

An Analytical Approach To Detecting Insurance Fraud Using Logistic Regression

J. Holton Wilson
Central Michigan University

Abstract

Insurance fraud is a significant and costly problem for both policyholders and insurance companies in all sectors of the insurance industry. In this paper our focus is on auto insurance fraud, which occurs in both auto physical damage (APD-collision and comprehensive) and injury claims (Personal Injury Protection-PIP). We look at various situations within APD and PIP claims and various tactics that insured people use to defraud insurance companies. We then apply logistic regression as a statistical tool to help identify fraudulent claims.

Insurance companies typically employ a claims investigation unit to investigate fraudulent activities. The investigation unit gathers supporting information to deny claims that are fraudulent, or to authorize payment to claims where there is insufficient supporting evidence to draw a fraud conclusion. By some industry estimates between 10% and 15% of all dollars spent on insurance premiums are spent supporting those that commit fraud. (Snider, 1996) Identifying and denying fraudulent claims may lead to increased corporate profitability and keep insurance premiums at a level below where they would be otherwise for insured's.

Keywords: Insurance, Fraud, Prediction, Logistic, Regression

Auto Insurance Fraud: Introduction

The purpose of auto insurance is to indemnify an insured who sustained a loss, or to restore an insured to the same financial position he/she had prior to the loss. People who engage in insurance fraud attempt to receive benefits from an insurance policy that will over indemnify them. Automobile insurance fraud is not just a problem in the U.S. but rather it is a global problem as indicated in by an analysis of auto insurance fraud in Spain. (Artis, 2002) While our focus is on auto insurance, fraud is prevalent with other forms of insurance as well. For example, it has been estimated that health care insurance fraud costs Americans close to \$150 billion annually. (Krammer, 2003)

There are a variety of ways in which auto insurance fraud may be accomplished. Generally there are two situations that might cause an insured to commit insurance fraud. The first is a condition in which a person intentionally attempts to cause a loss or exaggerate a loss that has occurred. For example, a person may have sustained a legitimate loss but after thinking about the premiums they have paid over the years, they become opportunistic and attempt to include other prior damage in the same loss, possibly avoiding an additional deductible (e.g.: claiming a parking lot door ding along with a collision loss because it is all in the same vicinity on the vehicle). In such cases, the insured has suffered a legitimate loss but is now opportunistic and would like to be in a better financial position by covering everything in one claim and possibly saving the cost of additional deductibles.

The second type of situation that may result in a fraudulent claims is one in which an insured is less than careful, or even reckless, knowing that they have insurance coverage. It is not intended by the insured to create or exaggerate a loss. However, they engage in actions that they would not normally engage in if the possible adverse outcome were their personal burden. An example is when people take a vehicle off-roading and encounter situations that are potentially damaging. Chances are if they were not insured they would not take the chance of damaging the vehicle due to a potentially large financial burden on their part. However, since they are insured and damages would be covered by an insurance company, they take the risk of damaging the vehicle.

Some insurance fraud is directly intentional. The insured may stage a loss in order to receive benefits provided under an insurance policy. In the case of auto insurance it has been found that ten percent of reported thefts were not actually thefts. ("Fraudulent Auto Theft ...", 2002)

Other insurance fraud, however, is less deliberate. The defrauders may rationalize that there is nothing wrong with their actions. An example is a parent who's child just received a driver's license and, knowing that the family's insurance rates will skyrocket, allows their child to operate a vehicle without notifying the insurance company that the 16 year old driver has not yet been rated. This form of insurance fraud is known as misrepresentation and it induces an insurance company to make an underwriting decision that would not normally be made.

Results from a survey conducted by Accenture that was released in 2003 show that many people have attitudes that condone or are not completely negative about insurance fraud. ("News" 2003) About 24 percent say that it is quite or somewhat acceptable to overstate the value of a claim. Eleven percent say it is quite or somewhat acceptable to submit claims for items

not actually lost or treatments not actually received. Thirty percent agree or strongly agree that people are more likely to submit fraudulent claims during periods of economic downturn and 49 percent believe that they can get away with such fraud. These findings suggest that developing better ways to detect insurance fraud is an important goal.

Insurance Fraud: Costs

Insurance fraud is costly to individuals and the insuring companies. The following are examples of the cost of insurance fraud to individuals:

- The average household pays an higher auto and home owner insurance premiums to cover the cost of fraud
- The price of consumer goods rise as businesses are paying higher premiums due to increased insurance cost due to theft claims
- Cost of health insurance rises due to fraudulent injury claims, particularly in states that have unlimited medical coverage
- Innocent insured's are scrutinized more carefully and may incur longer periods to settle claims while under investigation

Even though insurance companies typically pass the costs of insurance fraud on to the consumer in order to operate at a profit, insurance companies are directly impacted by insurance fraud. The following are examples of the costs of fraud to insurance companies:

- Every dollar that is spent on insurance fraud directly impacts the profitability for the company as claim costs rise.
- Insurance companies incur increased human resource costs by employing fraud units to investigate claims.
- Insurance companies that do not effectively prevent fraud may lose business when their rates increase due to fraud.

Insurance companies also lose investment income when a fraudulent claim is filed. Insurance companies, in essence, have two bank accounts; one which is interest bearing, the other which is not. When a claim is filed, an insurance company must transfer funds from the interest bearing account to the non-interest bearing account in the form of reserves to satisfy the potential claim. These funds are held in the non-interest bearing account so there is no risk to the insured that the claim will not be satisfied. A fraudulent claim ties up these reserved non-interest bearing funds while the claim is being investigated (which can be a lengthy period) and eventually denied or paid. The following are representative of the dollar amounts that must be moved from interest bearing to non-interest bearing accounts each time a claim is filed:

- Collision \$2,750
- Total Auto Theft \$12,000
- Partial Auto Theft (wheels, stereos, etc) \$1,550
- Total Theft Recovered \$6,500

The investment opportunity lost can be damaging to insurance companies since it can be a large part of operating revenues.

Common Types of Fraud

Auto Physical Damage (APD) Fraud

One of the largest types of insurance fraud committed is misrepresentation. An insured gives information to the insurance company that induces the company to make an underwriting decision that it would not otherwise make. Misrepresentation is very common, in part because insurance companies do not normally prosecute for this type of fraud. They normally re-rate the policy and if it realized while a claim is pending, will charge backdated premiums if the insured wants to have the claim satisfied, or rescind the policy and return all paid premiums if the misrepresentation is 'material' (over \$500).

Common types of misrepresentation would include:

- Un-rated driver in household. An insured may have a high-risk driver in the household and intentionally withhold this information.
Uninsurable interest. This occurs when a person insures a vehicle that does not have any relationship to them (does not belong to them or anyone in their household).
- Theft Misrepresentation. In the event of a total theft of an insured's vehicle, the insurance company must offer fair market value based on the insured's best assessment of the condition and mileage of the vehicle (since it can not be verified, unless service records are available). An insured may not be completely honest with the condition of the vehicle, knowing it will affect the settlement amount, and since it can't be verified they won't get caught, unless the vehicle is recovered. Consider the following antidotal incident: *An insured person's vehicle was stolen and subsequently recovered 5 days later with damage sufficient to deem it a total loss. The insured wanted the value to be based on 15,000 fewer miles than the reading on the odometer, stating that this was the amount of miles that the thief put on the vehicle over a 5-day period. When it was explained that the thief would have to travel non-stop, without refueling, without restroom breaks, at a rate of 125 mph for a five-day period, the insured revoked his statement.* If a vehicle is stolen and subsequently recovered, an insured may claim that the vehicle was in perfect condition before the loss. Even though there may be separate incidences that damaged the vehicle, it is very difficult or impossible to prove what occurred while it was stolen and what happened over the normal course of the vehicles life.
- Chop Shops and Theft Rings. Theft rings are often very advanced and sophisticated operations. In many cases the insured is involved, is responsible for the vehicle being 'stolen', and files a theft report for the stolen vehicle. The insured is wholly indemnified for the loss, while the organization (to which the insured person may belong) disassembles the vehicle and sells it for parts.
- Damage to Own Vehicle. Claims may be filed when an insured encounters costly maintenance that is required on a vehicle. If they cannot afford it, they may purposely

damage the vehicle to create a total loss situation in order to obtain compensation from the insurance company. Some common methods are to burn the vehicle or to 'accidentally' let it roll into the water at a boat launch.

- Claiming Unrelated Damages. An insured may have a legitimate claim of loss, and may also have damage in the vicinity of this loss from a prior occurrence and claim that the earlier damage was related to the legitimate loss.

Personal Injury Protection (PIP) Fraud

Injury protection with respect to auto insurance claims can be a particularly lucrative proposition for insurance fraud. Injury protection may result in three types of payments: 1. A payment (settlement) made when a person suffers a lifelong injury or impairment of a bodily function; 2. Compensation to an insured party for wages lost due to injury; or, 3. Hospitalization due to injuries sustained in an accident and the ongoing treatment of this sustained injury as long as the injury continues to exist.

Many states carry a maximum medical payout that limits the medical exposure but in some states there is unlimited coverage for medical expenses, which can result in potentially large cost for insurers. Therefore, identifying fraudulent claims can potentially save millions of dollars. Some examples of PIP fraud include:

- Wage Loss Scams. In certain situations an insured can make more money by collecting wage loss payments from insurance companies than they actually earn at their job. For example, in some states insurance companies must pay 85% of an insured's gross wages. This is 100% less 15% for an adjustment based on what taxes may be. The 85%, however, is non-taxable. Unless an insured is normally taxed less than 15% they would be bringing home more than they would have with their employer. Further, payments from the employer and the insurance company are not always coordinated. An insured can receive sick leave, and/or disability benefits from an employer without reducing the benefits from the insurance company. This provides no incentive to get better, or when a person is healed, they may not disclose such information.
- Misrepresentation of Current Medical Coverage Status. When a new policy is written, it is written based on whether a prospective insured has other health insurance coverage. Potential medical expenses to an insurance company are a large factor in the underwriting decision. A policy for which the insured has no other medical coverage has a significantly higher premium than one for which they have other medical coverage that is primary (paid first) in the event of a loss (coordinated medical benefits-CMB). At policy inception a insured may claim to have other insurance coverage, reaping the lower rates of a CMB policy and in the event of a loss, when it is discovered no other insurance coverage exist, they are subject only to a deductible (perhaps \$300) on the medical coverage. The deductible is trivial compared to the savings of a CMB policy over time.
- Claiming other ailments not related to loss-With the risk of generalization, this type of fraud may fall among the older generation that can receive better medical benefits from an insurance company than they can from MEDICARE. Sometimes insureds will group all existing ailments into an auto loss claiming that is when the ailment began. Or they

may claim that a future ailment is related to the past incident which was covered by insurance.

- **Injury Only Claim-** Persons without personal medical insurance are still required to have PIP and liability coverage to operate an automobile in most areas. These persons may find that they cannot afford medical coverage and then try to pass the cost of an injury on to an auto insurer. By law, as long as the person is using the vehicle, an insurer must pay medical benefits. Law precedent has stated that if a person is exiting, entering or even touching the vehicle, it is being used in its prescribed manner. This generality has opened the door for persons without health insurance to pass costs on to auto insurers as long as they can provide a believable story.

Internal Investigation of Fraud

In the fight against insurance fraud, many insurance companies employ special investigation units to look at those claims that may be fraudulent. Insurance companies have a goal to make such investigations as objective as possible, but human judgments often prevail. Typically, claims are investigated when a field adjuster makes a recommendation to a claims investigation unit (CIU). Some adjuster's may make a fraud investigation recommendation while others may not. The following are representative of the indicators that might be used to evaluate claims. (Artis, 2002, p. 328; Tennyson, 2002, pp. 302-303; Viaene, 2002, pp. 377-379)

General Indicators

A claims history with previous thefts
New business, add car or recent endorsements
Comprehensive coverage only
Insured address is a PO Box, or different than the policy
Vehicle is stolen from a mall or large parking lot
Insured delays filing police report
Vehicle is recovered by the insured or association of insured
Insured's age
Urban vs. non-urban location
Same family name for insured and other party

Vehicle Indicators

Recovered burned and intact
Recovered with collision damage only
Ignition or steering column not defeated
Vehicle is 'clinically' stripped
Recovered condition does not match condition on report of loss
Damage to expensive stereo equipment
Keys with vehicle

Title/Ownership

- Late model with no lien
- Title holder and insured not the same
- Signs of VIN tag or Federal sticker tampering
- Insured wishes to retain salvage on an obvious total

A Model For Fraud Detection

It would be useful for insurance companies to have an objective model that could be used to help narrow the possible number of claims to be investigated for fraud by directing attention to those with relatively high probabilities of being fraudulent. Such a model could be developed based on known characteristics of fraudulent claims. The first type of model that might come to mind is standard (OLS) regression. However, this statistical procedure requires that the dependent variable be either integer or ratio data, but for our application the dependent variable is nominal: a claim is either fraudulent (a value of 1) or it is not fraudulent (a value of 0).

Logit (Logistic) Regression

Logistic regression is similar to OLS regression but is specifically designed to deal with a dichotomous dependent variable such as the one with which we are concerned. (Aldrich, 1984; Liao, 1994; Menard, 1995; Mendenhall, 1996) In this case the dependent variable is either zero (the claim is not fraudulent) or one (the claim is fraudulent). While we need not get too involved with the mathematical/statistical aspects of the logistic model in this paper, it may be helpful to show the general form of the model. It is:

$$P_i = P (Y=1 | X_{ik}) = \exp(b_k X_{ik}) \div (1 + \exp(b_k X_{ik}))$$

Where, P_i is the probability that $Y=1$ (a claim is fraudulent) given some set of characteristics of the set of independent variables (X_{ik}), and “exp” represents exponentiation. For example, $\exp(2)$ means to raise the Naperian number “e” to the second power [$\exp(2) = e^2 = 2.718^2 = 7.388$]. The natural logarithm of the odds (probability) is called the logit of Y. The logit of Y is estimated statistically then converted back to the odds by exponentiation ($P(Y=1 | X_{ik}) = e^{\text{logit}(Y)}$). We show explicit examples of this later in the paper.

Figure 1, illustrates the difference between a standard (OLS) regression linear function and the “S-curve” that results from a logistic function. Note that the “S-shaped” curve is bounded by zero and one and that it is relatively flat at the extremes and more steep in the middle. This suggests that the degree of the effect of a unit change in an independent variable will decrease near the upper and lower boundaries, where-as with linear regression the effect of a unit change in any independent variable is constant throughout.

 Figure 1 Goes About Here

The intent of using a mathematical model for fraud detection is not to make a definite decision about a claim being fraudulent at the time of the report of a claim, but rather to

determine whether there is statistically significant evidence that a claim is likely to be fraudulent. A logistic model can help to identify claims that may have a higher likelihood of fraud potential. This will allow for better prioritization of claims that are to be investigated.

Analysis

When a claim is filed with an insurance company, the insured must report specific details concerning the loss. Such details include, but are not limited to, the type the loss, the date of loss, whether a police report was filed, and various other details of the loss. In our analysis we take information from the initial report of loss and the claims history (CHIS), and use the information to determine whether the claim has the potential to have an element of fraud. That is: Does specific information reported or available at the time of the loss report, suggest that the claim may be fraudulent? Flagging these abnormal reports as they are submitted would provide support to the claims professional that these claims deserve particular attention.

The Sample

To estimate the logistic regression, data must be available for both fraudulent claims as well as legitimate claims. In cooperation with an insurance company's Claim Investigation Unit (CIU), claims have been provided that have been investigated, deemed fraudulent, and subsequently denied on the basis of fraud. Legitimate claims were taken from past claims that had not been denied based on fraudulent activity. The sample of legitimate claims are claims that did not have any CIU involvement or investigation. (Note: there is the possibility that a claim that was not investigated may have still contained an element of fraud. This is an unavoidable limitation of our analysis).

The category of claims that are analyzed for this paper were stolen and recovered vehicle claims. The abundance of information from these types of reports provides an opportunity to consider various scenarios. Stolen unrecovered vehicle reports are not be used, because there is a greater level of difficulty in denying these claims due to the lack of physical evidence for to review. The sample used consists of an equal number of claims denied on the basis of auto theft fraud and as well as auto theft claims that have been authorized and paid (49 of each). The sample size is small but these were the only data available. It would have been good to have a hold-out sample to test the model against but with less than 100 total observations this was not deemed to be practical.

Hypotheses

The dependent variable is whether or not the claim is fraudulent: 1 if yes and 0 otherwise. Six independent variables are considered. The first independent variable that is considered is the number of years an insured has been with the company (YRS). It is hypothesized that the longer an insured has been with the company, the stronger the relationship and loyalty that exists between the company and the insured. As the years of association increases, the lower the odds will be that the insured will file a fraudulent auto theft claim. Thus, we expect an inverse relationship (the sign of the YRS coefficient should be negative).

The second independent variable considered is the claims history (CHIS) of the insured. Claims history will be used as the total number of claims that have been filed with the company. It is predicted that the more claims that an insured files, the greater chance the insured is using the policy for opportunistic reasons, instead of indemnification. As the number of claims filed goes up, so should the odds of fraud. Thus, we expect a direct relationship (the sign of the CHIS coefficient should be positive).

The third independent variable is the number of claims submitted per year (CLMSYEAR). This is the total number of claims that an insured has filed divided by the number of years an insured has been a member. This relationship should be positive and is similar to CHIS but may serve as a better proxy of opportunism since a longtime member has a longer time period to have filed claims.

The fourth independent variable considered is whether an insurance policy is a Joint Underwriting Association policy (JUA). JUA policies are high-risk policies that an insurance company would not normally accept as a risk, but are placed by the State. Since insurance is mandatory, insurance companies must accept these placed high-risk policies. Because these policies are high risk, it is hypothesized that if a person holding a JUA policy files an auto theft claim, the odds that the claim will be fraudulent is higher than otherwise. JUA is measured as a dummy variable, coded as 1 for JUA and 0 for not JUA, so a positive relationship (coefficient) is expected.

Fifth, new business (NEWBUS) is considered as an independent variable. NEWBUS is a policy that is 1 year old or less. It is hypothesized that if a policy is new business the odds of fraud will go up. This is based on the idea that the insured has switched companies for reasons that may have been due to an adverse relationship with a prior insurance carrier. We would expect a positive relationship between FRAUD and NEWBUS.

The last independent variable is the time between when the insured filed the claim with the company and when the claim was reported to the police (DATEGAP). It is hypothesized that a person committing a fraudulent claim may not want to call the police, as they would be committing a felony if the claim were not legitimate. The longer the person waits to file a police report the longer they are considering whether their fraudulent actions are worth it. A person who files a legitimate claim would not worry about a felony and would file a police report immediately. We expect a positive relationship with respect to fraud. The longer and insured wait to fill a police report, the greater the likelihood of fraud.

Measures of Significance

Significance for the models will be determined by several measures. First, a Pearson Chi Square will be used to determine whether there is a relationship between the dependent and independent variables. The Chi Square statistic in this application tests the hypothesis that the independent variables being considered are not associated with the dependent variable (FRAUD). A large Chi Square indicates a lack of support for the notion that the independent variables being investigated are not associated with FRAUD, which means there is a significant relationship.

The second measure will be the -2 log likelihood. This measure provides for an indication that the model is better than it would have been without the addition of the independent variable. Therefore, with the addition of an independent variable to the equation, a

log likelihood value that decreases indicates that the model has been improved with the addition of the variable.

Pseudo R square values are also calculated (Cox & Snell and Nagelkerke pseudo R squares). This value is an indicator of the percentage of the variance in the dependent variable that is explained by the model.

Results¹

To avoid multicollinearity we first consider whether the independent variables are correlated. The Pearson product moment correlations for the set of six potential independent variables are shown in Table 1 (these correlations are low enough that multicollinearity is not a concern).

Table 1. Pearson Product Moment Correlations for Independent Variables

	DAT EGAP	Y RS	C HRIS	CLMS YEAR	J UA	NE WBUS
DATE GAP	1					
YRS	-.006	1				
CHIS	.031	.744	1			
CLMS YEAR	.025	-.419	-.015	1		
JUA	-.009	.248	.175	.226	1	
NEW BUS	.166	.412	.216	.388	.449	1

When we look at the association between FRAUD and each independent variable alone, using the (linear) Pearson product moment correlation we find the results shown in Table 2.

¹ Complete results for all analyses are shown in the Appendix.

Table 2. Correlations between FRAUD and each independent variable.
Values in parentheses are one tailed significance levels.

	DATEGAP	RS	Y	CH	CLMS YEAR	J UA	NE WBUS
Fraud	.142 (.082)	.247 (.07)	- (.01)	- (.23)	.321 (.000)	.298 (.001)	.399 (.00)

We see that the time between an incident and the report of the event (DATEGAP), and the claims history (CHIS) do not have a significant association with FRAUD (α .05). Comparable results are found, of course, when each independent variable is evaluated using a bivariate logistic regression model.

The logistic regression results using all six independent variables are shown in Table 3. In this model we see that the only significant variable is NEWBUS (α .05, one tail).

Table 3. Variables in the Equation - Full Model
Values in the top of each cell are the coefficients for the respective variables.
Values in parentheses are one tailed significance levels.

CONSTANT	DATEGAP	RS	Y	CH	CLMS YEAR	J UA	NE WBUS
-1.054 (.043)	.139 (.136)	.021 (.310)	- (.90)	.01 (.3)	.543 (.107)	.182 (.365)	1.06 (.038)

Evaluating alternative models the best results are obtained when both claims per year (CLMSYEAR) and NEWBUS are included. The results are shown in Table 4.

Table 4. Variables in the Best Model.
Values in the top of each cell are the coefficients for the respective variables.
Values in parentheses are one tailed significance levels.

CONSTANT	CLMSYEAR	NEWBUS
-1.135 (.003)	.671 (.028)	1.601 (.002)

The resulting equation is: $\text{Logit } Y = -1.135 + 0.671(\text{CLMSYEAR}) + 1.601(\text{NEWBUS})$.
The signs for the coefficients are consistent with expectations and the coefficients are statistically

significant. When the above model is used to classify claims as fraudulent or not fraudulent the classifications shown in Table 5 result.

Table 5. The Final Classification Table.

	Predicted		
Observed	No Fraud	Fraud	Percentage Correct
No Fraud	40	9	81.6
Fraud	20	29	59.2
Overall Percentage Correctly Classified			70.4

The percentage correctly predicted is 70.4%, which is better than a random prediction which would yield 50%, therefore the model does provide better explanation using these two independent variables. We also see that the model does a better job predicting legitimate claims (81.6%) vs. fraudulent claims (59.2%).

Values for the independent variables can be substituted in the formula to provide a likelihood estimation (odds) of fraud. Using the estimated equation with observed values for the independent variables a estimation of the likelihood of fraud could be computed. First, let us consider a claim for a policy holder who has an average of 1 claim per year and does not represent a new policy:

$$\text{Logit } Y = -1.135 + 0.671(\text{CLMSYEAR}) + 1.601(\text{NEWBUS})$$

$$\text{Logit } Y = -1.135 + 0.671(1) + 1.601(0)$$

$$\text{Logit } Y = -0.464$$

$$e^{-0.464} = 0.629$$

Recall that $P_i = \exp(\beta_k X_{ik}) \div (1 + \exp(\beta_k X_{ik}))$, so we find that given this scenario, the probability that this person's claim is fraudulent is 38.6% ($= 0.629 \div 1.629$).

To contrast, let's consider the same situation, only this time with a person who is a new policy holder in the past year. We now have:

$$\text{Log odds} = -1.135 + 0.671(\text{CLMSYEAR}) + 1.601(\text{NEWBUS})$$

$$\text{Log odds} = -1.135 + 0.671(1) + 1.601(1)$$

$$\text{Log odds} = 1.137$$

$$e^{1.137} = 3.117$$

The percentage probability that this person's claim may be fraudulent rises to 75.7% ($= 3.117 \div 4.117$). Assuming a 50:50 cutoff is used, the first scenario would not warrant additional follow-up by an investigation unit, however, the second scenario would.

Managerial Implementations

Before a model is implemented which will determine which claims should be investigated and which should not, it should be determined whether the logistic model provides better

predictions than whatever current method is in use. Then during implementation the model should be run automatically from reports of loss as they are filed.

The model reported above is for automotive insurance. Because there is the possibility of fraud in all types of insurance claims a separate model should be specified for each category of insurance coverage. Once logistic models are created from sample data within each claim discipline, running probability percentages for potentially fraudulent claims would be relatively easy. Any spread sheet (such as Excel) could be programmed with the equation and the resulting odds and percentage probabilities would be computed automatically based on the logistic model. The report of potentially fraudulent claims printed based on the logistic model could then be used and claims could be prioritized based on the percentage probabilities that indicate higher odds of fraud. It should be stressed that this logistic model in predicting fraud is only a tool and that the determination of fraud would not be based on the model but only after complete and thorough investigation.

References

- Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models*. Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 45. Beverly Hills, CA:Sage.
- Artis, M., & Mercedes, A., & Montserrat, G. (2002, September). Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims. *The Journal of Risk and Insurance*.
- Fraudulent Auto Theft Claims Targeted By Office Of Insurance Fraud Prosecutor's Training Program. (2002, September 3). *Insurance Fraud News*. Retrived February 10, 2004, from www.njinsurancefraud.org/release/2002/video0903.htm.
- Liao, T. F. (1994). INTERPRETING PROBABILITY MODELS Logit, Probit, and Other Generalized Linear Models. *Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 101*. Beverly Hills, CA:Sage.
- Menard, S. (1995). *Applied Logistic regression Analysis*. Sage University Paper Series, No. 106. Beverly Hills, CA:Sage.
- Mendenhall, W., & Terry, S. (1996). *A Second Course in Statistics*. Prentice-Hall. Upper saddle River, NJ.
- "News." Coalition Against Insurance Fraud. (2003, February 13). Retrived February 15, 2004, from www.insurancefraud.org/study021303.htm. The survey reported was conducted by Taylor Nelson Sofres Intersearch and included 1,030 U. S. Adults. It was conducted in November 2002.
- Snider, H., & Tam, S. (June 3, 1996). Detecting Fraud With Insurance Data Warehouse. *National Underwriter*. Retrieved February 11, 2004, from www.businessguidance.com/Publications/Article_Detecting_Fraud.html.

Appendix: Complete Statistical Results for Final Model

Pearson Product Moment Correlations for Independent Variables

	DATEGAP	YRS	CHIS	CLMSYEAR	JUA	NEWBUS
DATEGAP	1					
YRS	-.006	1				
CHIS	.031	.744	1			
CLMSYEAR	.025	-.419	-.015	1		
JUA	-.009	-.248	-.175	.226	1	
NEWBUS	.166	-.412	-.216	.388	.449	1

Classification Table: No Independent Variables^a

Observed		Predicted				
		FRAUD		Percentage Correct		
		.00	1.00			
Step 0	FRAUD	.00		0	49	.0
		1.00		0	49	100.0
Overall Percentage						50.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Best Equation

Step		B	S.E.	Sig.
1 ^a	CLMSYEAR	.671	.350	.055
	NEWBUS	1.601	.546	.003
	Constant	-1.135	.405	.005

a. Variable(s) entered on step 1: CLMSYEAR, NEWBUS.

Model Summary: Best Model

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	115.470	.188	.250

Omnibus Tests of Model Coefficients: Best Model

Step		Chi-square	df	Sig.
1	Step	20.387	2	.000
	Block	20.387	2	.000
	Model	20.387	2	.000

Classification Table: Best Model^a

Observed		Predicted				
		FRAUD		Percentage Correct		
		.00	1.00			
	FRAUD	.00		40	9	81.6
		1.00		20	29	59.2
Overall Percentage						70.4

a. The cut value is .500

Figure 1. Comparison of the Linear and the “S-Shaped Logistic Regression Lines.

