important, while others are not.

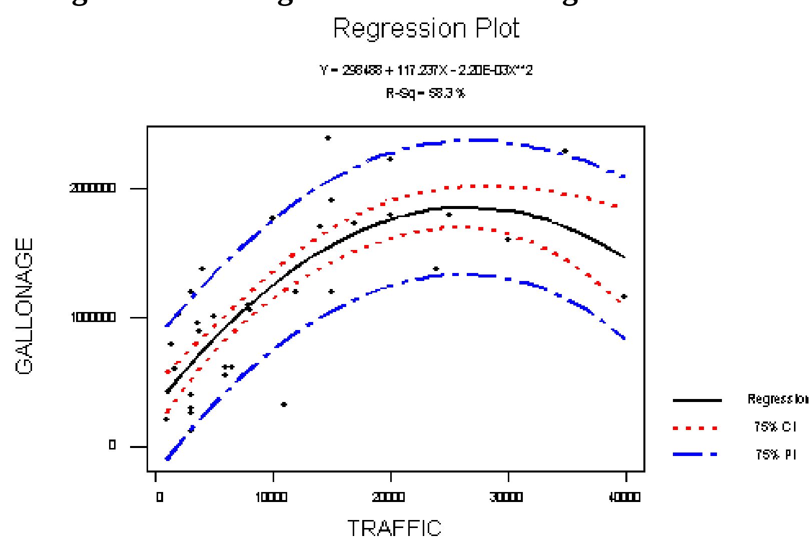
A step-wise regression will measure the contributory importance of each of the predictor variables. This along with some re-trials of the model and common sense should allow us to see how much variation can be accounted for in fuel sales with the independent variables used in this model. The measure will be the coefficient of determination, or r-squared value..

***Results of the Analysis***
Mini Tab was used to generate a statistical summary of the data. The descriptive statistical summary includes the number of observations, mean values, median values, standard

deviations, standard error of the means, minimum and maximum values, and first and third quartile values. The summary was generated before making any manipulations to the data, such as eliminating outliners. At this point, we are simply letting the raw data speak for itself. This data summary appears in Section 1 of the addenda. Steps 1 through 5 in the analysis below are an attempt to isolate and eliminate any aberrations in the data. After these steps are completed, the development of the model is then pursued.

<u>First Analysis: Scatter Diagram and Line Fit</u>
Next, a quadratic line graph was generated to get a rough idea of the pattern of the data. Here, traffic volume was plotted on the x-axis and gallonage was plotted on the y-axis. The line fit on this regression produced a r-squared value of 58.3%.
<u>Second Analysis: First Step-Wise Regression</u>
A step-wise regression using all data and all predictor variables was then completed. The step-wise regression produced an r-squared value of 53.25%. This result is somewhat

**Figure 1 First Regression: Scatter Diagram and Line Fit**




surprising in that the r-squared value feel slightly using all of the predictor variables as opposed to using traffic volume alone. My initial impression from this result is that the predictor variables other than traffic volume are not helping to account for the variance in fuel sales to any great extent.

**Figure 2: Second Analysis: First Step-Wise Regression**

**Stepwise Regression**

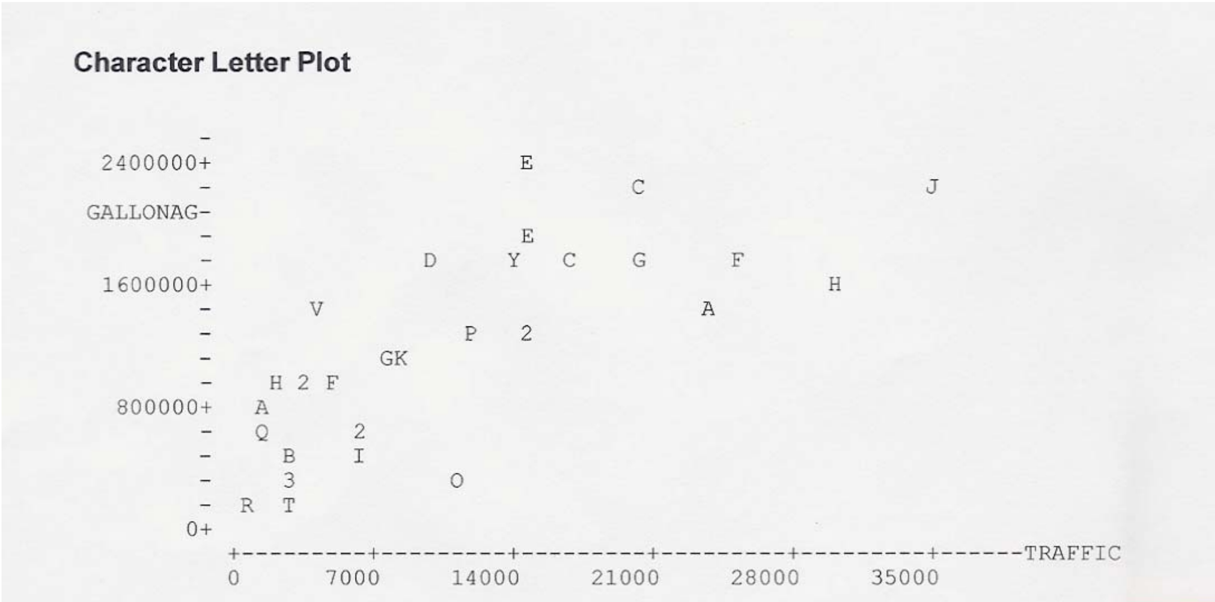
```
 F-to-Enter:      6.00    F-to-Remove:      0.00

 Response is GALLONAG on  6 predictors, with N =   35

    Step         1        2
Constant    657563   188839

TRAFFIC       41.2     30.3
T-Value       4.94     3.51

FP                    74457
T-Value                2.71

S           489290   448153
R-Sq         42.53    53.25
```

Third Analysis: Character Letter Plot

To identify the data that could be aberrations, a character graph was generated. Data points No. 2 and No. 15 were categorized as outliners based upon a visual inspection of the graph.

The logic of eliminating data No. 15 is that, although a quadratic function may best fit the data, we do not expect the line of best fit to rise and then turn downward as traffic volume increases. At least this is counter intuitive. The line may very well flatten out, but there is little reason for it to decline with this set of data. Indeed, most of the data do not suggest a decline. But, data No. 15 is unduly influencing the line plot downward for a the high traffic end of the graph line.

**Figure 3: Character Plot**




Fourth Analysis: Another Step-Wise Regression

Another step-wise regression was completed after eliminating the outliners. This indeed improved the coefficient of determination to 58.7% from the previous level of 53.25%.

**Figure 4: Fourth Analysis: Step-Wise Regression**

**Stepwise Regression**

Using all predictors

```
 F-to-Enter:      6.00    F-to-Remove:      0.00

 Response is GALLONAG on  6 predictors, with N =   33

    Step         1
Constant    578047

TRAFFIC       54.6
T-Value       6.64

S           417779
R-Sq         58.70
```

Fifth Analysis: Another Scatter Diagram and Line Fit

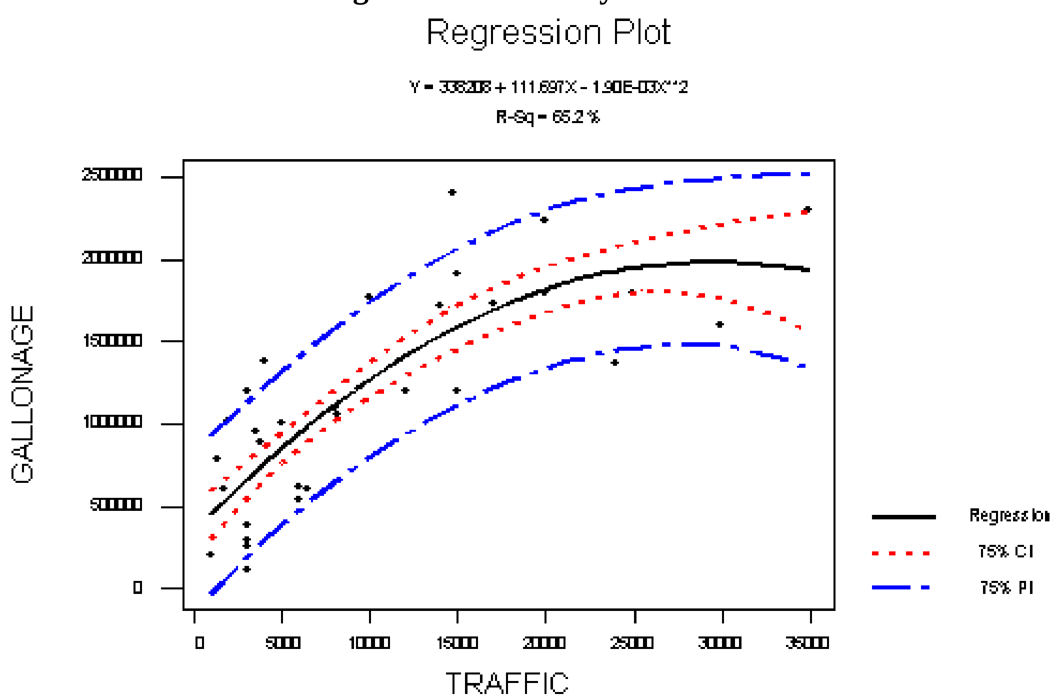
Tantalized by the higher r-squared value found in the quadratic line fit above, another line fit was made after eliminating the outliners. The r-squared value improved to 65.2%.

The step-wise regression in the fourth analysis found that only traffic volume was adding anything of much importance to the model. Even the predictor variable of the number of fueling positions was dropped after eliminating the outliners.

The higher r-squared value with the quadratic line fit shows that a curved line fits the data better than a straight-line function.

<u>Sixth Analysis: Best Subsets Regression</u>

**Figure 5: Fifth Analysis: Line Fit**



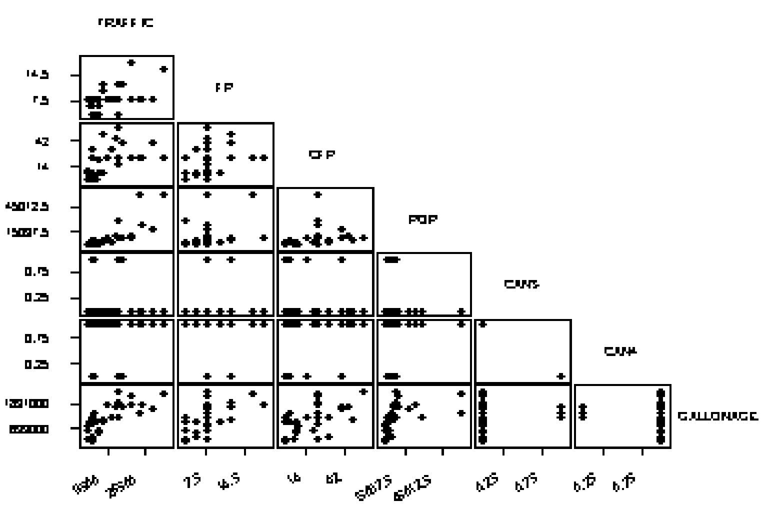

Once the outliners were eliminated, we looked for the best model using a Best Subsets routine.

A matrix plot was performed which contains a scatter plot for each pair of variables. The bottom row is the most important. The bottom row shows "Gallonage", our response variable plotted against each predictor variable. Note the two canopy design plots. They are nearly vertical. Mini Tab eliminated these variables because of this relationship.

The Best Subsets Regression generates regression models using the maximum r-squared value by first producing one predictor variable models and then adding each predictor variable in turn until all have been examined. This way we can select the best predictor model using the smallest number of predictor variables. Using the smallest number of predictor variables is desirable because multi-collinearity and extraneous data will be minimized.

The fifth row in the Best Subsets Regression produced the highest r-squared (67.0%) and the lowest standard error (386,086) using three predictor variables, traffic, fueling positions, and competitive fueling positions. Adding the predictor variable of trade area population increased r-squared to only 67.9%, a negligible difference.

**Figure 6: Sixth Analysis: Matrix Plot**




Using these three predictor variables (traffic, fueling positions and competitive fueling positions) another stepwise regression model was made. The r-squared value was 64.20%

**Figure 7: Sixth Analysis: Best Subsets Regression**

**Best Subsets Regression**

```
Response is GALLONAG

                                      T
                                      R
                                      A
                                      F
                                      F   C P
              Adj.                    I F F O
Vars   R-Sq   R-Sq    C-p          s  C P P P

   1   58.7   57.4    7.0    417779   X
   1   37.1   35.1   25.8    515553     X
   2   64.2   61.8    4.2    395373   X X
   2   62.4   59.9    5.8    405317   X   X
   3   67.0   63.6    3.8    386086   X X X
   3   65.9   62.4    4.7    392325   X X   X
   4   67.9   63.3    5.0    387740   X X X X
```

**Stepwise Regression**

```
 F-to-Enter:      3.00    F-to-Remove:      2.00

 Response is GALLONAG on  3 predictors, with N =   33

    Step         1        2
Constant    578047   254667

TRAFFIC       54.6     43.9
T-Value       6.64     4.77

FP                    54590
T-Value                2.15

S           417779   395373
R-Sq         58.70    64.20
```

**Stepwise Regression**

### *Conclusion and Recommendations*

We failed to find the “holy grail” of site selection statistical models. Indeed, perhaps none exists. After reviewing the data and analyses in this attempt to explain differences fuel sales volume, I am skeptical that any statistical model can achieve a great deal of accuracy in this area. Customer buying decisions simply have too much random variableness for analysts to ever create an infallible mathematical model of why they buy gasoline at a particular location. The reader can see that the prediction intervals on the line fit in Figure 5 are too wide to be of much practical use. For example, at a 75% level of confidence, the prediction interval at 10,000 vehicles per day is between 700,000 gallons and 1,700,000 gallons.

Physical, measurable attributes may explain up to 50% to 60% of how much fuel volume a particular site will, or should achieve. But, upwards of the half of all the reasons as to why a site will perform at a certain level of fuel sales is not explained. One additional reason that makes me confident of this statement is that only 29 data were collected originally. The statistical analyses and tests as described above were initially run on only these 29 data. Later, six additional outlets were added to the original 29, bringing the total sample to 35. The r-squared value did not change, it remained the same. These additional data made the original regression analysis no better or no worse than using 29. This could imply that even if we added a thousand data point, the explanation of variation would still be no better than 60%, or so.

For those proprietary companies that sell site analysis software and advice to the convenience industry, I am doubtful that they have any product that is doing better than a 60% to 70% level of confidence. The customer is probably never privy to those details of the model.

This means that retail locations will continue to be sorted out on a trial and error basis. Stores will fail at poor locations, and better locations will perform well. The knowledge of which sites are poor and which are good will only come after the fact when the final arbiter, the random voice of consumer choice, has been heard.

*Recommendations for Further Study*

•A larger sample set would add confidence to our conclusions here. More data should be collected and analyzed.

Industry surveys have consistently shown that price is the number reason that customers give as to where they purchase gasoline. So, incorporating the retail price of gasoline should be a promising predictor variable. However, the retail prices at a particular outlet fluctuate on a daily basis. If it were possible to record an average price per year for each retail outlet, then perhaps this type of information could be utilized and more accuracy could be obtained.

•Stratifying the data may attempted to see if better results could be obtained. For example, perhaps a common dynamic exists for retail outlets selling less than 500,000 gallons of fuel per year. That group of data could be analyzed separately. Or, perhaps traffic locations above 15,000 vehicles per day could be grouped together to see if the r-squared values improve.

•Perhaps some additional predictor variables cold be analyzed, such as a measure for access and visibility. However, based on what I have seen here, I am doubtful that the coefficients of determination would improve substantially.

# Addenda

## Section 1

### DESCRIPTIVE STATISTICS
### FOR EACH PREDICTOR VARIABLE

**Statistical Analysis for Real Estate Appraisal**

Class Project 2003

Dan Brick, PhD
Mail #OSS402
University of St. Thomas
2115 Summit Avenue
St. Paul, MN 55105

Jeffery Johnson, MAI
Integra realty Resources-
Minneapolis/St.Paul
222 N. Ninth Street
Suite 320
Minneapolis, MN 55402

# Investigating Fuel Sales in the Convenience Industry

by
Robert E. Bainbridge MAI, SRA
151 S.W. 1st Street
Ontario, OR 97914
(541) 823-0029 VOICE
(541) 823-0079 FAX
bbmw@fmtc.com