IMPACT OF PRODUCT VARIETY ON INVENTORY PERFORMANCE: AN EMPIRICAL ANALYSIS OF PHARMACEUTICAL AND MEDICAL DEVICE INDUSTRIES

A THESIS

SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Ali Yağmur Aydınlı

September, 2012

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Alper Şen(Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Osman Alp

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Doğan Serel

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

IMPACT OF PRODUCT VARIETY ON INVENTORY PERFORMANCE: AN EMPIRICAL ANALYSIS OF PHARMACEUTICAL AND MEDICAL DEVICE INDUSTRIES

Ali Yağmur Aydınlı
M.S. in Industrial Engineering
Supervisor: Asst. Prof. Dr. Alper Şen
September, 2012

In this thesis, we construct empirical models to analyze the effect of various financial measures and product variety on inventory performance. We apply multiple regression models and simultaneous equations models using 2010 financial data for 128 U.S. firms in pharmaceutical and medical device industries. Using multiple regression models, we investigate the correlation of inventory turnover with product variety, gross margin, capital intensity and type of the firm. Product variety data are obtained from U.S. Food and Drug Administration. The best multiple regression model explains 38.50% of the total variation in inventory turnover and we observe that inventory turnover is negatively correlated with product variety. Since inventory turnover is the ratio of cost of goods sold (sales) to inventory level, a change in product variety might have an effect both on inventory level and cost of goods sold. In order to investigate the effects of product variety on cost of goods sold and inventory level separately, we employ simultaneous equation models. The best simultaneous equation model explains 96.22% of the variation in inventory level and shows that inventory level is positively associated with product variety.

Keywords: Inventory turnover, inventory level, product variety.

ÖZET

ÜRÜN ÇEŞİTLİLİĞİNİN ENVANTER PERFORMANŞINA ETKISI: İLAC VE TIBBİ CİHAZ ENDÜSTRİSİ AMPİRİK ANALİZİ

Ali Yağmur Aydınlı
Endüstri Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Asst. Prof. Dr. Alper Şen
Eylül, 2012

Bu çalışmada, çeşitli finansal ölçütlerin ve ürün çeşitliliğinin envanter performansı üzerinde etkilerini analiz eden ampirik modeller geliştirilmiştir. Bu kapsamda, A.B.D sağlık sektöründe faaliyet gösteren 128 ilaç ve tıbbi cihaz firması verileri, çoklu regresyon modelleri ve eş-zamanlı denklemler modelleri kullanılarak analiz edilmektedir. Çoklu regresyon modelleri ile, envanter dönüşüm hızının; brüt kâr marjı, sermaye yoğunluğu, ürün çeşitliliği ve firmanın faaliyet gösterdiği sektör arasındaki korelasyonları incelenmektedir. Ürün çeşitliliği verileri Amerikan Gıda ve İlaç Kurumu'ndan (FDA) alınmıştır. En iyi çoklu regresyon modeli, envanter dönüşüm hızındaki varyasyonun %38,5'ini açıklayabilmektedir. Ayrıca, ürün çeşitliliği ile envanter dönüşüm hızı arasındaki korelasyonun, beklendiği üzere, negatif olduğu görülmüştür. Fakat, envanter dönüşüm hızı satışların maliyetinin envanter seviyesine oranı olduğundan, ürün çeşitliliğinin hem satışların maliyetine hem de envanter seviyesine etkisi bulunmaktadır. Ürün çeşitliliğinin, envanter seviyesine ve satışların maliyetine olan etkisini ayrı ayrı analiz edebilmek için eş-zamanlı denklemler modelleri kullanılmıştır. En iyi eş-zamanlı denklemler modelinin envarter seviyesindeki varyasyonunun %96,22'sini açıkladığı gözlemlenmiş ve envanter seviyesinin ürün çeşitliliği ile bağlantısının pozitif olduğu görülmüştür.

Anahtar sözcükler: Envanter dönüş hızı, envanter seviyesi, ürün çeşitliliği.

To my parents...

Acknowledgement

First, I would like to express my deepest gratitude to my advisor Asst. Prof. Dr. Alper Şen for his invaluable support, guidance, knowledge and patience throughout my studies. I attribute the level of my Masters degree to his encouragement and effort. I consider myself lucky to have had the opportunity to work under his supervision, one simply could not wish for a better supervisor.

I am indebted to Assoc. Prof. Dr. Osman Alp and Asst. Prof. Dr. Doğan Serel for devoting their valuable time to read and review this thesis and their substantial suggestions.

I would also like to thank my colleague and friend Korhan Aras from North Carolina State University for working with us in this project and providing us the data.

I am deeply grateful to my parents, if not for their moral support and academic guidance, I would not be able to complete neither my graduate studies nor my thesis. I want to thank my twin, Alp Aydınlı, for his unbending love and support not only throughout my studies but also through life. My family's love, belief and support has been a strength for me in every part of my life. Also, I would like to thank to my fiancee Işıl Ünal, who has also believed in me in every step.

Many thanks to my friends Ece Demirci, Mehmet Arıkan and Ali Can Ergür for their moral support and help during my graduate study and all other friends I failed to mention here for their friendship and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Drug market is a rapidly growing market. Standard and Poor's estimates that the generic drug market will grow from $ 83 billion in 2009 to $ 135 - $ 150 billion in 2015 [1]. The largest share of the market belongs to Teva with 21%. Novartis AG (10%), Mylan Inc. (6%), Watson Pharmaceuticals Inc. (6%), STADA Arzneimittel AG (3%), Actavis (2%), Ranbaxy Laboratories Ltd. (2%) making up the pack following Teva. The rest of the market is shared by many firms, which add up to 50% [1].

Prescription drugs are another key driver of growth for health care firms. Pfizer, AstraZeneca, Bristol-Myers and GlaxoSmithKlein are the top sellers of prescription drugs. Many drugs' patents will expire in the coming years, the prescription drugs will become generic drugs, hence they will be open for competition. Therefore, health care companies will be considering new cost cutting strategies for growth. The cost cutting initiatives include but are not limited to reduction of overhead cost, reduction of inventories held and other cost cutting programs. We analyze the inventory management and supply chain strategies that will benefit decision makers in the industry and guide them in their search for efficiency and cost cutting strategy.

In this scope, we have obtained public financial data of the related companies from Wharton Research Data Services's (WRDS) [2] and product variety data

from Food and Drug Administration's web site. FDA monitors the manufacturing process, the storage process and the transportation process of health care products in USA. In this highly regulated environment, it is becoming difficult for companies to cut costs and effectively manage the company. Apart from cutting overhead costs, to get one step ahead in a highly competitive environment, companies need to improve their supply chain structures and inventory systems. Hence, optimizing and understanding the variables that effect inventory turnover and inventory level becomes a critical issue for success.

Inventory managers usually measure the inventory turnover, which is the ratio of cost of goods sold to the inventory level, to understand the performance of the company's inventories. The inventory turnovers of major firms vary widely from each other, representing opportunities in efficiently managing inventories and the related supply chain. For example, the annual inventory turnover of Teva Pharmaceutical Industries Ltd. is around 1.5530, whereas the values for Mylan Inc., Watson Pharmaceuticals Inc. and Bristol-Myers Squibb Co. are around 2.2574, 2.9576 and 3.6710, respectively.

Even though other factors within a company can influence inventory turnover, we are mainly interested in the effects of product variety. To our knowledge a research of this scale has not been put forward.

Our thesis uses public financial data obtained from Wharton Research Data Services's (WRDS) CompuStat Database and product variety data obtained from FDA to understand and evaluate the relationship between the inventory turnover and product variety. The WRDS provides gross margin, capital intensity, inventory level, cost of goods sold and sales data of the chosen firms from medical device and pharmaceutical industries. We use the existing literature to formulate the hypothesis and use the data we obtained to draw conclusions about them. We have set up a similar hypothesis as Gaur et al. [3]. The primary contribution of this thesis is the inclusion of product variety to explain the variation in inventory turnover, Hypothesis 3, and inventory level, Hypothesis 4.

To test these hypotheses, first, we apply multiple regression models and observe that only 30%-39% of the variability in inventory turnover can be explined

by the variables. Our main findings from the multiple regression models point out that the inventory turnover rate can be explainable by the exogenous variables. The best multiple regression model we have can explain approximately 39% of the variation.

The relationship between product variety and inventory turnover is vague because a change in product variety affects both cost of goods sold and inventory level, which together make up the ratio of inventory turnover. One main goal of this paper is to explain this vague relationship among product variety, inventory level and cost of goods sold. We develop Hypothesis 4 in order to understand the relative effect of product variety on inventory levels. We believe that product variety increases inventory level relatively more than cost of goods sold.

We incorporate simultaneous equations models, in particular two stage least square estimation to analyze the relationship between product variety and inventory level. The models we built can explain approximately 90 to 96% of the total variations in inventory level.

This thesis contributes, empirically, to inventory management and supply chain research in health care industry. As the competition increases in drugs and devices industry tackling the problem of inventory management becomes a significant issue. We believe that our model explains the role of product variety on inventory turnover ratio and inventory level across firms.

The rest of the thesis is organized as follows; in Chapter 2 the pertinent literature is summarized, in Chapter 3 the methods and models used is explained throughly, Chapter 4 provides information on the data sources we used, Chapter 5 describes the hypotheses, in Chapter 6 we construct multiple regression models and simultaneous equations models. After giving each multiple regression model's results we test for the regression assumptions for each model. For SEM, we solve them by using two stage least square estimation, test for endogeneity bias and overidentification then summarize our findings. Finally, we draw conclusions of the results and possible extensions in Chapter 7.

Chapter 2

Literature Review

In this chapter, we review the important empirical and theoretical literature related to our research. Inventory management has been an area of interest in operations research. There is a vast amount of literature in this area (starting from the EOQ model that was developed nearly a century ago and advanced models that include stochastic demand, multiple products and multiple echelons). However, empirical research in this area is relatively new. We have reviewed the papers that use actual macroeconomic data and theoretical papers, whose main focus is inventory management and supply chain management.

Gaur et. al [3] incorporate cost of goods sold, inventory level and gross margin as endogenous variables to improve the firm-level sales forecast. Data of retailers are driven from the CompuStat database. The data consist of 230 observations from 2004 to 2007 of retail firms. Authors define simultaneous equations model to represent the forecasting model's basis. The authors use the analysts' consensus forecasts to compare with their forecast results. Absolute percentage error (APE) and mean absolute percentage error (MAPE) are used to compare the results with analysts' forecasts. Furthermore, the authors investigate the reasons behind the differences between their forecasts and analysts' forecasts by examining inventory and gross margin residuals. The authors have managed to get more accurate forecasts than the analysts, since analysts do not take into account the inventory and gross margin data in their sales forecasts. However, the authors fail to

incorporate the effects of product variety on inventory, cost of goods sold and gross margin.

Gaur et. al [4] develop econometric models to analyze the effect of gross margin, capital intensity and sales surprise on inventory performance. The authors use data from 311 publicly listed retail firms to test their hypothesis. The authors manage to show that inventory turnover, in retail services, have a high correlation with gross margin, capital intensity and sales surprise. The paper proves that as gross margin increases inventory turnover decreases, capital intensity and sales suprise are positively correlated with inventory turnover. Moreover, they conclude that inventory should not be treated as the only variable that is used in performance analysis. In our thesis, we prove authors' Hypothesis 1 and Hypothesis 2 using multiple regression models for health care industry. The authors suggest to use product variety as another explanatory variable. We have extended this study and incorporated product variety as one of the explanatory variables in multiple regression models to analyze its effect on inventory turnover.

Rajagopalan and Malhotra [5] study the trends in finished goods materials and work in process inventory ratios of 20 U.S manufacturing companies during 1961-1994. The authors use a simple linear regression model to investigate the rate of change in inventory ratios over years and setup hypotheses to test the regression model. The hypothesis tests formed are related to the inventory ratios and their relation with the time period and the respective industry sector. The article goes on to find out that as the inventory theory improved, materials and work in process inventories started to decrease in most of the cases studied. Moreover, industry sectors such as furniture and fabricated metal products showed improvement in work-in-process inventory ratios, while sectors such as textiles, lumber, printing/publishing, rubber, and machinery showed improvement in work-in-process inventory, finished goods inventory and materials and supplies inventory.

Cachon and Olivares [6] measure the effect of days of supply, sales rate, production flexibility and number of dealerships on finished goods inventory levels in U.S. auto manufacturing industry, mainly comparing Toyota, Chrysler, Ford and

GM. The authors find that, the differences between the levels of inventory among Toyota, Chrysler, Ford and GM arise from differences in production flexibility and number of dealerships. Flexibility in production enables the firm to easily track the demand of the customer. The authors argue that this can be accomplished by changing the production levels when there is a change in sales level. Fewer number of dealerships allows the firms to pool demand in fewer locations. The paper concludes that Toyata has these set of advantages, which enables it to have a competitive advantage over rival firms in U.S. auto industry.

Roumiantsev and Netessine [7] examine absolute and relative inventories of companies based on an empirical data of 722 U.S. public companies for the years, between 1992-2002. The authors' multiplicative model explains 85 percent of the absolute inventory in these firms. Moreover, the paper argues that as gross margin increases so does the inventory level. Companies that can attain economies of scale have less inventory in relative terms. Therefore, the paper concludes that aggregate inventory level is positively linked to mean demand and demand uncertainty. Additionally, relative inventory levels are negatively associated with company size, since both absolute and relative inventory increase with lead times and product margins.

Cachon et. al [8] study the bullwhip effect using empirical industry level U.S data. The bullwhip effect is defined as the increase in variability of demand as one goes upstream in the supply chain. It is usually observed in forecast driven distribution channels. The paper conducts its empirical research on 6 retail, 18 wholesale and 50 manufacturing industry sectors, whose data are taken from the U.S Census Bureau. The authors measure the variance of production and the variance of demand of customers in downstream for each sample. The paper concludes that there is no bullwhip effect among retailers and manufacturers. They go on to find that seasonality has great influence on bullwhip effect.

Kekre and Srinivasan [9] explore the advantages and disadvantages of having a broader product variety using 1400 firms. The authors use Strategic Business Unit data obtained from Profit Impact Strategy (PIMS) and use econometric analysis to test the hypotheses. One of the major hypothesis the authors put

forward is; as product variety increases so does level of inventory. But they, also, believe that an increase in product variety will increase the market share of the firm and enable it to use economies of scale by pushing suppliers to meet its terms, thus, reducing inventory level. But we mainly believe that variety increase brings an increase in inventory level, in health care industry. The key finding of the authors is; higher product variety leads to a higher market share and increased profitability. However, the cost effect of having higher product variety cannot be tested using empirical analysis.

Ton and Raman [10] analyze the effect of product variety on the level of inventory held and the amount of sales made in Borders stores across USA. The authors believe that having high product variety and high inventory levels lead to higher sales in stores. However, there are some drawbacks of maintaining a high variety and a high level of inventory. The authors prove with an econometric model that an increase in product variety leads to an increase in phantom products, i.e, the products that are in the inventory but not on the shelf, therefore the customers are not able to make a purchase. Increase in inventory levels leads to an increase in the percentage of unidentified phantom product percentage by customers. Lastly, as the number of phantom products increases there is a decrease in the sales of the related store. The paper uses regression analysis to prove the hypothesis on phantom products.

Mahajan and van Ryzin [11] use a multinomial logit model to identify the purchase quantities for each of the product's varieties. The paper tries to capture the trade-off between the costs of having a high product variety and a low variety. The authors use a stochastic choice process for customer purchase decision using a multinominal logit random utility model. For the retailers' supply process the authors use the newsboy model. The paper proves that high variety becomes profitable in high volume of business because big businesses can utilize economies of scale more efficiently.

We also analyze the theoretical work behind the variables effecting inventory levels. These papers include Marvel and Peck [12], Balakrishnan et. al [13] and Cachon [14]. These papers enable us to develop our hypotheses and effectively

identify and measure variables that induce inventory management systems.

Marvel and Peck [12] study the vertical misalignment of inventories and demonstrate its impact on product variety. The paper uses game theory to analyze the problem. The game consists of three stages; choice of the manufacturer wholesale price, choice of price for retailer for each of the product varieties and choice of the products that retailers want to carry. The game is studied over an infinite time horizon. Every retailer must first choose its price then its inventory level of the product. The paper concludes that a retailer carrying few product varieties and high inventory turnover can achieve high market share. The authors suggest that high inventory turnover ratio can be achieved by reducing prices to keep the inventory levels low.

Balakrishnan et. al [13] analyze the effect of stock keeping on firm's profit level. The authors define the objective function as a profit maximization problem. Then, they use EOQ model to incorporate the effects of cost and revenue ingredients. The paper finds that the excess inventory can, indeed, increase demand. Additionally, the authors prove that for items that have low holding costs, it is optimal to have inventory levels above the reorder level. Since the classical EOQ model does not include the demand stimulation of inventory it orders too little and too frequent making it a suboptimal policy.

Cachon [14] investigates supply chain demand volatility using a model with a single supplier and N retailers with stochastic demand. The paper identifies that using scheduled ordering policies, like ordering every T periods with an integer multiple of a fixed batch size, usually lowers the total supply chain costs. Moreover, if the supplier's demand volatility decreases the supplier's average inventory will decrease. In order to do this, two strategies have been identified. First, one can increase the number orders by increasing the period, T, and reduce the order sizes. Second, the retailer can balance its order intervals. However, the authors note that by increasing the order period the supply chain costs increase.

Chapter 3

Methodology

Our main contribution in this research is to study the effect of product variety in addition to gross margin, capital intensity and cost of goods sold on inventory performance of a firm to identify the effects on inventory turnover and firm's performance. We use empirical data retrieved from Wharton Research Data Services (WRDS) [2], CompuStat database. We limit our research to analyzing pharmaceutical drug and medical device companies. We use STATA software [15] to implement the multiple regression method and the simultaneous equation models. We use the STATA command *regress* for multiple regressions and *ivregress 2sls* to handle simultaneous equation models. In this chapter we review the methods used in our analysis.

3.1 Multiple Regression

We have used the multiple linear regression model to test some of our hypothesis. This model has the advantage of identifying the independent variables that have an effect on the dependent variable. The generic form of a multiple linear regression model is given as follows;

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_k x_{tk} + \epsilon_t. \tag{3.1}$$

In the above equation y_t is the dependent variable and $x_{t2}, x_{t3}, ..., x_{tk}$ are called independent variables. The ϵ_t term is the disturbance term or the error term in

the equation. Greene [16] gives the assumptions of this model as:

1. **Linearity**; the linear relationship between the dependent variable and the independent variables.

2. **Full Rank**; there is no or little correlation between the independent variables.

3. **Exogeneity and independent variables**; the expected value of the error term given all the independent variables equal to zero. That is;

$$E[\epsilon_t | x_{t2}, ..., x_{tk}] = 0,$$

meaning that the independent variables do not carry information about ϵ_i.

4. **Homoscedasticity and non-autocorrelation**; ϵ_i the error terms, have the same variance, σ^2, and they are uncorrelated with the other error terms ϵ_j.

5. **Normal distribution**; the error terms are normally distributed.

Furthermore, since the dependent variable, y_t depends on the random error term ϵ_t, it is also a random variable. The properties of the dependent variable y_t are as follows:

1. When we take the expected value of the multiple regression model, since $E[\epsilon_t] = 0$, the model becomes;

$$E[y_t] = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk}.$$

2. $var(y_t) = \sigma^2$. The variance of the dependent variable y_t does not differ with each and every observation.

3. $cov(y_t, y_s) = 0$, meaning that any two observations on the dependent variable are uncorrelated. If any of the observations is above the expected value of the dependent variable $E[y_t]$, any other subsequent observation will be above

$E[y_t]$.

4. The values of the dependent variable, y_t, are assumed to be normally distributed.

The multiple regression analysis enabled us to estimate, forecast and construct test statistics to determine the significance of the independent variables on the model using hypothesis testing.

In order to fully understand the least square estimation procedure consider the following multiple regression model with two independent variables:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \epsilon_t.$$

Carter [17] describes the least square estimation procedure for the multiple regression model as a procedure that minimizes the sum of squared differences between the observed values of y_t and their expected values. Thus, the above multiple regression model becomes;

$$E[y_t] = \beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3.$$

Then, we minimize the sum of square function $S(\beta_1, \beta_2, \beta_3)$ that is;

$$S(\beta_1, \beta_2, \beta_3) = \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_{t2} - \beta_3 x_{t3})^2. \tag{3.2}$$

The solution to the least square estimation is the least square estimates b_1,b_2,b_3. The deviations of each variable from their means are;

$$y_t^* = y_t - \bar{y}, x_{t2}^* = x_{t3} - \bar{x}_3. \tag{3.3}$$

Thus, b_1,b_2 and b_3 can be found by solving;

$$b_1 = \bar{y} - b_2\bar{x_2} - b_3\bar{x_3}, \tag{3.4}$$

$$b_2 = \frac{(\sum y_t^* x_{t2}^*)(\sum x_{t3}^*) - (\sum y_t^* x_{t3}^*)(\sum x_{t2}^* x_{t3}^*)}{(\sum x_{t2}^{*2})(\sum x_{t3}^{*2})(\sum x_{t2}^* x_{t3}^*)^2}, \tag{3.5}$$

$$b_3 = \frac{(\sum y_t^* x_{t3}^*)(\sum x_{t2}^{*2}) - (\sum y_t^* x_{t2}^*)(\sum x_{t3}^* x_{t2}^*)}{(\sum x_{t2}^{*2})(\sum x_{t3}^{*2})(\sum x_{t2}^* x_{t3}^*)^2}. \tag{3.6}$$

These formulas are used to calculate the least square estimates for (3.2) using the given data. In implementation, we use STATA software to calculate b_1, b_2 and b_3. To estimate the variance (σ^2) of the error term, ϵ_t an estimation procedure has been developed. This procedure uses the least square residuals, which represents the sample information about the error term. The least squares residuals method for (3.2) are written as;

$$\widehat{\epsilon_t} = y_t - \widehat{y_t} = y_t - (b_1 + x_{t2}b_2 + x_{t3}b_3), \tag{3.7}$$

An estimator of σ^2 would be

$$\widehat{\sigma}^2 = \frac{\sum \widehat{\epsilon_t}^2}{T - K}, \tag{3.8}$$

where K is the number of parameters estimated in the multiple regression model and T is the sample size.

In the multiple regression model it is assumed that the least square estimators are unbiased. If the variances are high, then the probability of producing estimates close to the true parameter value will be higher. For example when we have only three parameters to estimate, meaning $K = 3$ we can formulate the variance and covariance of the least square estimator of b_2 as such;

$$var(b_2) = \frac{\sigma^2}{\sum (x_{t2} - \overline{x}_2)^2(1 - r_{23}^2)}, \tag{3.9}$$

where r_{23} is the correlation coefficient between the values x_2 and x_3. The formula of the correlation coefficient, r_{23} is given as;

$$r_{23} = \frac{\sum (x_{t2} - \overline{x}_2)(x_{t3} - \overline{x}_3)}{\sqrt{\sum (x_{t2} - \overline{x}_2) \sum (x_{t3} - \overline{x}_3)^2}}. \tag{3.10}$$

Other variances and covariances have similar formulas. Factors affecting the variance of b_2 can be listed as:

1. Since variance (σ^2) is the uncertainty in the model, the greater the variance of the least square estimator is, the greater the error variance will be.

2. As the sample size increases the variance of the least square estimator decreases. Because the larger the sample size is the larger the value of $\sum (x_{t2} - \overline{x}_2)^2$

will be. Thus, the more accurate the estimator becomes.

3. The larger $\sum (x_{t2} - \overline{x}_2)^2$ becomes, which is the variation of the explanatory variable around its mean, the smaller the variance of the least square estimator.

4. As the correlation coefficient, r_{23} grows larger, variance of b_2 increases.
For example, if we wanted to estimate three parameters of a multiple regression model the variance and covariance matrix, $\text{cov}(b_1, b_2, b_3)$ would be like:

$$cov(b_1, b_2, b_3) = \begin{pmatrix} var(b_1) & cov(b_1, b2) & cov(b_1, b3) \\ cov(b_1, b2) & var(b_2) & cov(b_2, b3) \\ cov(b_1, b3) & cov(b_2, b3) & var(b_3). \end{pmatrix}$$

3.1.1 Significance Test

We have constructed a hypothesis test to identify the significance of our variables. The null and alternative hypotheses we have specified are;

H_o: variable is significant

H_a: variable is not significant

Based on the test statistics we either reject the null hypothesis or fail to reject it. Moreover, when evaluating the results of the hypothesis tests we have calculated the p-value in order to either reject or accept the null hypothesis. If the p-value of a hypothesis test is smaller than the chosen significance level, α, then the conclusion is to reject the null hypothesis. Consequently, if the p-value is greater than the significance level we fail to reject the null hypothesis. For all of our hypotheses, we have tested whether the independent variable has an effect on the dependent variable or not.

We take the logarithms of the variables in order to smooth out the fluctuations in our observations. This usually gives us a more reliable estimation.

3.1.2 R-Squared and Adjusted R-Squared

In a multiple regression model, R^2 measures the proportion of variation in the dependent variable explained by all the explanatory variables in the model [17]. The logic and formulas behind R^2 in a multiple linear regression model is exactly the same as the simple linear regression model. To explain R^2, consider a simple regression model;

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t. \tag{3.11}$$

For this model, R^2 states how much of the variation in x_t can explain the variation in the dependent variable, y_t. To measure the variation in y_t we divide the dependent variable, y_t, into explainable and unexplainable components. Further assume that;

$$y_t = E(y_t) + \epsilon_t, \tag{3.12}$$

where ϵ_t is the unexplainable portion of y_t, while $E[y_t] = \beta_1 + \beta_2 x_t$ is the explainable portion of the dependent variable. We estimate β_1 and β_2 by decomposing y_t into;

$$y_t = \widehat{y}_t + \widehat{\epsilon}_t, \tag{3.13}$$

where $\widehat{y}_t = b_1 + b_2 x_t$ and $\widehat{\epsilon}_t = y_t - \widehat{y}_t$.

Then, by subtracting sample mean of the least squares fitted line, $\overline{y}$ from both sides of the equation

$$y_t - \overline{y} = (\widehat{y}_t - \overline{y}) + \widehat{\epsilon}_t \tag{3.14}$$

In the above statement $(\widehat{y}_t - \overline{y})$ is the explained part of the equation, while $\hat{\epsilon}_t$ is the noise part. Moreover, to measure the total variation in a variable we square the differences between y_t and its mean value $\overline{y}$ and sum it over the entire sample.

That is;

$$\sum_t (y_t - \overline{y})^2 = \sum_t [(\widehat{y}_t - \overline{y}) + \widehat{\epsilon}_t]^2$$
$$= \sum_t (\widehat{y}_t - \overline{y})^2 + \sum_t \widehat{\epsilon}_t^2 + 2\sum_t (\widehat{y}_t - \overline{y})\widehat{\epsilon}_t$$
$$= \sum_t (\widehat{y}_t - \overline{y})^2 + \sum_t \widehat{\epsilon}_t \tag{3.15}$$

Since the term $\sum (\widehat{y}_t - \overline{y})\widehat{\epsilon}_t = 0$ it has been dropped out of (3.15). In order to calculate the R^2 we need to identify the sums of squares;
1. Total sum of squares, SST, is the measure of total variation in y about its sample mean and it is given as; $\sum_t (y_t - \overline{y})$
2. Explained sum of squares, SSR, is the part of total variation in the dependent variable, y, about its sample mean that is explained by the regression model and is shown by $\sum (\widehat{y}_t - \overline{y})^2$
3. Error sum of squares, SSE, is the total variation in y about its mean that cannot be explained by the regression $\sum \widehat{\epsilon}_t^2$.

The equation (3.15) becomes $SST = SSR + SSE$. The degrees of freedom for the above sum of squares are given in the below table:

Source of Variation	DF	Sum of Squares	Mean Square
Explained	1	SSR	SSR/1
Unexplained	N-2	SSE	SSE/(N-2)
Total	N-1	SST	

Table 3.1: Analysis of Variance

In Table 3.1, the first row of the Mean Square column is the ratio of SSR to its degrees of freedom and the second row is the ratio of SSE to its degrees of freedom, which equals to $\widehat{\sigma}^2$.

Lastly, the proportion of variation in the dependent variable, y, is explained by x within the regression model, is shown by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3.16}$$

If R^2 equals to 1 then all the sample data fall on the fitted least squares line and SSE =0. If y and x are not correlated then the least square fitted line is horizontal and identical to $\overline{y}$ and $SSR = 0$ in which case the $R^2 = 0$. When $0 < R^2 < 1$, the percentage of variation in the dependent variable, y, about its mean that is explained by the regression model [17].

3.3 Simultaneous Equations Models

We, also use simultaneous equations models (SEM). Simultaneous equations models consist of a set of equations rather than just a single equation and there are two or more dependent variables in each model . Ordinary least square estimation procedure is not appropriate in these models since these models do not adress endogeneity bias. In order to fully understand the simultaneous equations models, consider a simple SEM such as [18]:

$$y_1 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_k x_k + \beta_1 y_2 + \epsilon, \tag{3.17}$$

$$y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \beta_{k+1} z_1 + ... + \beta_{k+m} z_m + \nu, \tag{3.18}$$

$$y_2 = \hat{y_2} + \nu, \tag{3.19}$$

where y_2 is an endogenous variable. There are m instruments, namely $z_1, z_2, ..., z_m$ that are correlated with y_2. Let $z = (1, x_1, ..., x_K, z_1, ..., z_m)$. $\hat{y_2}$ is a linear projection of y_2. Note that, $\hat{y_2}$ is not correlated with ϵ but ν is correlated with ϵ. When there is a small change in the error term, say ϵ, this effect is directly transmitted to the value of x_k. The failure of the least square estimation for the first equation arises cause of the effects triggered by the changes in the error terms. This happens because we do not observe the change in the error term but rather the change in x_k, which results from its correlation with the error term. The least square estimator of β_k will underestimate/overestimate the true parameter value in the model. Therefore, the least square estimator of parameters lead to a biased and inconsistent estimator, due to the correlation between the random error and endogenous variables on the right hand side of the equation.

Apart from the classic assumptions of least square we postulate that the set of variables, z, have two properties:

1. For vector z, $E(z'\epsilon) = 0$.
2. Relevance: z's are correlated with the independent variable, X.

Then $\hat{y_2}$ with Z can be written as;

$$\hat{Y_2} = (Z\hat{\beta}) = Z(Z'Z)^{-1}Z'Y_2. \tag{3.20}$$

Note that Z is a matrix. We use y_2 as a variable in X and project X with Z. That is;

$$\hat{X} = Z(Z'Z)^{-1}Z'X, \tag{3.21}$$

The IV estimation becomes;

$$\hat{\beta_{2SLS}} = (\hat{X}'X)^{-1}\hat{X}'Y. \tag{3.22}$$

This is the two stage least square estimator.

The properties of two stage least squares are: 1. Two stage least square has a consistent estimator.
2. Two stage least square is normally distributed in large samples.
3. Variance and covariance of two stage least square estimators are known.

3.3.1 Endogeneity Test

We use Durbin-Wu-Hausman statistics to determine whether the regression analysis has any autocorrelation.

The Durbin-Wu-Hausman test was proposed first by Durbin in 1954[19], then by Wu[20] in 1973 and separately by Hausman[21] in 1978. The test allows to decide for the presence of a correlation between an explanatory variable and the error term, ϵ_i. The test is performed by comparing the estimate of the ordinary least square estimation and the estimate from the IV estimation, in our case two stage least square estimation. Therefore, the test statistics becomes:

$$H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})^T[\hat{Avar}(\hat{\beta}_{IV}) - \hat{Avar}(\hat{\beta}_{OLS})]^{-1}(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \ \sim \ \chi^2_J$$

where J is the number of endogenous regressors. Note that, $Avar$ is the asymptotic variance-covariance matrix.

The hypothesis test is:

$$H_o : cov(x, \epsilon_i) = 0 \tag{3.23}$$

$$H_a : cov(x, \epsilon_i) \neq 0 \tag{3.24}$$

If the null hypothesis cannot be rejected, we conclude that the instrumental variables are consistent otherwise if the hypothesis test becomes statistically significant then the chosen instruments are not valid.

Furthermore, Hill [17] describes an alternative method, where the author uses the following regression method to explain it:

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t. \tag{3.25}$$

We want to know whether x is correlated with the error term, ϵ that is $cov(x, \epsilon) = 0$. To calculate the correlation we use z_{t1} and z_{t2} as instrumental variables for x. Hill [17] carries out a couple of steps; firstly incorporates the instrumental variables z_{t1} and z_{t2} and obtain the residuals. Secondly, uses the residuals computed in the first step as a dependent variable in the above equation. Then, estimates the regression by least square estimation and uses the t-test for the hypothesis significance. Additionally, if more than one variable is suspected of having a correlation with the error term, one must use an F-test of joint significance of the coefficients on the included residuals.

We also need to test the endogeneity that arises due to autocorrelation with the 2SLS method.

In STATA we use "estat endog" command to test endogeneity problem. The estat endog command reports us a Durbin chi-square and a Wu-Hausman F-statistics and the respective p-values with null hypothesis being;

H_o: variables are exogenous
H_a: variables are endogenous

If the p-value is less than our confidence interval we reject the null hypothesis, which means that the instrumented variable is endogenous.

3.3.2 Overidentification Test

We use "estat overid" command of STATA to test the validity of the instruments used in the model. The estat overid command reports a Sargan chi-square and a Basmann chi-square statistics and the respective p-values. Sargan and Basmann chi-square test checks whether the statistical model used over identifies restrictions, meaning if the instruments used are valid. If the null hypothesis cannot be rejected then the instruments used are valid. Moreover, if the p-value is greater than our confidence interval we conclude that the over identifying restriction is valid.

3.3.3 Significance Tests: Wald Statistics

STATA, also, gives the results of the Wald statistics for two stage least square estimation procedure. Wald test is used to test one or more restrictions in a regression model. For example, consider the OLS estimated equation

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$$

Suppose that we consider testing the restriction $\beta_1 + \beta_2 = 1$. We estimate the variance with

$$\widetilde{\upsilon}^2 = \widetilde{\sigma}^2(x^{22} + x^{33} + 2x^{23})$$

The square of a standard normal distribution is a χ^2 distribution with one degree of freedom. Thus,

$$W = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{\widetilde{\upsilon}}$$

W is the Wald statistics for the restriction. Wald test can also be utilized to test more than one restriction.

Chapter 4

Description of data and variables

For our analysis, we use the U.S pharmaceutical and medical device manufacturing companies' 2010 data. The dataset is collected from the Wharton Research Data Services (WRDS) database. In WRDS, we have exerted the "CompuStat North America- Annual Updates" to obtain the cost of goods sold, inventory, total asset and current asset amounts of the respective pharmaceutical and medical device manufacturing companies. Additionally, we have used the Food and Drug Administration's (FDA) data on medical devices and pharmaceutical drugs. The medical device manufacturers' dataset is obtained from the products and medical procedures downloadable 510(k) files. Data regarding the pharmaceutical drugs are obtained from the "drug approvals and databases" national drug code directory.

The FDA databases of medical device manufacturers and pharmaceutical drug companies' datasets contain the product varieties of each of the respective firms.

Table 4.1 describes the data obtained from the national drug code directory. Section 510(k) of FDA requires manufacturers to register their respective medical devices. The file descriptions for releasable 510(k)s are presented in table 4.2.

The WRDS's Compustat database has each firm's data that are classified via the help of Standard Industry Classification (SIC) codes that are assigned by

Definition	**Description**
Product NDC :	The labeler code and product code of the related National Drug Code number
Product Type Name :	Contains the type of the product and indicates whether the drug is a human prescription drug or human OTC drug
Proprietary Name :	Trade name of the drug
Proprietary Name Suffix :	Distinguishes the characteristic of a product such as extended release (XR) or intravenous (IV) etc.
Dosage Form Name :	The dosage form submitted by the company
Non-priorietary Name :	Generic name of the drug
Route Name :	Describes how the product is used
Starting Marketing Date :	The date the marketing of the drug begun
End Marketing Date :	The date when the drug will no longer be available
Marketing Category Name :	The marketing category chosen by the company from a list, which contains NDA/ANDA/BLA, OTC etc.
The Application Number :	Contains the number reported by the companies that have the corresponding marketing category
Labeler Name :	The name of the company that produces the drug
Substance Name :	The active ingredient of the drug
Strength Number :	The strength values of the respective drug
Strength Unit :	The units of the strength number described preciously
Pharm Classes :	The pharmaceutical class categories corresponding to substance name
DEA Schedule :	The assigned DEA schedule number reported by the company

Table 4.1: Notation and Descriptions of the National Drug Code data

Definition	Description
Applicant :	The company, which wants to approve the medical device
Contact :	The name of the respective company
Street1, street2, city, state and zip :	The open address of the respective firm
Device Name :	The name of the device registered
Date Received :	The application date of the drug
Decision Date :	The date of the decision taken by the FDA
Decision Column :	Decision taken by the FDA for the related drug
Review Advise :	Contains the advisory committees decision on the drug
Product Code :	Code given to the respective drug
State or Summary Column :	if it is labeled as summary it indicates safety and effectiveness information is available from FDA if it is labeled as statement it indicates that the information can be obtained from the applicant firm
Class Advise Committee:	The code number under which the respective product was classified
Type :	Indicates the type of 510(k) submission
Third Party Column :	Indicates whether or not the application was reviewed by a third party

Table 4.2: Notations and Descriptions of the 501 (k)data

the U.S. Department of Commerce. The U.S. Department of Commerce identifies pharmaceuticals and medical devices companies by four-digit SIC codes as: *pharmaceutical preparations* (SIC:2834), *in vitro and in vivo diagnostics substances* (SIC:2835), *biological products (no diagnostic substances)* (SIC:2836), *perfumes, cosmetics and other toilet preparations* (SIC:2844), *surgical and medical instruments and apparatus* (SIC:3841), *orthopedic, prosthetic, surgical appliances and supplies* (SIC:3842), *dental equipment and supplies* (SIC:3843), *x-ray apparatus, tubes and related irradiation apparatus* (SIC:3844) and *electromedical and electrotherapeutic apparatus* (SIC:3845).

The CompuStat database includes respective inventory, cost of goods sold (COGS), current assets, total assets and sales/turnover ratio of the firms. The CompuStat database offers ten different inventory valuation methods. The most common inventory valuation methods used by the pharmaceutical and medical device manufacturing firms are first in first out (FIFO) and last in first out (LIFO) methods. Cost of goods sold includes costs incurred during the production of the goods such as labor, raw material and overhead costs. Even though the COGS line can vary from company to company, it does not include the research and development costs for pharmaceutical companies. Current assets are the cash and other assets that are expected to be realized in short term, which consists of items realized less than 12 months or used in the production of revenue within 12 months. Sales/turnover ratio is total gross sales minus cash discounts trade discounts, and sales returns and allowances for the credit given to customers, for each operating segment. We have collected the financial data for the companies in accordance with their SICs for the fiscal year 2010.

In our research, we identify the effects of product variety on gross margin, capital intensity and inventory turnover. Therefore, we match the CompuStat data with the data obtained from FDA. We have pursued the following steps to match the data. First, we basically pair the companies with the same name. Some companies had only a single official name, which make them easier to pair. However, in the FDA data set some of the firms have multiple names. To give an example, Roche is registered under the names of Roche, Roche Diagnostic System Inc., Roche Diagnostics, Roche Diagnostics Corp., Roche Diagnostics

Corporation, Roche Professional Diagnostics etc. We combine all the Roche data in the FDA database into a single Roche and match it with Roche Holding AG in the CompuStat database. Moreover, if the parent company has any subsidiaries we group the product varieties' of the subsidiaries under the parent company. The problem we encountered was that some firms in the CompuStat database did not match one-to-one with the FDA database. Thus, we eliminated the non-matching data.

Thereafter, we remove the firms that have product variety less than five. Since, these firms probably have other goods or services that generate income, which cannot be captured solely by product variety. We omit the firms that have missing data. After applying these rules, we take a step further and identify the pharmaceutical and medical device firms by analyzing the SICs of the firms. Firms' respective SICs are obtained from the CompuStat database. Subsequently, the number of pharmaceutical companies we determine is 36 and medical device firms is 92. However, there are firms that are included in both the pharmaceutical SIC and medical device SICs in the CompuStat database. We have separated these firms into pharmaceutical or medical device via examining the products on their web sites. We have followed these procedures in order to increase the sample size of our study. Furthermore, we generate another dataset that combines pharmaceutical and medical device firms and contains 128 firms. From this dataset we have omitted 2 outliers, which we think will effect the assumptions of regression. We have identified outliers using STATA command "lvr2plot", which plotted the outliers of the data set. By removing 2 firms we are left with 126 firms that is 36 pharmaceutical firms and 90 medical device firms. This data set, also, has only the firms with product variety greater than or equal to five. We use the combined data as the master data set in our study. Tables 4.3, 4.4, 4.5 and 4.6 describe the firms we have used and their respective SIC.

Company Type	**SIC Codes**	**Firms**
Pharmaceutical Drug Firms (PDF)	2834	Teva Pharmaceutical Industries Ltd Mylan Inc Glaxosmithkline PLC Pfizer Inc Watson Pharmaceuticals Inc. Novartis AG Perrigo Company Par Pharmaceutical Companies Inc. Merck and Co Inc. Astrazeneca PLC Eli Lilly and Company Lannett Company Inc. Allergan Inc. Bristol-Myers Squibb Company Taro Pharmaceutical Industries Ltd Warner Chilcott Plc Shire PLC ProPhase Labs Inc Cephalon Inc Endo Pharmaceuticals Holdings Inc Akorn Inc. Salix Pharmaceuticals Ltd Medicis Pharmaceutical Corp. Cornerstone Therapeutics Inc. Celgene Corp. Columbia Laboratories Inc. United Therapeutics Corp. Amylin Pharmaceuticals Inc.
PDF	2836	Baxter International Inc. Amgen Inc. Gilead Sciences Inc.
PDF	2844	Avon Products Inc. Colgate-Palmolive Company L'Oreal S.A., Paris CCA Industries Inc. Estee Lauder Companies Inc.

Table 4.3: SIC Code Classification of Pharmaceutical Firms

Company Type	**SIC Codes**	**Firms**
Medical Device Firms (MDF)	2834	Roche Holding AG Abbott Laboratories Johnson and Johnson Elan Corp PLC Hospira Inc. Derma Sciences Inc. Rockwell Medical Technologies Inc.
MDF	2835	American Bio Medica Corp. OraSure Technologies Inc. Hemagen Diagnostics Inc. Quidel Corp. Corgenix Medical Corp. Gen-Probe Inc. Meridian Bioscience Inc. Chembio Diagnostics Inc. SeraCare Life Sciences Inc. Trinity Biotech PLC
MDF	2836	Kinetic Concepts Inc.
MDF	3841	Boston Scientific Corp. Becton, Dickinson and Company NuVasive Inc. Tornier NV Merit Medical Systems Inc. ICU Medical Inc. Teleflex Inc. Bard (C.R.) Inc. Orthovita Inc. AtriCure Inc. LeMaitre Vascular Inc. Mako Surgical Corp. CareFusion Corp. Rochester Medical Corp. Cardica Inc. Orthofix International NV Atrion Corp. Synergetics USA Inc. Cardiovascular Systems Inc. Retractable Technologies Inc.

Table 4.4: SIC Code Classification of Medical Device Firms

Company Type	**SIC Codes**	**Firms**
MDF	3842	Smith and Nephew PLC Zimmer Holdings Inc. Wright Medical Group Inc. Exactech Inc. Vascular Solutions Inc. Edwards Lifesciences Corp. Invacare Corp. Stryker Corp. Integra LifeSciences Holdings Corp. Alphatec Holdings Inc. Kensey Nash Corp. Allied healthcare Products Inc. Regeneration Technologies Inc. Span-America Medical Systems Inc.
MDF	3843	Dentsply International Inc. Sirona Dental Systems Inc. Biolase Technology Inc.
MDF	3844	Hologic Inc.

Table 4.5: SIC Code Classification of Medical Device Firms Continued

Company Type	**SIC Codes**	**Firms**
MDF	3845	ArthroCare Corp. St.Jude Medical Inc. Resmed Inc. Varian Medical Systems Inc. Cynosure Inc. Masimo Corp. Conmed Corp. Zoll Medical Corp. Covidien Plc Mindray Medical International Ltd. Intuitive Surgical Inc. NxStage Medical Inc. Syneron Medical Ltd. Given Imaging Ltd. Palomar Medical Technologies Inc. SonoSite Inc. Spectranetics Corp. Volcano Corp. Cambridge Heart Inc. Stereotaxis Inc. Accuray Inc. CAS Medical Systems Inc. Trimedyne Inc. IRIDEX Corp. Utah Medical Products Inc. Digirad Corp. Natus Medical Inc. Cutera Inc. Dynatronics Corp. Bovie Medical Corp. Paradigm Medical Industries Inc. Hansen Medical Inc. Electromed Inc. Fonar Corp.

Table 4.6: SIC Code Classification of Medical Device Firms Continued

We combine the dataset of pharmaceutical drug companies and medical device companies into a single dataset, which we used in our analysis. Hence, our final test sample consists of 126 observations taken from year 2010. Then, we generat hypothesis for test samples. In order to understand the hypotheses, we first describe the setup and the notations of the problem and how we used the data obtained from WRDS Compustat and FDA database in our calculations. Data description is available Table 4.7.

Notation	**Description**
IL_i	Inventory amount of firm i
$COGS_i$	Cost of goods sold of firm i
PV_i	Product Variety of firm i
ST_i	Sales turnover ratio of firm i
A_i	Total Assets of firm i
CA_i	Current Assets of firm i
S_i	Sales, net of markdowns in dollars for firm i ($ million)
GFA_i	Gross fixed asset of firm i
T_i	Type of firm i
	1: Pharmaceutical Drug Firm
	0: Medical Device Firm

Table 4.7: Notations and Descriptions

To compare the effects of the variables we have used performance variables described below:

Inventory Turnover ratio shows how many times a company's inventory is sold and replaced and is calculated by dividing inventory to cost of goods sold.

$$IT_i = \frac{COGS_i}{IL_i}.$$

Gross Margin is the ratio of company's total sales revenue minus cost of goods sold, divided by the total sales revenue and it is expressed as a percentage. The higher this percentage, the more company retains on each dollar of sales and is represented as;

$$GM_i = \frac{S_i - COGS_i}{S_i}.$$

Capital Intensity is the ratio of gross fixed assets to the total assets

$$CI_{\ i} = \frac{GFA}{IL_i + GFA_i}.$$

Gross Fixed Assets is total assets minus the current assets.

$$GFA_i = A_i - CA_i.$$

Table 4.8 describes the statistics of our dataset. We have analyzed a total of 126 firms. These firms include 36 pharmaceutical drug firms and 90 medical device firms. Average inventory turnover of 126 firms is 2.78 with a standard deviation of 2.07 and a median of 2.42. Average gross margin is approximately 0.63 having a standard deviation of 0.17 and a median of 0.64. Average capital intensity is 0.74 with a standard deviation of 0.18 and a median of 0.79. Average product variety is 70.15, has a deviation of 120.01 and a median of 26. Note that, the standard deviations of each variable for each type of firm are presented in brackets.

Description	Pharmaceutical Drug Firms	Medical Device Firms	Combined
SIC Codes	2834, 2836 2844	2834, 2835, 2836 3841, 3842, 3843 3844, 3845	
Number of Firms	36	90	126
Average Inventory Turnover	2.5366	2.8751	2.7784
Standard Deviation	(1.3947)	(2.2855)	(2.0706)
Maximum	8.3488	16.4467	16.4467
Minimum	0.7404	0.5228	0.5228
Median	2.3157	2.5106	2.4153
Average Gross Margin	0.7134	0.5909	0.6259
Standard Deviation	(0.1619)	(0.1571)	(0.1682)
Maximum	0.9307	0.8824	0.9308
Minimum	0.3672	0.1889	0.1889
Median	0.7597	0.6072	0.6374
Average Capital Intensity	0.8378	0.7039	0.7422
Standard Deviation	(0.1519)	(0.1815)	(0.1833)
Maximum	0.9740	0.9657	0.9740
Minimum	0.2795	0.2362	0.2362
Median	0.8752	0.7585	0.7926
Product Variety	119.5833	50.3667	70.1429
Standard Deviation	(177.6148)	(80.2396)	(120.0120)
Maximum	721	402	721
Minimum	5	5	5
Median	49	22.0	26.0

Table 4.8: Summary Statistics of the Dataset

Chapter 5

Hypothesis

We have developed hypotheses to test the relations among inventory turnover, gross margin, capital intensity and product variety using the combined data of 126 pharmaceutical and medical device manufacturing firms in their fiscal year 2010. Gaur et. al [4] have previously studied the correlation among inventory turnover, gross margin, capital intensity and sales surprise for the period 1985-2000. We test some of the hypothesis that are proposed by Gaur et al. [4] for our dataset. Furthermore, we have developed hypothesis to understand the effect of product variety on inventory turnover and inventory level.

5.1 Gross Margin

Hypothesis 1 *Inventory turnover is negatively correlated with gross margin*

To explain the hypothesis let us consider the classical newsboy model. Newsboy model allows us to determine the order quantity to satisfy demand during a short period with stochastic demand and no replenishment opportunities. In the classical newsboy setting, it can be proved that as gross margin increases, inventory level increases, which implies a decrease in the inventory turnover. Thus, an increase in the gross margin decreases the inventory turnover. Additionally, price can be a factor affecting gross margin. As price of the products increase the gross margin will increase. However, simply by taking into account the supply demand

curve, the demand will fall due to increase in price; thus, decreasing inventory turnover.

5.2 Capital Intensity

Hypothesis 2 *Inventory turnover is positively correlated with capital intensity*

Capital intensity is the amount of capital in relation to the factors of production. Capital investments in information technologies, warehouse, supply chain systems might improve the inventory management systems, hence, increasing the inventory turnover. To prove this assertion, let us consider a depot-warehouse system with independently identically distributed demand at the warehouses as stated in the model by Eppen and Schrage [22]. Inserting a depot before the warehouses will decrease the amount of inventory held at the warehouses [4]. This effect is called the joint ordering effect [23]. As the average inventory holding decreases, this will in turn increase the inventory turnover. Additionally, any capital investment on information systems will allow better management of inventory by decreasing lead times, effectively forecasting customer demand and reducing order batch sizes.

5.3 Product Variety

Hypothesis 3 *Inventory turnover is negatively correlated with product variety*

Given that the total demand does not change, an increase in product variety increases coefficient of variation, hence the inventory level. Coefficient of variation allows us to compare two or more different magnitudes of variation and is represented by the ratio of standard deviation divided by its mean, $(\frac{\sigma}{\mu})$. Therefore, the increase in inventory level, in return decreases the inventory turnover ratio.

We can also explain the validity of the hypothesis using the concept risk pooling effect. Eppen [23] indicates that risk pooling allows firms to manage uncertainty, while driving costs down and increasing the profit level. Risk pooling,

in products, is to decrease the number of variety in a product line. Lower product variety through delayed differentiation helps increase inventory turnover. Moreover, lower variety will allow firms to forecast demand accurately. Kekre and Srinivasan [9] point out that an increase in variety will increase inventory level. However, the authors also state that, if the company can take advantage of increasing product variety, while achieving a higher market share, total inventory level will decrease. While the market share expansion may be the case for consumer goods and industrial goods, we do not expect an increase in market share for the health care industry. So, an increase in product variety will only increase inventory level. By the formulae of inventory turnover, an increase in inventory level would definetly bring forth a decrease inventory turnover. Therefore, our hypothesis will hold.

Also, fewer number of product variety indicates less variability, lower $\frac{\sigma}{\mu}$, since the aggregation effect becomes more significant, which in return allows for better forecasts. As the number of products increase it gets more difficult to forecast demand.

We use graphs, in order to have an initial understanding of the impact of product variety on inventory turnover. First, we plot the histogram of logarithm of inventory turnover and logarithm of product variety for medical device companies in graphs 5.1 and 5.2. The heights of the bars drawn over the $\ln(IT)$ and $\ln(PV)$ invervals indicate the frequency of the respective intervals. Notice that the intervals have equal width, since the area of the bar is proportional to its relative frequency. Figure 5.1 describes the frequency of $\ln(IT)$ variable for 90 medical device firms. 1 is the most observed logarithm of iventory turnover value for medical device companies. The latter histogram 5.2 shows the frequency of logarithm of product variety for medical device companies. We observe that the logarithm of product variety values tend to pile up between 2 and 3.

We plot similar histograms for 36 pharmaceutical companies. Histogram figure 5.3 plots the occurance rate of $\ln(IT)$ for pharmaceutical companies. According to the graph, $\ln(IT)$ value around 1 has the highest occurance rate. Histogram figure 5.4 shows the frequency distribution of $\ln(PV)$ for for pharmaceutical companies.

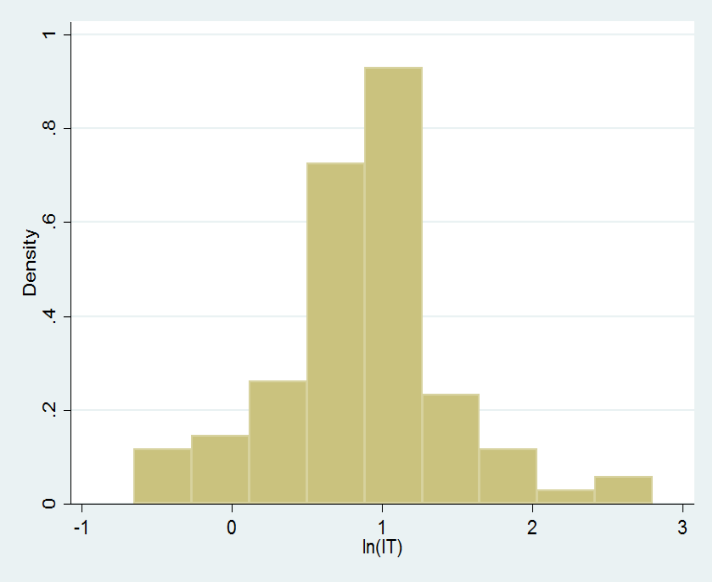

Figure 5.1: Histogram of ln(IT) medical device companies

From this figure we observe that $\ln(PV)$ values accumulate between 3 and 4.

We plot a scattered graph to show the relationship between IT and PV for medical device companies, which is shown in figure 5.5. We also provide a fitted curve. STATA estimates a single relationship between IT and PV for all observations and then plots the fitted values of IT and PV. In order to fit a line in the scatter graph, we use the "lfitci" command of STATA [24], which draws a fitted regression line for IT and PV values. Furthermore, we provide confidence intervals to forecast the actual value of $(IT|PV)$ by using the "stdf" command in STATA in figure 5.5. Even though there are some outliers in the dataset, the fitted line in figure 5.5 shows that there is a vague negative relationship between IT and PV. Using the similar graph plotting method as in the previous figure we have draw the scattered plot graph, fitted the values of $\ln(IT)$ and $\ln(PV)$ and constructed the confidence bands for predicting the actual values of $\ln(IT)$ and $\ln(PV)$ in figure 5.6. We observe that as the logarithm of PV increases the logarithm of IT tends to decrease. Additionally, note that the widths of the confidence interval is large since the values gets further away from the mean of our dataset. Although the relationship between inventory turnover and product variety is obscure in the graphs, both figures 5.5 and 5.6 weakly support hypothesis

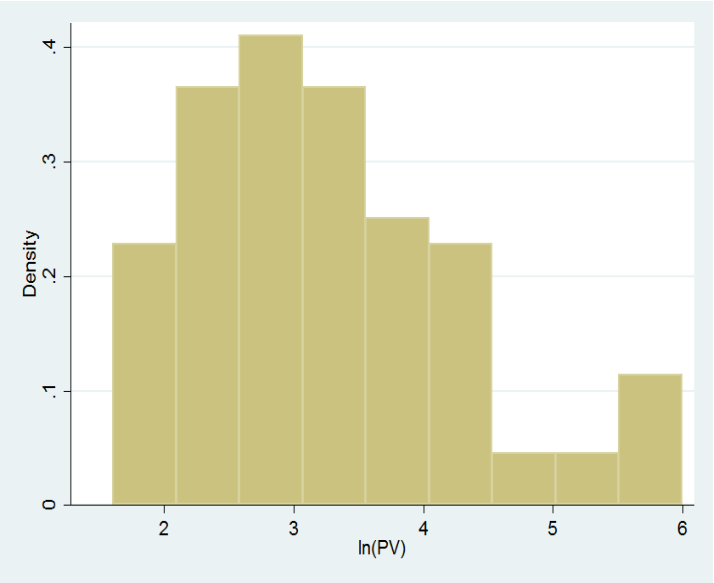

Figure 5.2: Histogram of ln(PV) for medical device companies

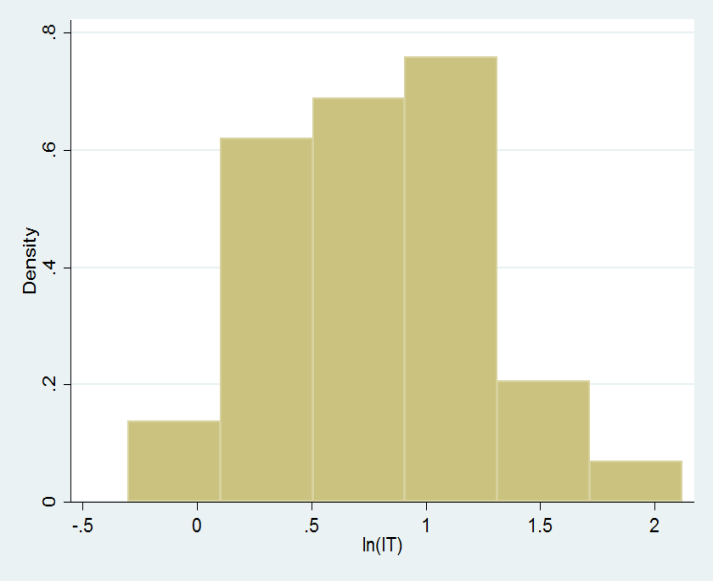

Figure 5.3: Histogram of ln(IT) for pharmaceutical companies

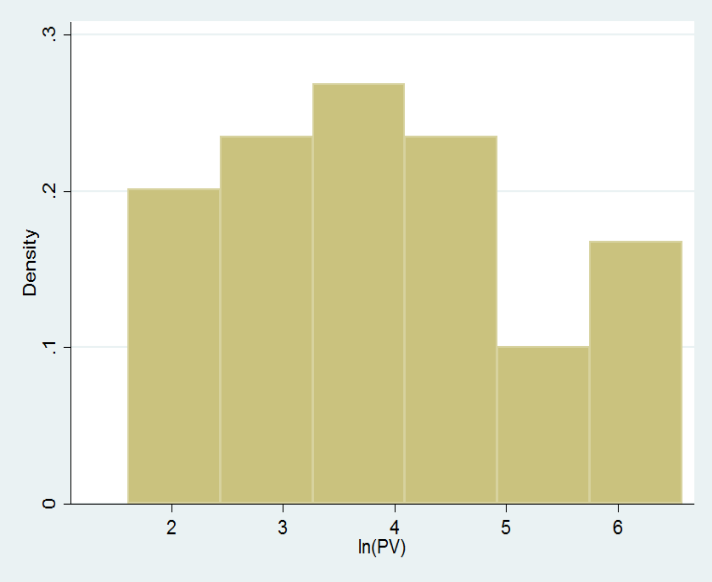

Figure 5.4: Histogram of ln(PV) for pharmaceutical companies

3.

We have drawn scattered plot graphs of IT versus PV for pharmeceutical companies in figure 5.7. When we analyze the scatter plots, we discover an unclear relationship between IT and PV but the fitted curve shows a negative relationship between the two variables. Moreover, over the range of the values, the width of the interval changes as we go further from the mean value of PV. This figure weakly supports hypothesis 3. Another graph we have plotted for pharmaceutical companies is the scattered plot graph of $\ln(IT)$ and $\ln(PV)$, which is displayed in figure 5.8. There does not exist a clear pattern in the fitted regression line for $\ln(IT)$ and $\ln(PV)$ for pharmaceutical firms, hence, no correlation between them. In order to understand the correlation between product variety and inventory level we have developed hypothesis 4. Another scattered plot figure we have plotted for medical device firms is figure 5.9. On the y-axis we have $\ln(IT)$ minus the estimate of $\ln(IT)$ and on the x-axis we have $\ln(PV)$. We would have expected that in cases where $\ln(IT)$ minus estimate of $\ln(IT)$ are positive PV has to be higher and in cases where it is negative, PV has to be low. However, this clearly is not the case. We have constructed a similar scattered graph for pharmaceutical companies displayed in figure 5.10. The fitted

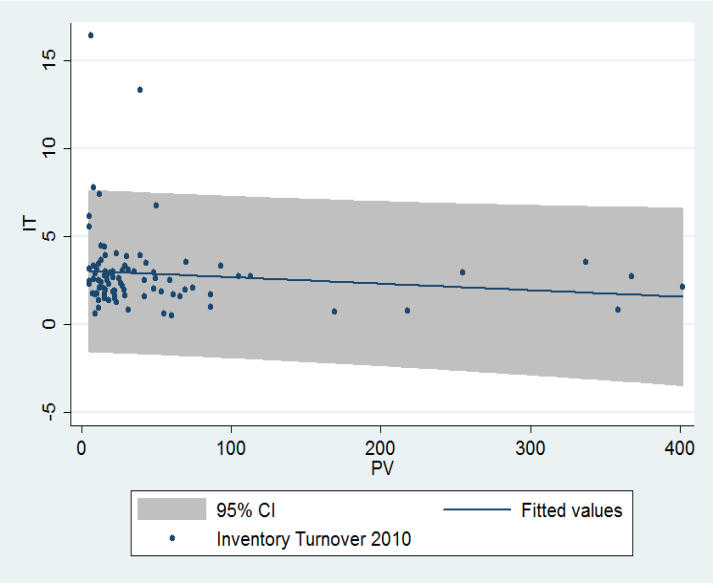

Figure 5.5: IT versus PV for medical device companies

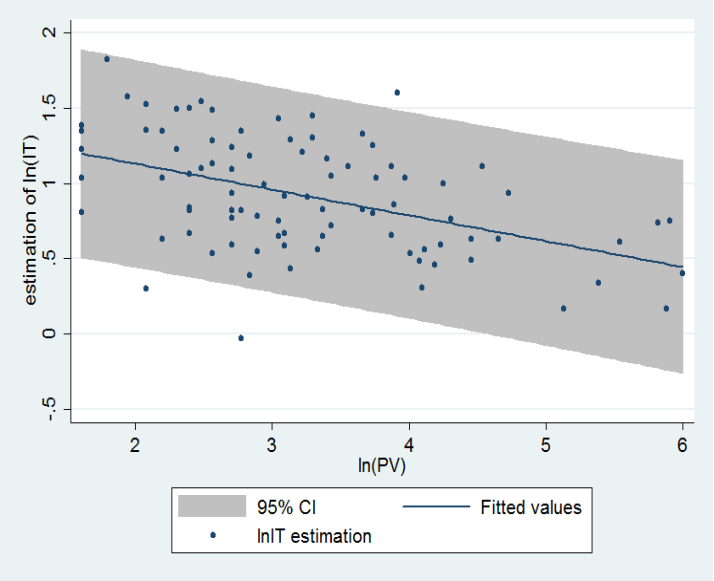

Figure 5.6: ln(IT) versus ln(PV) for medical device companies

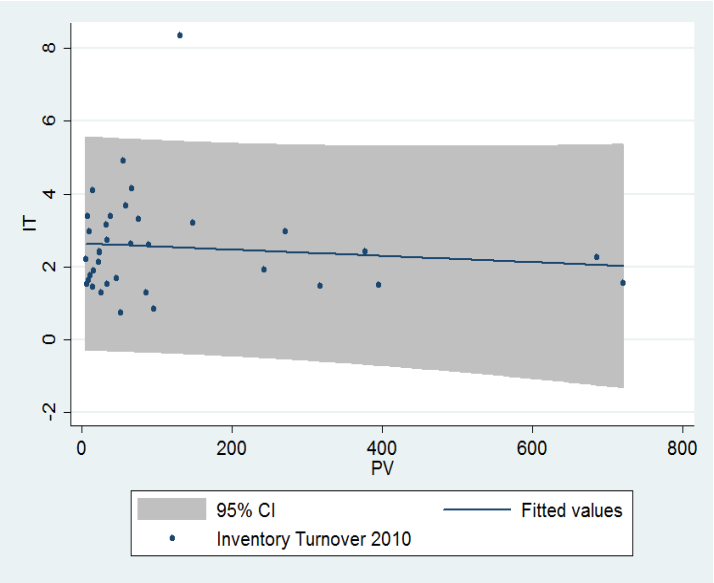

Figure 5.7: IT versus PV for pharmaceutical companies

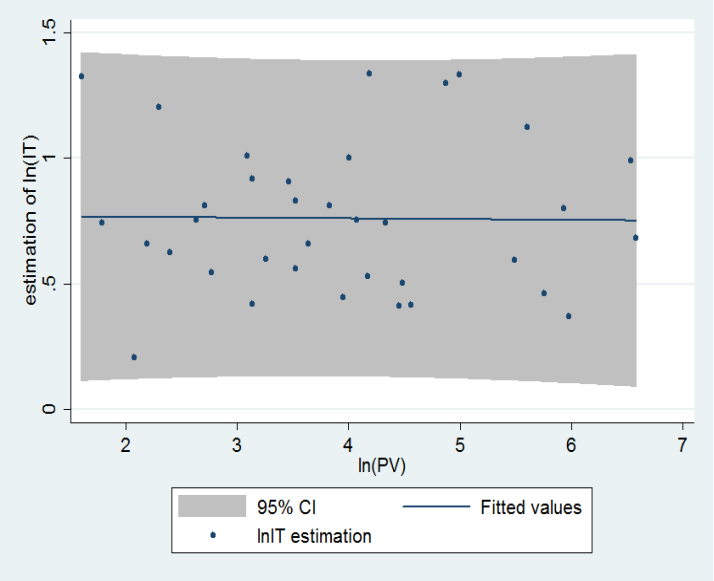

Figure 5.8: ln(IT) versus ln(PV) for pharmaceutical companies

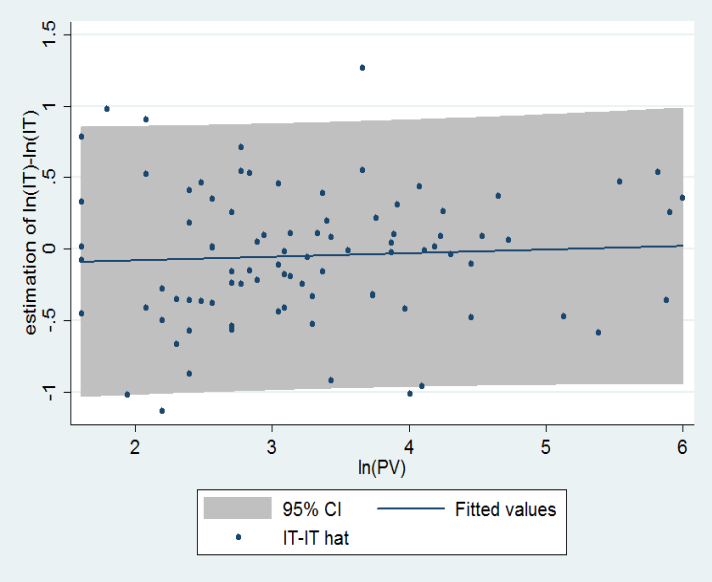

Figure 5.9: Difference between ln(IT) and the estimate of ln(IT) vs ln(PV) for medical device companies

regression line plotted using "lfitci" command of STATA, shows that there exist no correlation between these two variables. It may be because product variety influences inventory level and cost of goods sold simultaneously, the relationship between product variety and inventory cannot be sought clearly using graphical methods.

Hypothesis 4 *Inventory level is positively correlated with product variety*

Assuming that the demand is constant, an increase in product variety will divide the same amount of demand to a larger number of products increasing the variability, which in return increases the inventory level. Additionally, increase in the variety results in an increase in the number of different inventory items kept [25]. The relationship between inventory levels and product variety has been studied by Ton and Raman [10]. The authors indicate that there is a trade-off between product variety and the optimal level of inventory held at a specific store, since variety would increase the inventory levels of the related store. A paper by van Ryzin and Mahajan [11] argues that there is an implicit cost related with having a large amount of product variety and this cost is mostly derived from the level of inventory kept at the retailer's stock. Kekre and Srinivasan [9] indicate

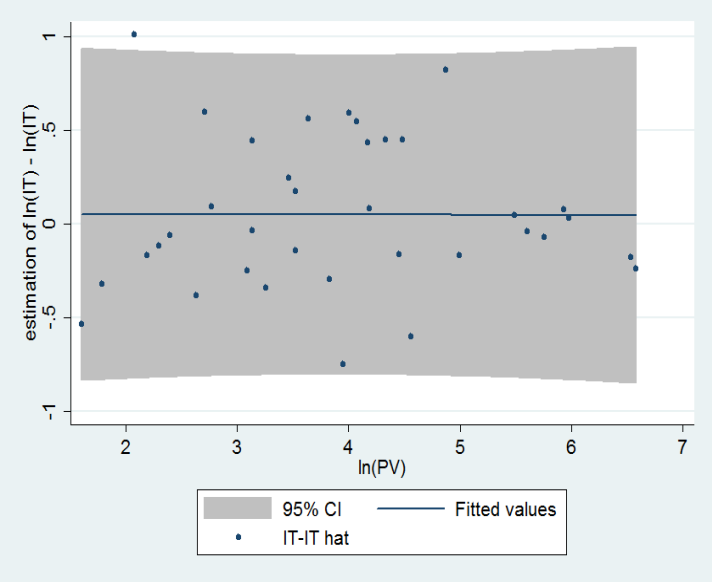

Figure 5.10: Difference between ln(IT) and the estimate of ln(IT) vs ln(PV) for pharmaceutical companies

that as product lines get broader a variety of finished goods inventory and safety stock will be needed, which in return increase the inventory level. But the authors note that having a broader product line will increase market share therefore, reduce total level of inventory. We mainly state that, in health care industry, broader product line would lead to an increase in cost of goods sold, which has an indirect positive effect on inventory level. However, note that, product variety, also, has a direct increasing effect on inventory level. This direct impact is due to failure of taking advantage of risk pooling effect as product variety increases. Furthermore, an increase in product variety will increase setup times, which in return would increase total inventory level. We argue that an increase in cost of goods sold due to an increase in product variety will be relatively lower than the effect of an increase in product variety on inventory level. Leading to a conclusion that inventory level is positively correlated with product variety.

Chapter 6

Model Specifications and Analysis

6.1 Regression Models

We use the following multiple linear regression and semi-log models to test hypothesis 1 through 3.

Model 1

We begin by analyzing the following model;

$$IT_i = \alpha_1 + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 PV_i + \alpha_5 T_i + \epsilon_i \tag{6.1}$$

In the above equation (6.1), IT_i is the dependent variable representing inventory turnover of firm i, GM_i is the gross margin of firm i, CI_i is the capital intensity of firm i, PV_i is the product variety of firm i all calculated using the formulas presented in Chapter 4. T indicates the type of the firm. It takes the value of 1 if the firm is a pharmaceutical firm and 0 if the firm is a medical device manufacturer. We initially assume that inventory turnover is linearly related to gross margin, capital intensity, type of the firm and product variety of firm i. In this multiple regression model, α_1 is the value of the dependent variable when each of the independent, explanatory variables takes the value zero and the remaining

items α_2, α_3,α_5 and α_4 describe the dependence of inventory turnover on gross margin (GM), capital intensity (CI) and product variety (PV), respectively. The sign of α's indicate the relationship between inventory turnover and the related exogenous variable.

The results of equation 6.1 are given in Tables 6.1 and 6.2:

Number of observations	126
F(4,121)	13.39
Prob > F	0.0000
R-Squared	0.3068
Adjusted R-Squared	0.2839
Root MSE	1.7522

Table 6.1: Results of multiple regression model 1

IT	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
GM	-7.654113	1.084114	-7.06	0.000	[-9.800403, -5.507824]
CI	3.838169	1.015568	3.78	0.000	[1.827584, 5.848754]
PV	-0.0025566	0.0013834	-1.85	0.067	[-0.0052955, 0.0001823]
T	0.2620416	0.3816994	0.69	0.494	[-0.493633, 1.017716]
cons.	4.825068	0.752984	6.41	0.000	[3.334338, 6.315799]

Table 6.2: Results of multiple regression model 1

The explanatory variables can only explain about 30.68% of the variation in the inventory turnover. More importantly, PV is not a significant variable, since its p-value is greater than 0.05. Furthermore, T is, also, not a significant variable.

However, before jumping into any conclusions we have to test the model for the assumptions of multiple regressions we have stated in Chapter 3. First, we use the "vif" command, which stands for variance inflation factor, of STATA to test for colinearity of the model. VIF measures how much the variance of the coefficient of estimate is inflated by multicolinearity. VIF values greater than 10 may pose some problems, since the variable is considered as a linear combination of other independent variables. $1/VIF$ is the tolerance level and is used to check for the degree of colinearity. The results of the multicolinearity test is displayed in Table 6.3. According to the results for model 1, the model passes the multicolinearity

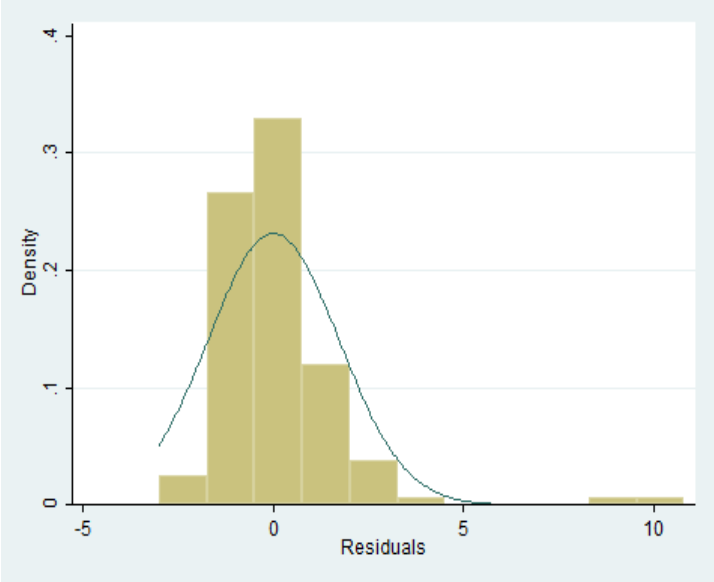

Figure 6.1: Histogram of residuals for model 1

test. Note that, even when a model shows extreme multicolinearity the estimates are still the best linear unbiased estimators (BLUE).

Variable	VIF	1/VIF
CI	1.41	0.708632
GM	1.35	0.738490
T	1.22	0.819546
PV	1.12	0.891061
Mean VIF	**1.28**	

Table 6.3: Test for multicolinearity for model 1

We use histogram to asses the residuals visually using Figure 6.1. We use STATA command "histogram res, normal" to draw a histogram and superimpose a normal line on it. The graph shows that the errors are not normally distributed.

We further use the "swilk res" command of STATA to numerically test for normality of the error terms. The null hypothesis is; data is normally distributed. We reject the null hypothesis and conclude that the residuals are not normally distributed. The results are given in Table 6.4.

To test for homoscedasticity we plot residuals against the fitted values. If the model is well-fitted there should not exist any pattern between residuals and the

Variable	Obs	W	V	z	Prob > z
Residuals	126	0.73308	26.768	7.384	0.00000

Table 6.4: Shapiro-Wilk W test for normal data for model 1

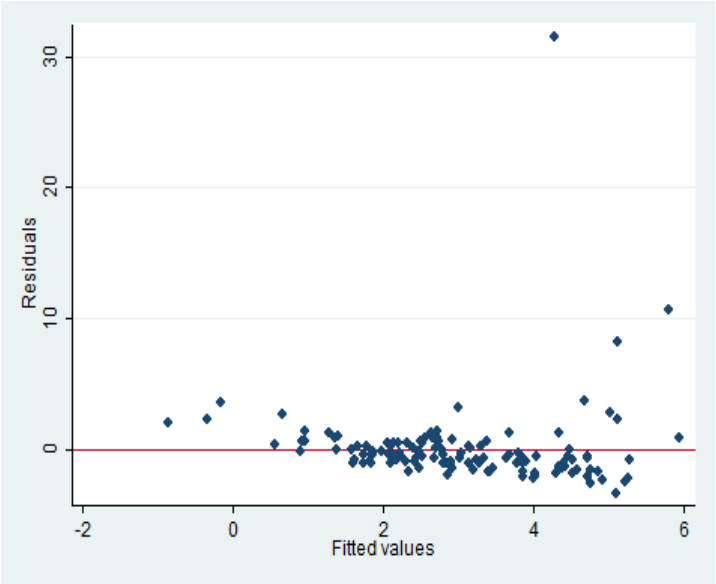

Figure 6.2: Residuals against fitted values for model 1

predicted values. The figure is presented in 6.2. Note that we put a reference line at $y = 0$. We observe that pattern of the data point is getting wider toward the end, which is a sign of heteroscedasticity. We, also, test for heteroskedasticity in the model using "hettest" command of STATA. The test gives Breusch-Pagan/Cook-Weisberg test statistics for the model. The observed p-value is 0 therefore, we reject the null hypothesis that the variables are homoscedastic. The detailed results of the test is provided in Table 6.5.

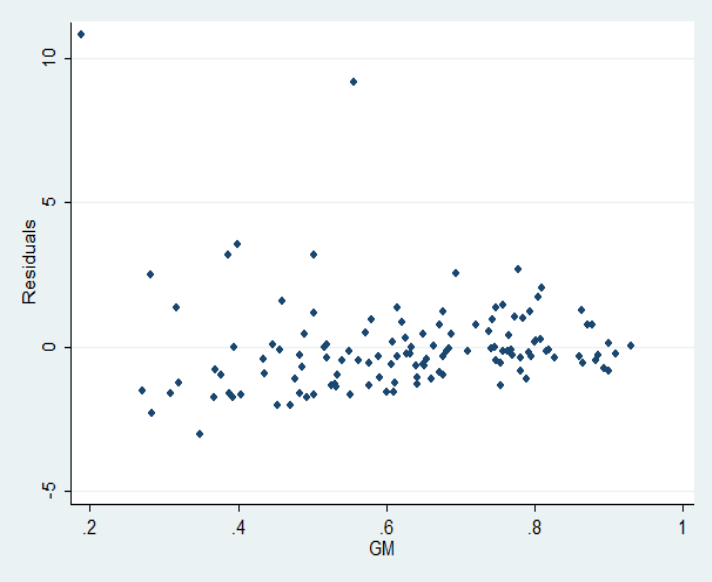

Figure 6.3: Scatter plot graphs of Residuals vs GM for model 1

$\chi^2(1)$	105.21
Prob $>\chi^2$	0.0000

Table 6.5: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for model 1

Another assumption we test is the linearity assumption of the model. In order to test this assumptions we draw scatter graphs, which plot residuals against the independent variables. These graphs are shown in Figures 6.5, 6.3 and 6.4. The figures do not indicate a strong departure from linearity. We conclude that the model 1 cannot pass the multiple regression assumptions. Hence, cannot be used as a proper model.

Model 2

Another multiple regression model we test is the following linear model;

$$IT_i = \alpha_1 + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 \ln(PV_i) + \alpha_5 T_i + \epsilon_i \tag{6.2}$$

We give the multiple regression results of equation (6.2) in Tables 6.6 and 6.7.

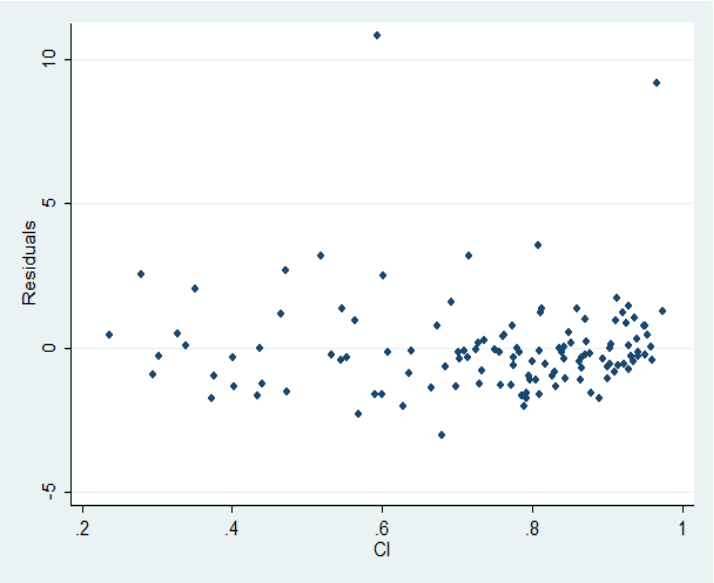

Figure 6.4: Scatter plot graphs of Residuals vs CI for model 1

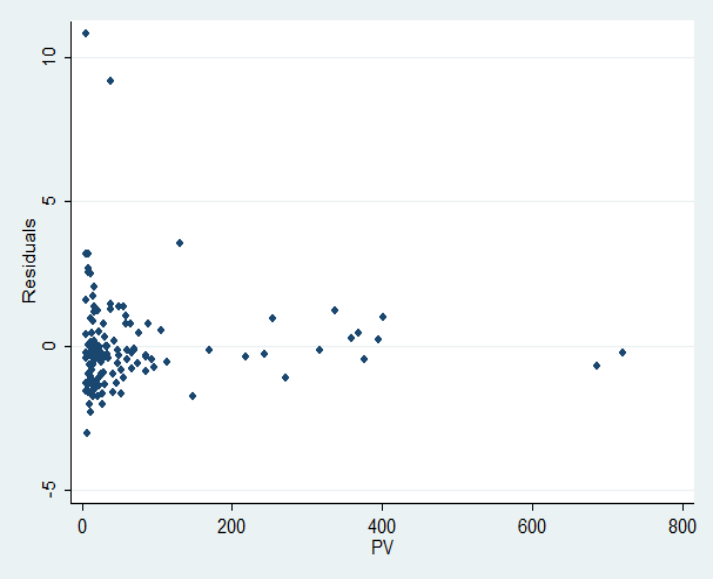

Figure 6.5: Scatter plot graphs of Residuals vs PV for model 1

Number of observations	126
F(4,121)	13.42
Prob > F	0.0000
R-Squared	0.3073
Adjusted R-Squared	0.2844
Root MSE	1.7515

Table 6.6: Results of multiple regression model 2

IT	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
GM	-7.441569	1.084294	-6.86	0.000	[-9.588214, -5.294923]
CI	3.973327	1.031491	3.85	0.000	[1.931218, 6.015437]
ln(PV)	-0.2740662	0.1461138	-1.88	0.063	[-0.563337, 0.0152045]
T	0.2210755	0.3777569	0.59	0.559	[-0.5267938, 0.9689449]
cons.	5.369552	0.790067	6.80	0.000	[3.805406, 6.933699]

Table 6.7: Results of multiple regression model 2

30.73% of the variation in inventory turnover can be explained by the explanatory variables in the model. Additionally, by analyzing our results we observe that inventory turnover is negatively correlated with gross margin just as we stated in Hypothesis 1. Moreover, capital intensity is positively correlated with inventory turnover, just like in Hypothesis 2. Type of the firm and product variety is negatively correlated wih inventory turnover. However, since the p-value of $\ln(PV)$ and T is greater than 0.05 they are not significant.

We test the regression assumptions for model 2. We start by testing multicolinearity shown in Table 6.8. Judging by the results, there does not exist a variable that is a linear combination of other variables and the model passes the multicolinearity test.

Variable	VIF	1/VIF
CI	1.46	0.686353
GM	1.36	0.737633
T	1.20	0.836048
ln(PV)	1.19	0.839806
Mean VIF	**1.30**	

Table 6.8: Test for multicolinearity for model 2

To test the normality assumptions we use various methods. First, we draw a

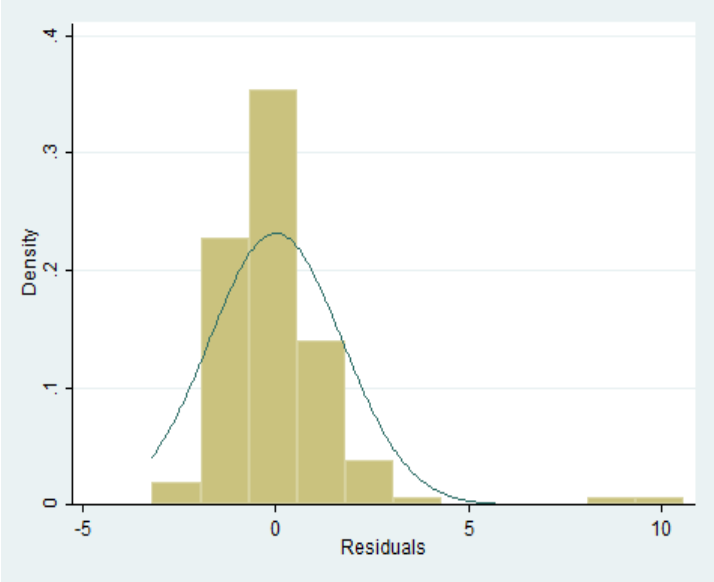

Figure 6.6: Histogram of residuals for model 2

histogram displayed in Figure 6.6. The normal line on the histogram is skewed and indicates non-normality. In order to statistically understand the normality assumption, we use Shapiro-Wilk W test. Again, we observe that the p-value is below 0.05 level. Thus, we reject the null hypothesis that the residual terms are normally distributed. The results are presented in Table 6.9.

Variable	Obs	W	V	z	Prob > z
Residuals	126	0.74148	25.925	7.312	0.00000

Table 6.9: Shapiro-Wilk W test for normal data for model 2

We plot residuals against fitted values to check for heteroscedasticity in Figure 6.7. As in model 1, by analyzing the graph's pattern, we observe that toward the end of the figure the pattern is widening. This is a sign of heteroscedasticity. We further use the "hettest" command to make sure that the residuals are heteroscedastic. The results are displayed in Table 6.10. Since the p-value equals to 0, we reject the null hypothesis that the residuals are homoscedastic.

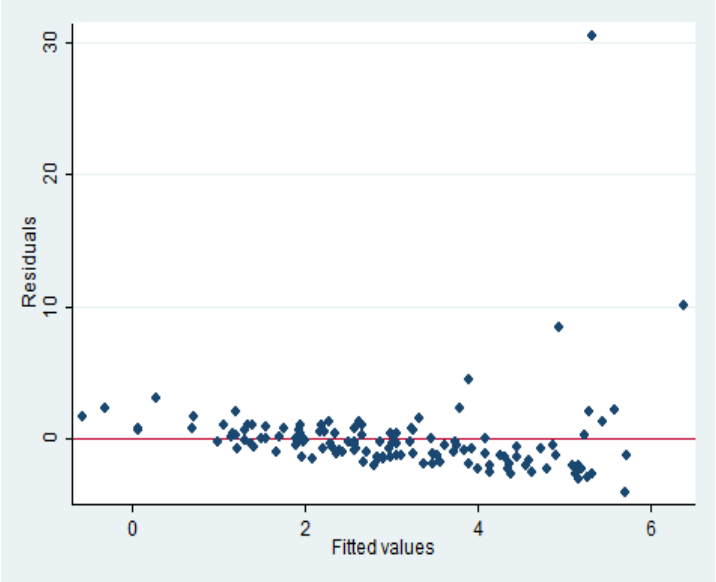

Figure 6.7: Residuals against fitted values for model 2

$\chi^2(1)$	112.65
Prob $>\chi^2$	0.0000

Table 6.10: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for model 2

Analyzing the linearity assumption is not so straightforward in multiple regression models. We use scatter plots of GM, CI and $\ln(PV)$ to analyze the linearity assumption. The Figures are presented in 6.8, 6.9 and 6.10. The three residuals versus the independent variables do not indicate a strong departure from linearity even though there are some outliers.

Model 2 fails the normality test and hence is not a valid multiple regression model. The failure of normality assumption may be due to fluctuations in inventory turnover data. We take the logarithm of inventory turnover to eliminate these fluctuations in the rest of the models.

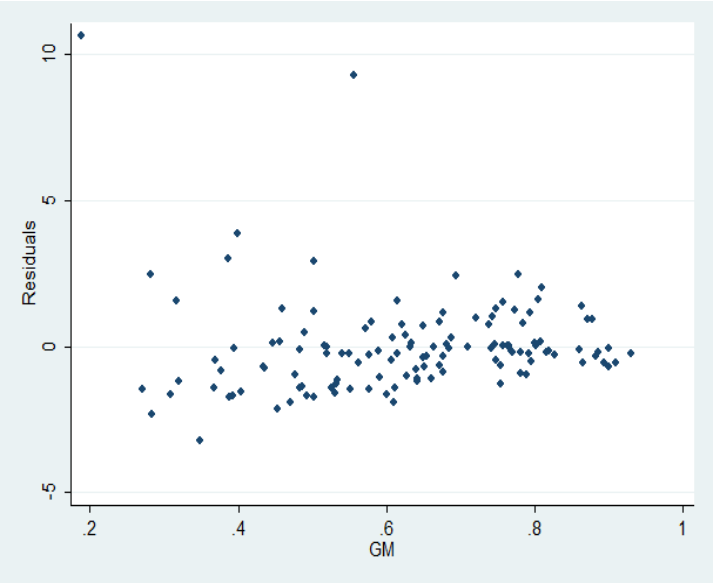

Figure 6.8: Scatter graph residual vs GM for model 2

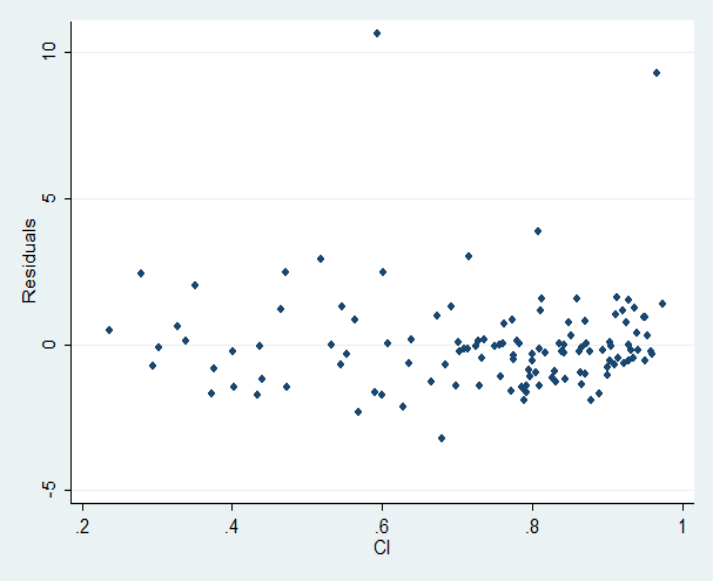

Figure 6.9: Scatter graph residual vs CI for model 2

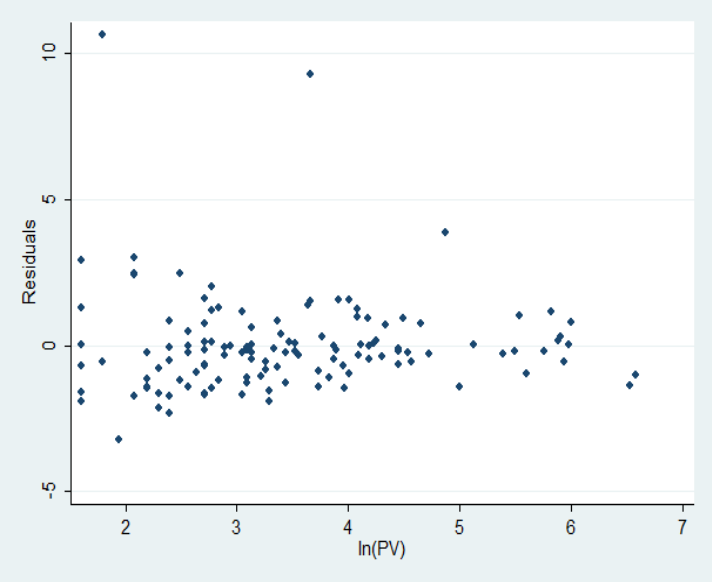

Figure 6.10: Scatter graph residual vs ln(PV) for model 2

Model 3

Next, we analyze the following multiple regression model in which we have taken the logarithm of inventory turnover;

$$\ln(IT_i) = \alpha_1 + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 \ln(PV_i) + \alpha_5 T_i + \epsilon_i \tag{6.3}$$

The results of the multiple regression model are given in tables 6.11 and 6.12.

Number of observations	126
F(4,121)	19.58
Prob > F	0.0000
R-Squared	0.3929
Adjusted R-Squared	0.3729
Root MSE	0.44945

Table 6.11: Results of multiple regression model 3

Again we observe that the gross margin is negatively correlated with inventory turnover. Furthermore, capital intensity is positively correlated with inventory

ln(IT)	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
GM	-2.307173	0.2782342	-8.29	0.000	[-2.858011, -1.756335]
CI	1.241072	0.2646848	4.69	0.000	[0.7170587, 1.765086]
ln(PV)	-0.0914869	0.0374934	-2.44	0.016	[-0.1657149, -0.0172588]
T	0.1219985	0.0969339	1.26	0.211	[-0.0699078, 0.3139048]
cons.	1.653766	0.2027344	8.16	0.000	[1.2524, 2.055132]

Table 6.12: Results of multiple regression model 3

turnover. The reasons behind it were explained fully in hypothesis 1 and hypothesis 2. We have used a log specification for product variety and observe that the logarithmic variable product variety is negatively correlated with inventory turnover. We find out that the log-linear model can explain 39.29% of the total variation in inventory turnover.

We need to test the regression assumptions for model 3. We start by checking for multicolinearity, displayed in Table 6.13. The data indicates that there is no multicolinearity because the variance of inflation factors are smaller than 10. Thus, there is no strong correlation between predictor variables.

Variable	VIF	1/VIF
CI	1.46	0.686353
GM	1.36	0.737633
T	1.20	0.836048
ln(PV)	1.19	0.839806
Mean VIF	**1.30**	

Table 6.13: Test for multicolinearity for model 3

We continue checking the assumptions by analyzing whether the error terms are normally distributed. The Histogram is shown in Figure 6.11. The histogram shows that the error terms are normally distributed. Additionally, we use the Shapiro-Wilk W test to confirm that the normality assumptions stands. Table 6.14 shows the results of the normality test. The p-value is greater than 0.05, therefore, we do not reject the null hypothesis and conclude that the residuals are normally distributed.

We test for homoscedasticity. First, we use graphical methods for detecting

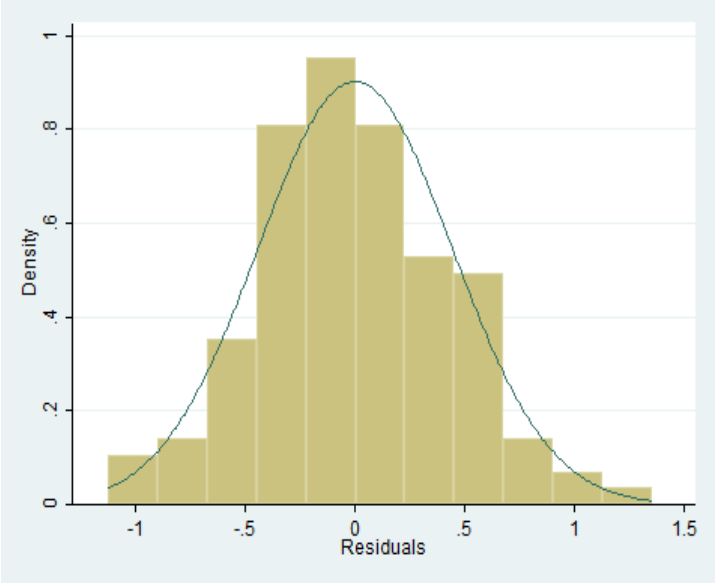

Figure 6.11: Histogram of residuals for model 3

homoscedasticity in Figure 6.12. We observe that the scatter plot is evenly distributed along the x-axis. Then, we use a non-graphical method, Breusch-Pagan / Cook-Weisberg method to test for heteroskedasticity. The details of the results are displayed in Table 6.15. Prob $>\chi^2(1)$ is greater than 0.05, hence, we do not reject the null hypothesis. The data is homoscedastic meaning all residuals have the same variance.

Variable	Obs	W	V	z	Prob > z
Residuals	126	0.99363	0.639	-1.007	0.84299

Table 6.14: Shapiro-Wilk W test for normal data for model 3

$\chi^2(1)$	0.02
Prob $>\chi^2$	0.8839

Table 6.15: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for model 3

To test the linearity assumption we draw scatter plots of each variable against

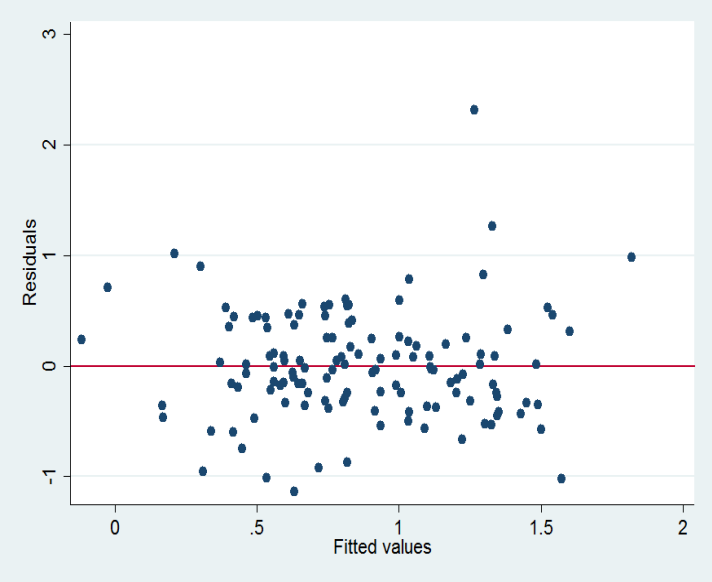

Figure 6.12: Residuals against fitted values for model 3

the residuals. The Figures are listed in 6.13, 6.14 and 6.15. There does not seem to be a strong deviation from linearity, the model satisfies the linearity assumption. Model 3 passes the assumptions of the regression model. We can use model 3 as a legitimate model in our study.

Model 4

For the fourth multiple regression model, we drop variable T from model 3;

$$\ln(IT_i) = \alpha_1 + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 \ln(PV_i) + \epsilon_i \tag{6.4}$$

We have omitted type of the firm variable, T, since it did not produce significant effects in any of the multiple regression equations. The results are given in Tables 6.16 and 6.17.

This model seems to be a better fit than the other models. Approximately 39% of the variation in $\ln(IT)$ can be explained by the model and it supports hypothesis 1, 2 and 3 and each variable is significant. The type of the firm, indeed, has a minimal effect on R^2.

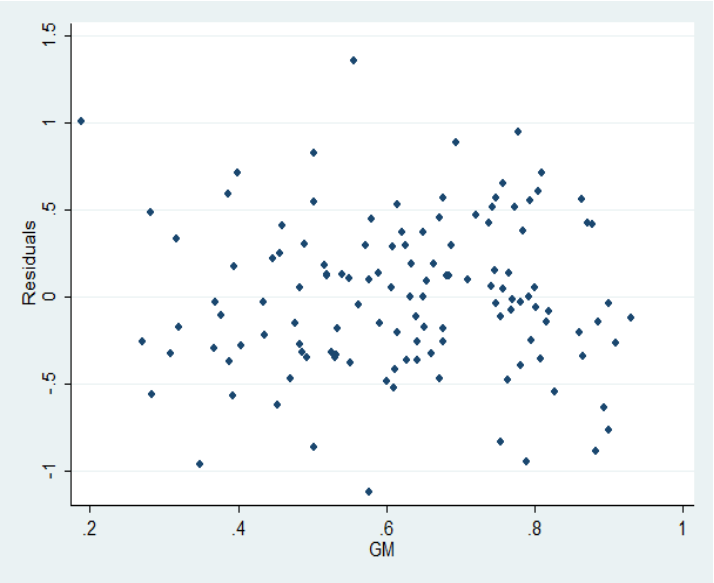

Figure 6.13: Scatter graph residual vs GM for model 3

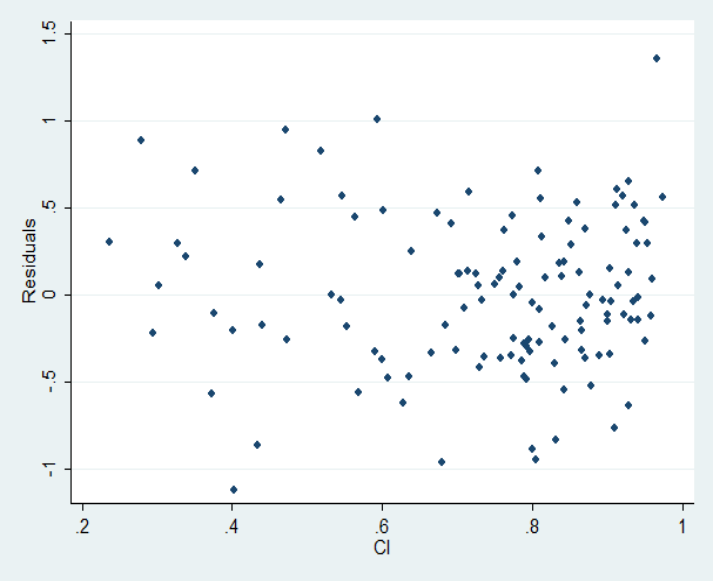

Figure 6.14: Scatter graph residual vs CI for model 3

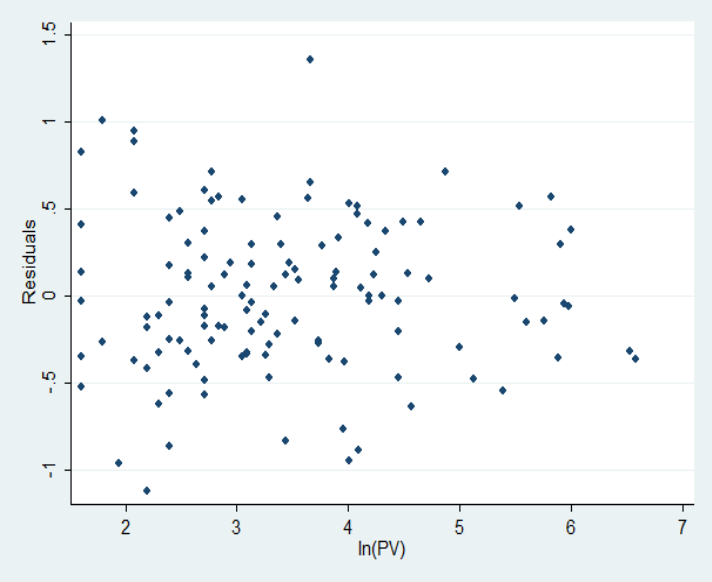

Figure 6.15: Scatter graph residual vs ln(PV) for model 3

Number of observations	126
F(3,122)	25.46
Prob > F	0.0000
R-Squared	0.3850
Adjusted R-Squared	0.3699
Root MSE	0.45052

Table 6.16: Results of multiple regression model 4

But we need to check whether the model passes the assumptions of regressions. We begin by addressing the multicolinearity issue. Figure 6.18 shows the results of the VIF calculated. According to these results multicolinearity problem does not exist.

We draw histogram of the error terms to test for normality displayed in Table 6.16. The figure shows that the error terms are normally distributed. We also use a numerical method to check for normality assumption. Table 6.19 provides the results of the normality test. From the results, we conclude that error terms are normally distributed, since the p-value is greater than our critical value of 0.05.

To check the homoscedasticity of residuals we plot the residuals against the fitted values. The resulting scatter plot is displayed in Figure 6.17. The scatter

ln(IT)	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
GM	-2.23789	0.2733861	-8.19	0.000	[-2.779085, -1.696695]
CI	1.29564	0.2617338	4.95	0.000	[0.7775118, 1.813768]
ln(PV)	-0.0850832	0.0372353	-2.29	0.024	[-0.1587943, -0.0113721]
cons.	1.582666	0.19517	8.11	0.000	[1.196308, 1.969025]

Table 6.17: Results of multiple regression model 4

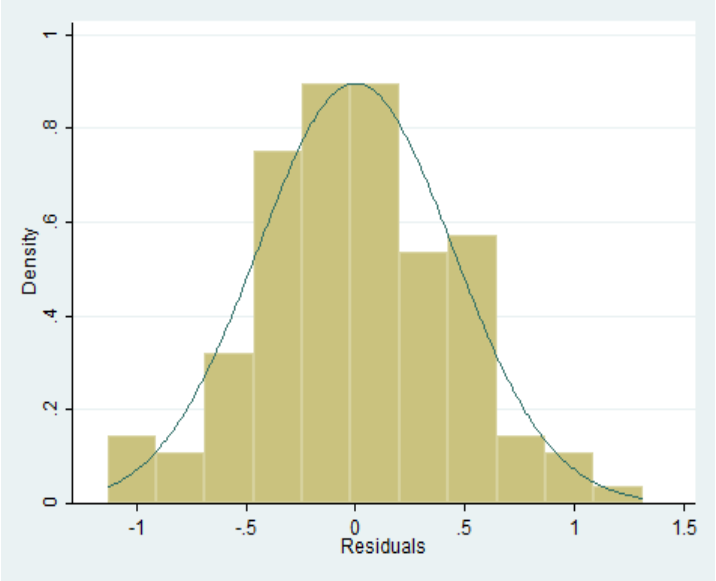

Figure 6.16: Histogram of residuals for model 4

plot seems normal. Moreover, we use a numerical method to test homoscedasticity. The results are given in Table 6.20. We do not reject the null hypothesis that the variance of the residuals is homogeneous.

Variable	VIF	1/VIF
CI	1.42	0.705277
GM	1.30	0.767683
ln(PV)	1.17	0.855562
Mean VIF	**1.30**	

Table 6.18: Test for multicolinearity for model 4

We check linearity assumption using GM, CI and $\ln(PV)$ in Figures 6.18, 6.19

Variable	Obs	W	V	z	Prob > z
Residuals	126	0.99308	0.694	-0.821	0.79409

Table 6.19: Shapiro-Wilk W test for normal data for model 4

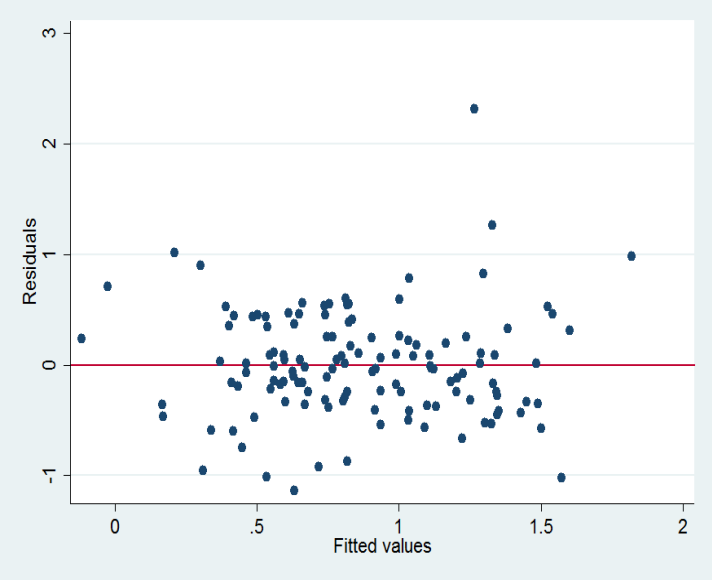

Figure 6.17: Residuals against fitted values for model 4

and 6.20.

The model passes the assumptions of regression. We observe that R^2 value has only changed 0.79%, meaning that type of the firm has an insignificant effect on the equation.

Model 5

Last multiple regression we analyzed is;

$$\ln(IT_i) = \alpha_1 + \alpha_2 GM_i + \alpha_3 CI_i + \epsilon_i \tag{6.5}$$

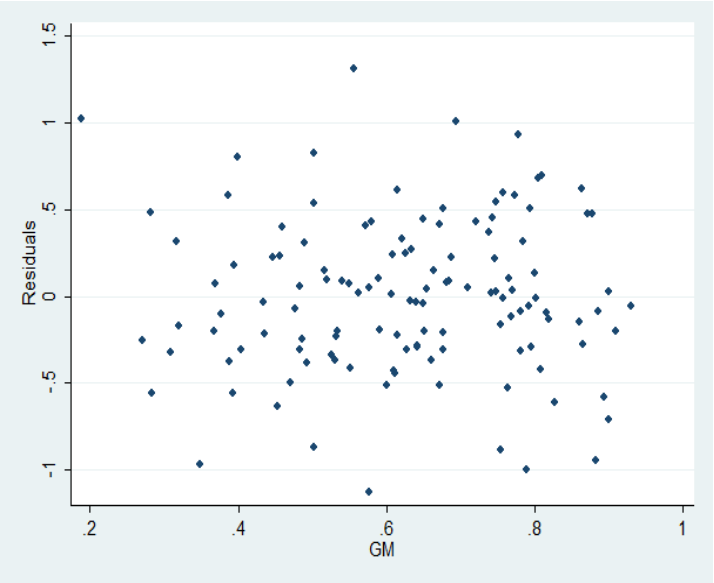

Figure 6.18: Scatter graph residual vs GM for model 4

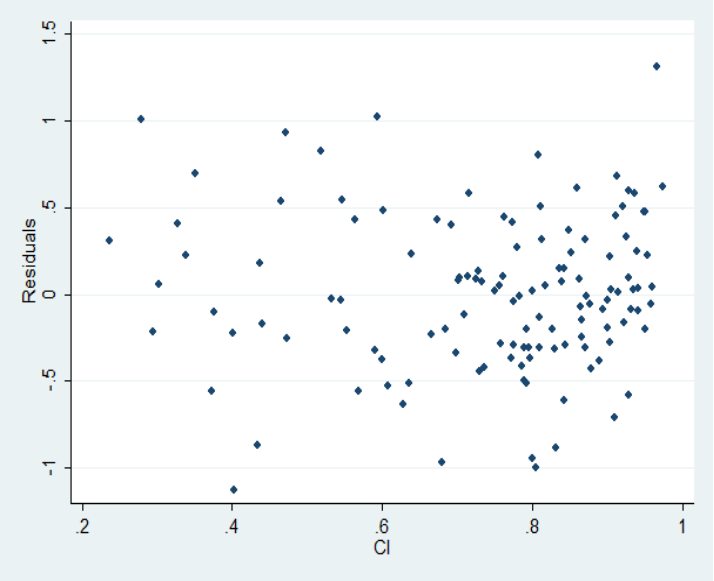

Figure 6.19: Scatter graph residual vs CI for model 4

$\chi^2(1)$	0.08
Prob $>\chi^2$	0.7719

Table 6.20: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for model 4

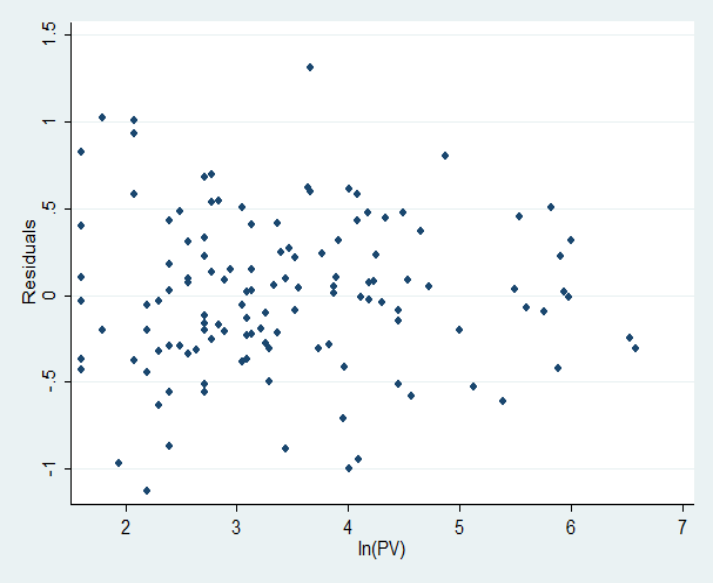

Figure 6.20: Scatter graph residual vs ln(PV) for model 4

Number of observations	126
F(2,125)	34.39
Prob > F	0.0000
R-Squared	0.3587
Adjusted R-Squared	0.3482
Root MSE	0.45819

Table 6.21: Results of multiple regression model 5

ln(IT)	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
GM	-2.292538	0.2769718	-8.28	0.000	[-2.840787, -1.744289]
CI	1.117901	0.2541604	4.40	0.000	[0.6148057, 1.620995]
cons.	1.455274	0.1902203	7.65	0.000	[1.078745, 1.831804]

Table 6.22: Results of multiple regression model 5

The model explains 35.87% of the variation on the dependent variable. Gross margin is negatively correlated with the logarithm of inventory turnover, whereas capital intensity is positively correlated. These results strongly support hypothesis 1 and hypothesis 2. Both variables are statistically significant since the p-value is smaller than our critical value of 0.05.

We use the same tests to check whether the model passes the multiple regression assumptions. Results of the colinearity test is displayed in Table 6.23. There does not exist a strong correlation between results, the model passes the multicolinearity test.

We check whether the errors are normally distributed using a histogram and a numerical test. The histogram is presented in Figure 6.21. The figure signals a normality of residuals. Furthermore, we use the numerical test of Shapiro-Wilk W given in Table 6.24. We observe that errors are normally distributed, since we cannot reject the null hypothesis.

Then we examine the residuals versus the fitted values of IT to understand if there is a heteroskedasticity problem. There does not exist an obvious pattern that deviates from homoscedasticity in Figure 6.22. Moreover, we use Breusch-Pagan / Cook-Weisberg test for heteroskedasticity. The results are shown in Table 6.25. We do not reject the null hypothesis and conclude that all of the residuals are from a distribution with the same variance.

Variable	VIF	1/VIF
CI	1.29	0.773604
GM	1.29	0.773604
Mean VIF	**1.29**	

Table 6.23: Test for multicolinearity for model 5

Variable	Obs	W	V	z	Prob > z
Residuals	126	0.99175	0.827	-0.426	0.66506

Table 6.24: Shapiro-Wilk W test for normal data for model 5

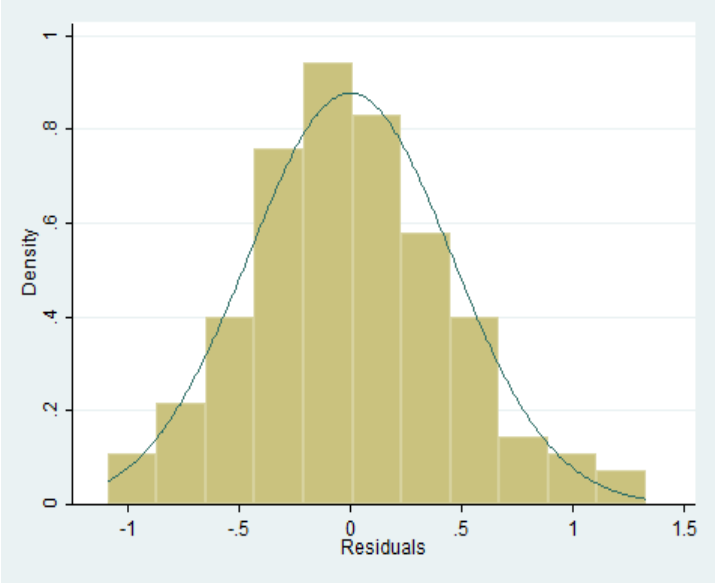

Figure 6.21: Histogram of residuals for model 5

To test for linearity assumption we use scatter plots of residuals versus GM and

$\chi^2(1)$	0.85
Prob $>\chi^2$	0.3564

Table 6.25: Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for model 5

T in Figures 6.23 and 6.24, respectively.

The absence of $\ln(PV)$ changes the R^2 value approximately 3%, which is a significant change.

Therefore, we accept Model 4 as the best multiple regression model with an R^2 value of 39%. Apart from validating Hypothesis 1 and Hypothesis 2, we conclude that as product variety increases the logarithm of inventory turnover decreases, which is stated in Hypothesis 3.

To confirm our findings we run a stepwise regression in the multiple regression model using "sw regress" command of STATA. STATA uses forward selection, which means that the model starts in an empty state with no variables and tests the significance of each variable by adding the variable that improves the

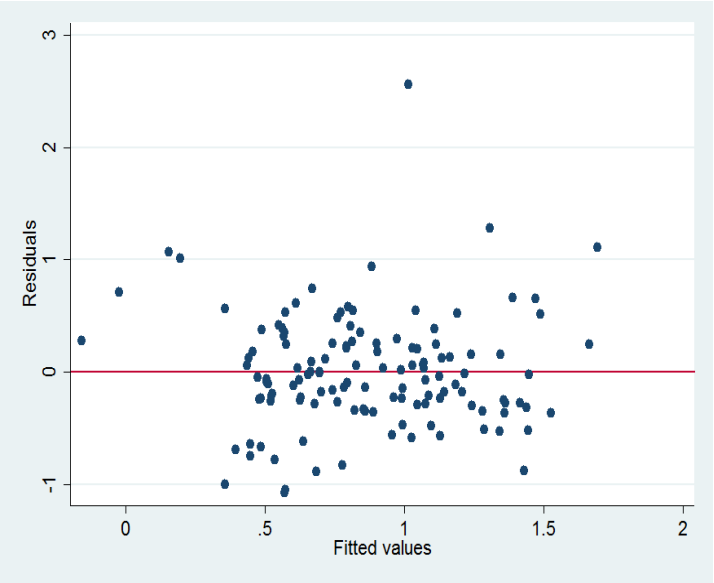

Figure 6.22: Residuals against fitted values for model 5

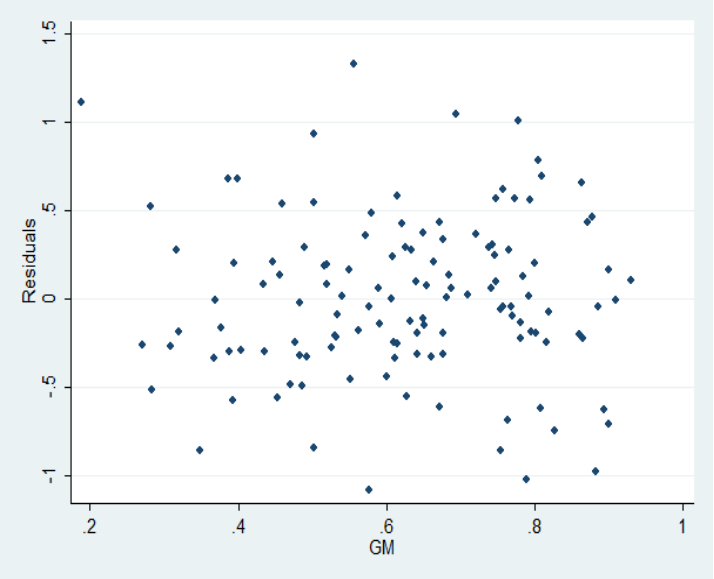

Figure 6.23: Scatter graph residual vs GM for model 5

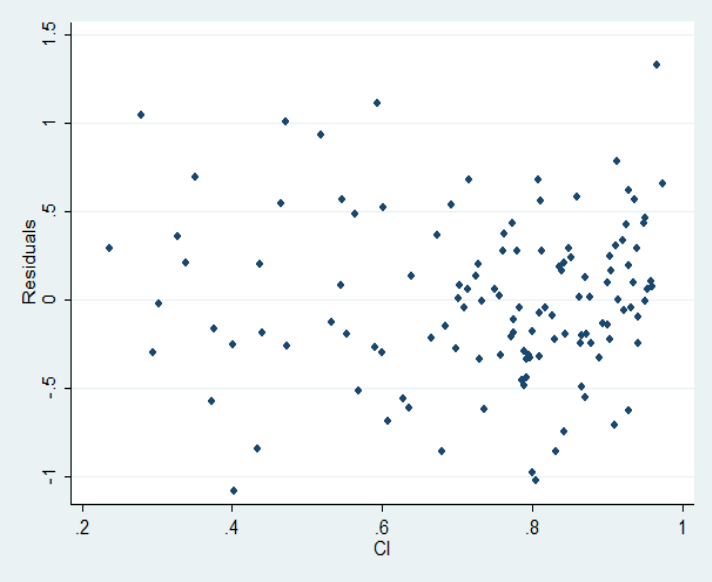

Figure 6.24: Scatter graph residual vs CI for model 5

model most. Stepwise regression model we use takes GM, CI, T and $\ln(PV)$ as explanatory variables and $\ln(IT)$ as the dependent variable. If the p-value of the explanatory variable is below 0.05 level the variable is added to the regression model and model is re-fitted with the newly added explanatory variable. This process goes on until no further variable can be added. In the stepwise regression model we build first, gross margin is added. The p-value of gross margin is 0, thus significant and added to the model. Then, the model fits CI in the regression, which has a p-value of 0. Lastly, the model includes $\ln(PV)$ that has a p-value of 0.0240 in to the regression model. Stepwise regression finds Model 4 as the best multiple regression model, just as we predicted.

Our results reports three main findings in line with the 3 hypotheses. First, gross margin is negatively correlated with inventory turnover (Hypothesis 1). Second, capital intensity is positively correlated with inventory turnover (Hypothesis 2). These results are in line with Gaur et. al [4] paper. In Gaur et. al paper 66.7% of within firm variation and 99.7% of total variation can be explained with the independent variables. Overall fit of their chosen model is statistically significant with a p-value smaller than 0.0001. Moreover, $log(GM_{sit})$, $log(CI_{sit})$ and $log(SS_{sit})$ are statistically significant in their model. The chosen model in

the paper finds segmentwise estimate of coefficients for $log(GM)$ that range from -0.153 to -0.571 and $log(CI)$ coefficients that range from 0.106 to 1.085.

Our model explains approximately 39% of the variation in inventory turnover and is significantly correlated with IT. Also, our model fit is statistically significant with a p-value of 0. We have all of our explanatory variables, GM, CI and $ln(PV)$ statistically significant. Note that GM and CI are highly statistically significant with p-values equal to 0. Our model finds GM coefficient estimate of -2.237 and a CI coefficient of 1.295. The reason that Gaur et. al model explains more significant part of the total variation may be due to the fact that they have larger sample size and use panel data instead of just a years data to explain the total variation, whereas we only use the data of 2010.

6.2 Simultaneous Equations Models

In order to validate Hypothesis 4, we need to analyze the effect of product variety on inventory level. We use simultaneous equations models to test the effect of product variety on inventory level. We cannot use OLS to explain the relationships, since product variety has a joint effect on inventory level and cost of goods sold, creating an endogeneity problem. Also, because we are dealing with a large sample analysis the least square estimator is not convenient. That is $E(\epsilon) = 0$ and $cov(x, \epsilon) = 0$ may not hold true. This can make the least square estimator biased and inconsistent. Hence, our model may have an endogeneity problem. Therefore, we use methods of moment principle of estimation as an alternative to the least square estimation. The usual assumptions of our linear and log-linear multiple regression model methods of moment leads us to the least square estimator. But when x is correlated with the error term, the method of moment leads to instrumental variable estimation that is the two stage least squares estimation, which works in large samples.

We use two stage least squares method to analyze simultaneous equations. We apply the "ivregress 2SLS" command of STATA. The commands inputs are the estimator, in our case the two stage least square estimator, $depvar_1$, the dependent variable for the model that has an endogenous regressors, $varlist_1$, which are

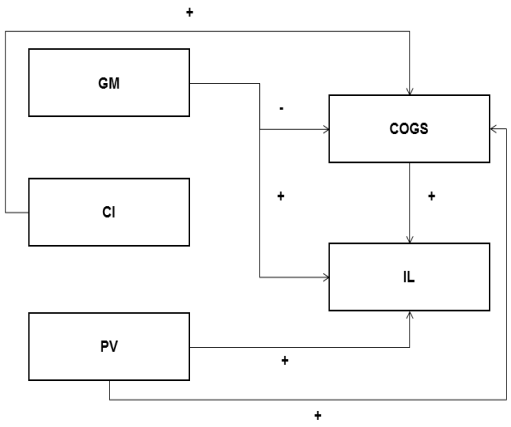

Figure 6.25: Direct effects of variables among each other

the exogenous variables in the model that has the endogenous regressors, $depvar_2$ is the endogenous regressors and $varlist_{iv}$ that are the exogenous variables that are believed to affect the endogenous regressors.

As we explained in Chapter 3, two stage least square method first estimates the parameters of the reduced form equations to obtain the predicted values then replaces the endogenous variables on the right hand side of the structural equation by their predicted values and estimates the parameters of this equation by least squares. We give the results for both of the estimations.

The hypothesized effects among GM, CI PV, $COGS$ and IL are displayed in Figure 6.25.

The rest of the chapter discusses the simultanoeus equations models and its results.

Model 1

The first regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(GM_i) + \beta_2 \ln(CI_i) + \beta_3 T_i + \beta_4 \ln(PV_i) + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(GM_i) + \alpha_3 \ln(CI_i) + \alpha_4 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.6}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, $\ln(GM)$, $\ln(CI)$ and T.

For the first stage of two stage least square estimation, R^2 equals to 0.67 and the overall F-statistic is 62.85, which has a p-value of 0. All of the estimated coefficients are statistically significant, indicating that the exogenous variables have an effect on the logarithm of inventory.

For the second stage of the estimation, R^2 equals to 96.5%. Note that, in the two stage least square estimation $\ln(CI)$ is negatively correlated with the inventory level and $\ln(COGS)$, $\ln(PV)$ and $\ln(GM)$ have positive correlations. However, $\ln(PV)$ and $\ln(CI)$ are not statisticaly significant. Other variables have statistically significant coefficient estimates.

By using the estat endogenous command of STATA, we test endogeneity of our model with the null hypothesis being tested is:
H_o: variables are exogenous.

The p-values for the endogeneity test of model 1 are more than 0.05 thus, insignificant. Hence, we reject the null hypothesis that variables are exogenous.

Furthermore, we use the "estat overid" command to test for overidentifying restrictions. The null hypothesis of the test is that instruments are uncorrelated with the error term. Since there are as many regressors as there are instruments, the model s just identified.

The detailed results of two stage least square estimation, endogeneity test and

overidentification test of model 1 is given in Table 6.26.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	62.85
Prob > F	0.0000
R-Squared	0.6751
Adjusted R-Squared	0.6643
Root MSE	1.3435

ln(COGS)	Coefficient	Std.Err.	t	P> $\|t\|$	[95 % Conf. Interval]
ln(PV)	1.015724	0.1119107	9.08	0.000	[0.7941677, 1.237281]
ln(GM)	-0.4030481	0.4427377	-0.91	0.364	[-1.279564, 0.473468]
ln(CI)	3.18577	0.4505428	7.07	0.000	[2.293801, 4.077738]
T	0.8994063	0.2852513	3.15	0.002	[0.3346761, 1.464136]
cons.	2.095058	0.5198044	4.03	0.000	[1.065968, 3.124148]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(4)$	2493.77
Prob > χ^2	0.0000
R-Squared	0.9645
Root MSE	0.44237

ln(IL)	Coefficient	Std.Err.	z	P> $\|z\|$	[95 % Conf. Interval]
ln(COGS)	0.8894393	0.1044282	8.52	0.000	[0.6847638, 1.094115]
ln(PV)	0.1815355	0.1174355	1.55	0.122	[-0.0486338, 0.4117048]
ln(GM)	1.144836	0.1440294	7.95	0.000	[0.8625432, 1.427128]
ln(CI)	-0.2481405	0.3822645	-0.65	0.516	[-0.9973651, 0.5010842]
cons.	-0.420395	0.2992105	-1.41	0.160	[-1.006837, 0.1660468]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	0.04579 (p = 0.8306)
Wu-Hausman F(1,120)	.043625 (p = 0.8349)

Table 6.26: Results for simultaneous equations model 1

Model 2

The second regression model we tested is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 ln(PV_i) + \beta_2 ln(GM_i) + \beta_3 \ln(CI_i) + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(GM_i) + \alpha_2 \ln(CI_i) + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.7}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, $\ln(GM)$, $\ln(CI)$ and T. For the first part of the least square estimation the R^2 equals to 0.675 and the overall F-statistic is 62.85, which has a p-value of 0. The logarithm of gross margin is not statistically significant since the respective p-value is bigger than 0.05. In the second stage of the estimation, we observe that the logarithm of cost of goods sold and gross margin have positive correlation with inventory level and they are statistically significant at 0.05 level. The estimation fits well with an R^2 of 96%.

The Durbin test uses an estimate of the error term's variable by assuming the variable being tested is exogenous. The Wu-Hausman test assumes that the variable being tested is endogenous. By examining the p-values of Durbin and Wu-Hausman statistics at 5 % test level we reject the null hypothesis and conclude that the variables are endogeneous.

We apply the overidentification test for model 2. The test is not statistically significant. Therefore, the regressors are overidentified.

Details of all the test results for model 2 are given in the Table 6.27.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	62.85
Prob > F	0.0000
R-Squared	0.6751
Adjusted R-Squared	0.6643
Root MSE	1.3435

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(GM)	-0.4030481	0.4427377	-0.91	0.364	[-1.279564, 0.473468]
ln(CI)	3.18577	0.4505428	7.07	0.000	[2.293801, 4.077738]
ln(PV)	1.015724	0.1119107	9.08	0.000	[0.7941677, 1.237281]
T	0.8994063	0.2852513	3.15	0.002	[0.3346761, 1.464136]
cons.	2.095058	0.5198044	4.03	0.000	[1.065968, 3.124148]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2139.40
Prob > χ^2	0.0000
R-Squared	0.9587
Root MSE	0.47738

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	1.042907	0.034953	29.84	0.000	[0.9744, 1.111413]
ln(GM)	1.176303	0.1538664	7.64	0.000	[0.8747301, 1.477875]
ln(CI)	-0.7474577	0.2206064	-3.39	0.001	[-1.179838, -0.315077]
cons.	-0.711891	0.2506946	-2.84	0.005	[-1.203243, -0.2205386]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	8.05295 (p = 0.0045)
Wu-Hausman F(1,121)	8.26139 (p = 0.0048)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	2.052 (p = 0.1520)
Basmann $\chi^2(1)$	2.00319 (p = 0.1570)

Table 6.27: Results of model 2

Model 3

The third regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 \ln(CI_i) + \beta_3 \ln(GM_i) + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(CI_i) + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.8}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(GM)$, $\ln(CI)$ and T. For the first stage of 2SLS of model 3, we observe an R^2 value of 67.5% and the F value is 62.85 with a p-value of 0. Furthermore, logarithm of gross margin is negatively correlated with cost of goods sold, whereas logarithm of capital intensity and product variety are positively correlated. Type of the company, also, has a positive effect on cost of goods sold. The last stage of 2SLS gives 94.3% value of R^2. A relatively high R^2 value implies that there is no weak instrument problem. Additionally, inventory level has a positive correlation with logarithm of cost of goods sold, capital intensity and product variety.

We do not reject the null hypothesis that the variables are exogenous, since the Durbin score and the Wu-Hausman test both has a p-value larger than the critical value of 0.05.

We have also tested whether the instruments used are correlated with the error term in our model by using the overidentification test. The results of the overidentification test indicate that the instruments are correlated with the error term. Hence we reject the null hypothesis and conclude that the instruments used cannot pass the overidentification test. The detailed results are presented in Table 6.28.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	62.85
Prob > F	0.0000
R-Squared	0.6751
Adjusted R-Squared	0.6643
Root MSE	1.3435

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	1.015724	0.1119107	9.08	0.000	[0.7941677, 1.237281]
ln(CI)	3.18577	0.4505428	7.07	0.000	[2.293801, 4.077738]
ln(GM)	-0.4030481	0.4427377	-0.91	0.364	[-1.279564, 0.473468]
T	0.8994063	0.2852513	3.15	0.002	[0.3346761, 1.464136]
cons.	2.095058	0.5198044	4.03	0.000	[1.065968, 3.124148]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(4)$	1515.88
Prob > χ^2	0.0000
R-Squared	0.9431
Root MSE	0.56016

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8028541	0.131512	6.10	0.000	[0.5450952, 1.060613]
ln(PV)	0.3134629	0.1472114	2.13	0.033	[0.0249339, 0.6019919]
ln(CI)	0.4601252	0.4707129	0.98	0.328	[-0.4624551, 1.382705]
cons.	-0.7890017	0.3743007	-2.11	0.035	[-1.522618, -0.0553858]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	0.685509 (p = 0.4077)
Wu-Hausman F(1,121)	.661908 (p = 0.4175)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	39.4037 (p = 0.0000)
Basmann $\chi^2(1)$	55.0584 (p = 0.0000)

Table 6.28: Results of model 3

Model 4

The fourth regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 \ln(GM_i) + \beta_3 \ln(CI_i) + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(GM_i) + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.9}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, $\ln(GM)$, $\ln(CI)$ and T. For model 4, the R^2 of the first stage is 67.5% and the F-value is 62.85 with 4 degrees of freedom. The related p-value for the F-statistics is 0. For the first part of the estimation, logarithm of gross margin is negatively correlated with the logarithm of cost of goods sold. However, the effect of gross margin is not significant since the p-value is above the critical level. The logarithm of product variety and capital intensity are positively correlated with cost of goods sold. Also, type of the firm is positively correlated with cost of goods sold and they are statistically significant.

The 2SLS results have an R^2 value of 96.22%. Note that the coefficient of logarithm of cost of goods sold, product variety and capital intensity are positive, indicating that as the logarithm of inventory increases so do these variables. Moreover, the p-values are statistically significant. The coefficient of product variety is 0.25. This model strongly supports hypothesis 4.

The variables have passed the endogeneity test. The p-values of Durbin and Wu-Hausman test are above the critical level of 0.05. Thus, we fail to reject the null hypothesis that the variables are exogenous.

Furthermore, we use the estat overid command to test for overidentifying restrictions because we have more instrumental variables than endogenous explanatory variables. Since the p-value of the test is above 0.05 we do not reject the null hypothesis that the instruments are uncorrelated with the error term.

The results are reported in Table 6.29.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	62.85
Prob > F	0.0000
R-Squared	0.6751
Adjusted R-Squared	0.6643
Root MSE	1.3435

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	1.015724	0.1119107	9.08	0.000	[0.7941677, 1.237281]
ln(GM)	-0.4030481	0.4427377	-0.91	0.364	[-1.279564, 0.473468]
ln(CI)	3.18577	0.4505428	7.07	0.000	[2.293801, 4.077738]
T	0.8994063	0.2852513	3.15	0.002	[0.3346761, 1.464136]
cons.	2.095058	0.5198044	4.03	0.000	[1.065968, 3.124148]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2343.94
Prob > χ^2	0.0000
R-Squared	0.9622
Root MSE	0.45625

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8268617	0.0414086	19.97	0.000	[0.7457023, 0.9080211]
ln(PV)	0.24595	0.0647736	3.80	0.000	[0.118996, 0.3729039]
ln(GM)	1.123042	0.1444569	7.77	0.000	[0.8399118, 1.406173]
cons.	-0.2578974	0.1690383	-1.53	0.127	[-0.5892065, 0.0734117]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	2.87072 (p = 0.0902)
Wu-Hausman F(1,121)	2.82107 (p = 0.0956)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 0.396123 (p = 0.5291)
Basmann $\chi^2(1)$	= 0.381603 (p = 0.5367)

Table 6.29: Results of simultaneous equation model 4

Model 5

The fifth regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.10}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, GM, CI and T. The reduced form of model 5 has a F-statistic value of 67.59, which has a p-value of 0. These values suggest that our model has statistically significant explanatory power with an R^2 of 69%. Furthermore, capital intensity and the logarithm of product variety have positive correlation with logarithm of cost of goods sold. Type of the firm has, also a positive correlation with cost of goods sold. However, the logarithm of gross margin is negatively correlated with cost of goods sold, indicating as gross margin decreases as cost of goods sold increases. But as the p-value points out gross margin is not a statistically significant variable.

The 2SLS result has a R^2 value of 96.48%. Logarithm of product variety has a coefficient equal to 0.22 but is not a statistically significant variable. The p-values indicate that the estimated slopes of the inventory curve is significantly different from zero.

According to the endogeneity test, we do not reject the null hypothesis that the variables are exogenous since the p-values are less than the critical value of 0.05.

There are no overidentifying restrictions in the model. The details of the results are displayed in 6.30.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	67.59
Prob > F	0.0000
R-Squared	0.6908
Adjusted R-Squared	0.6806
Root MSE	1.3106

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	0.9893609	0.1093316	9.05	0.000	[0.7729101, 1.205812]
GM	-1.187859	0.8113376	-1.46	0.146	[-2.794116, 0.4183974]
CI	5.902216	0.7718275	7.65	0.000	[4.37418, 7.430252]
T	0.7964782	0.2826617	2.82	0.006	[0.2368748, 1.356082]
cons.	-2.293615	0.5911784	-3.88	0.000	[-3.464009, -1.123221]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2539.74
Prob > χ^2	0.0000
R-Squared	0.9648
Root MSE	0.44063

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8625345	0.1193152	7.23	0.000	[0.6286809, 1.096388]
ln(PV)	0.2200516	0.1283106	1.71	0.086	[-0.0314327, 0.4715358]
GM	2.182231	0.281417	7.75	0.000	[1.630663, 2.733798]
CI	-0.4118744	0.7893893	-0.52	0.602	[-1.959049, 1.1353]
cons.	-1.980095	0.380406	-5.21	0.000	[-2.725677, -1.234513]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	0.262674 (p = 0.6083)
Wu-Hausman F(1,120)	0.250689 (p = 0.6175)

Table 6.30: Results of simultaneous equation model 5

Model 6

The sixth regression model we test using the two stage least square estimation model is

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 GM_i + \alpha_2 CI_i + \alpha_4 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.11}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, GM,

CI and T. The first stage results of model 6 has a R^2 value of 69% with an F-value of 67.59 and a p-value of 0. Capital intensity, type of the firm and the logarithm of product variety are positively correlated with the logarithm of cost of goods sold and these variables are statistically significant. Yet, gross margin is negatively correlated with cost of goods sold and is not statistically significant.

The two stage least square estimation has an R^2 value of 95.97%. Gross margin and logarithm of cost of goods sold have positive correlation with the logarithm of inventory level. Capital intensity is negatively correlated with inventory level. Cost of goods sold has a 105% additional influence on inventory level with a wide confidence interval.

According to the endogeneity test, we reject the null hypothesis. DWH test has values 0.0043 and 0.0045, respectively. Therefore, the variables are not exogenous.

Model passes the overidentification test, which means that chosen instruments are valid instruments.

The details of the results can be examined in Table 6.31

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	67.59
Prob > F	0.0000
R-Squared	0.6908
Adjusted R-Squared	0.6806
Root MSE	1.3106

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
GM	-1.187859	0.8113376	-1.46	0.146	[-2.794116, 0.4183974]
CI	5.902216	0.7718275	7.65	0.000	[4.37418, 7.430252]
ln(PV)	0.9893609	0.1093316	9.05	0.000	[0.7729101, 1.205812]
T	0.7964782	0.2826617	2.82	0.006	[0.2368748, 1.356082]
cons.	-2.293615	0.5911784	-3.88	0.000	[-3.464009, -1.123221]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2218.98
Prob > χ^2	0.0000
R-Squared	0.9597
Root MSE	0.47113

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	1.058744	0.0362086	29.24	0.000	[0.9877765, 1.129712]
GM	2.337936	0.2848068	8.21	0.000	[1.779725, 2.896147]
CI	-1.602815	0.4013527	-3.99	0.000	[-2.389452, -0.8161782]
cons.	-1.412448	0.2004703	-7.05	0.000	[-1.805362, -1.019533]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	8.15508 (p = 0.0043)
Wu-Hausman F(1,121)	8.37342 (p = 0.0045)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 2.57271 (p = 0.1087)
Basmann $\chi^2(1)$	= 2.52212 (p = 0.1123)

Table 6.31: Results of simultaneous equation model 6

Model 7

The seventh regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 GM_i + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.12}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, GM, CI and T. Capital intensity, type of the firm and logarithm of product variety are positively correlated with the logarithmic function of cost of goods sold. All of the aforementioned variables are statistically significant variables. The fitted values of capital intensity and logarithm of product variety, for the first stage, are high having wide range of confidence intervals. Gross margin is negatively correlated with the logarithm of cost of goods sold but it is not a statistically significant variable since it has a p-value of 0.146.

The second stage of the estimation has an R^2 of 96.15%. The logarithm of inventory level is positively correlated with gross margin, cost of goods sold and logarithm of product variety. The variables all have a p-value of 0, hence, statistically significant.

We apply Durbin and Wu-Hausman method to test for endogeneity. The results are significant, thus we reject the null hypothesis.

Furthermore, we use the estat overid command to test for overidentifying restrictions. The p-value of the overidentifying restrictions test is above our critical level of 0.05. Hence, we conclude that the instruments we have used have passed the test.

The results are displayed in 6.32

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	67.59
Prob > F	0.0000
R-Squared	0.6908
Adjusted R-Squared	0.6806
Root MSE	1.3106

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	0.9893609	0.1093316	9.05	0.000	[0.7729101, 1.205812]
GM	-1.187859	0.8113376	-1.46	0.146	[-2.794116, 0.4183974]
CI	5.902216	0.7718275	7.65	0.000	[4.37418, 7.430252]
T	0.7964782	0.2826617	2.82	0.006	[0.2368748, 1.356082]
cons.	-2.293615	0.5911784	-3.88	0.000	[-3.464009, -1.123221]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2321.67
Prob > χ^2	0.0000
R-Squared	0.9615
Root MSE	0.46083

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8036444	0.0404662	19.86	0.000	[0.7243321, 0.8829567]
ln(PV)	0.2789459	0.063812	4.37	0.000	[0.1538766, 0.4040151]
GM	2.119616	0.266219	7.96	0.000	[1.597836, 2.641396]
cons.	-2.156251	0.1833181	-11.76	0.000	[-2.515548, -1.796955]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	5.92394 (p = 0.0149)
Wu-Hausman F(1,121)	5.96952 (p = 0.0160)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 0.248888 (p = 0.6179)
Basmann $\chi^2(1)$	= 0.239485 (p = 0.6246)

Table 6.32: Results of simultaneous equations model 7

Model 8

The eighth regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 CI_i + \beta_3 GM_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 CI_i + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \quad (6.13)$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$ GM, CI and T. The results of the least square estimation of the first stage is similar to the previous model.

The 2SLS results have an R^2 value equal to 91.23%. The logarithm of inventoy level is positively correlated with capital intensity, logarithm of cost of goods sold and product variety. The p-values indicate that $\ln(PV)$ and $\ln(COGS)$ are significant, where as CI is not. Capital intensity has the biggest influence on invetory level with a coefficient value of 2.20. More importantly, product variety's regression coefficient is 0.54, supporting hypothesis 4. However, we need to be sure that the variables pass the endogeneity and overidentification tests.

The results of endogeneity test dictate that we should reject the null hypothesis that the variables are exogenous. The instruments we used in the model are not valid instruments. Since the p-values are 0.0153 and 0.0163, respectively.

Furthermore, we use the estat overid command to test for over identifying restrictions. The p-values of the test have both values 0. Therefore, overidentifying restrictions are rejected.

Using CI as an independent variable in the $\ln(IL)$ equation rather than GM, used in model 8, has decreased the R^2 value significantly. The detailed results of the model are given in Table 6.33.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	67.59
Prob > F	0.0000
R-Squared	0.6908
Adjusted R-Squared	0.6806
Root MSE	1.3106

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	0.9893609	0.1093316	9.05	0.000	[0.7729101, 1.205812]
CI	5.902216	0.7718275	7.65	0.000	[4.37418, 7.430252]
GM	-1.187859	0.8113376	-1.46	0.146	[-2.794116, 0.4183974]
T	0.7964782	0.2826617	2.82	0.006	[0.2368748, 1.356082]
cons.	-2.293615	0.5911784	-3.88	0.000	[-3.464009, -1.123221]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	996.14
Prob > χ^2	0.0000
R-Squared	0.9123
Root MSE	0.69519

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.5739998	0.1788589	3.21	0.001	[0.2234429, 0.9245568]
ln(PV)	0.5410516	0.1916143	2.82	0.005	[0.1654944, 0.9166087]
CI	2.198443	1.126524	1.95	0.051	[-0.0095039, 4.406389]
cons.	-2.220834	0.5981745	-3.71	0.000	[-3.393235, -1.048434]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	5.88732 (p = 0.0153)
Wu-Hausman F(1,121)	5.93081 (p = 0.0163)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 24.1567 (p = 0.0000)
Basmann $\chi^2(1)$	= 28.7005 (p = 0.0000)

Table 6.33: Results of simultaneous equations model 8

Model 9

The ninth regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 PV_i + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 PV_i + \alpha_2 GM_i + \alpha_3 CI_i + \alpha_4 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.14}$$

The instrumented variable is $\ln(COGS)$ and instruments are PV, CI, GM and T. First stage regression has an R^2 value of approximately 61%. Product variety is positively correlated with logarithm of cost of goods sold with a coefficient of 0.007. Capital intensity, also, has a high coefficient with 6.67.

The 2SLS estimation has an R^2 value of 96.04%. Additionally, gross margin, product variety and logarithmic functions of cost of goods sold have a positive effect on the logarithm of inventory level. Note that PV and CI are insignificant in the model with a p-value of 0.07 and 0.93.

The test of endogeneity fails to reject the null hypothesis and we conclude that the variables are exogenous. Hence, the chosen instruments are valid.

The model is not overidentified, thus, has no overidentifying restrictions.

The results are presented in Table 6.34

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	47.46
Prob > F	0.0000
R-Squared	0.6107
Adjusted R-Squared	0.5979
Root MSE	1.4705

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
PV	0.0073571	0.001161	6.34	0.000	[0.0050586, 0.0096557]
GM	-0.4896769	0.9098216	-0.54	0.591	[-2.290909, 1.311555]
CI	6.671151	0.8522959	7.83	0.000	[4.983806, 8.358496]
T	0.7489908	0.3203338	2.34	0.021	[0.1148056, 1.383176]
cons.	-0.3907765	0.6319272	-0.62	0.537	[-1.641843, 0.8602901]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	1989.69
Prob > χ^2	0.0000
R-Squared	0.9604
Root MSE	0.4674

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8348222	0.1359367	6.14	0.000	[0.5683912, 1.101253]
PV	0.0020303	0.0011313	1.79	0.073	[-0.000187, 0.0042476]
GM	2.330964	0.2825887	8.25	0.000	[1.7771, 2.884828]
CI	-0.0845111	0.9875192	-0.09	0.932	[-2.020013, 1.850991]
cons.	-1.561331	0.219445	-7.11	0.000	[-1.991435, -1.131227]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	0.766929 (p = 0.3812)
Wu-Hausman F(1,120)	0.734881 (p = 0.3930)

Table 6.34: Results of simultaneous equations model 9

Model 10

The tenth regression model we test using the two stage least square estimation model is given as;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 PV_i + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 GM_i + \alpha_2 CI_i + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.15}$$

The instrumented variable is ln($COGS$) and instruments are PV, GM, CI and T.

For the first stage of the least square estimation, we have an R^2 of 61.07% and an F value of 47.46 with a p-value of 0. Hence, the model is significant. The logarithm of cost of goods is positively correlated with product variety, type of the firm and capital intensity and negatively correlated with gross margin. However, gross margin is not a statistically significant variable.

The 2SLS estimation has a R^2 value of 0.9591, which means that we can explain approximately 96% of the variation in the logarithm of inventory level. When we analyze the estimation results, we observe that all but capital intensity have a positive relation with the logarithm of inventory level. In the second stage all of the variables are statistically significant.

We have applied the Durbin and Wu-Hausman test of endogeneity. The test results have a p-value smaller than 0.05 hence, we reject the null hypothesis that the variables used are exogenous.

Overidentifying restriction test is not rejected by Sargan $\chi^2(1)$ and Basmann $\chi^2(1)$ tests, which yield a p-value of 0.0772 and 0.0795, respectively.

Details of the results are displayed in Table 6.35.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	47.46
Prob > F	0.0000
R-Squared	0.6107
Adjusted R-Squared	0.5979
Root MSE	1.4705

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
GM	-0.4896769	0.9098216	-0.54	0.591	[-2.290909, 1.311555]
CI	6.671151	0.8522959	7.83	0.000	[4.983806, 8.358496]
PV	0.0073571	0.001161	6.34	0.000	[0.0050586, 0.0096557]
T	0.7489908	0.3203338	2.34	0.021	[0.1148056, 1.383176]
cons.	-0.3907765	0.6319272	-0.62	0.537	[-1.641843, 0.8602901]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	1926.11
Prob > χ^2	0.0000
R-Squared	0.9591
Root MSE	0.47466

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	1.065955	0.0441713	24.13	0.000	[0.9793804, 1.152529]
GM	2.338464	0.2869504	8.15	0.000	[1.776052, 2.900877]
CI	-1.663475	0.455423	-3.65	0.000	[-2.556088, -0.7708622]
cons.	-1.403695	0.2042251	-6.87	0.000	[-1.803969, -1.003421]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	5.32278 (p = 0.0210)
Wu-Hausman F(1,121)	5.33701 (p = 0.0226)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 3.12289 (p = 0.0772)
Basmann $\chi^2(1)$	= 3.07518 (p = 0.0795)

Table 6.35: Results of simultaneous equations model 10

Model 11

The eleventh regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 PV_i + \beta_2 GM_i + \beta_3 CI_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 PV_i + \alpha_2 GM_i + \alpha_3 \ln(\widehat{COGS_i}) + \epsilon_i \end{aligned} \tag{6.16}$$

The instrumented variable is $\ln(COGS)$ and instruments are $\ln(PV)$, GM, CI and T.

Just like the previous model, the least square estimation has a R-squared value equal to 0.6107.

The two stage estimation has a R^2 of 95.94%. The logarithm of inventory level is positively correlated with gross margin, product variety and the logarithm of cost of goods sold. Each of the variables is statistically significant.

Since the p-values of the Durbin and Wu-Hausman test are less than our chosen critical value of 0.05, we reject the null hypothesis. We conclude that the variables are not exogenous.

Test statistics is $\chi^2(1)$ distributed since the number of overidentifying restrictions equals to 1. Because p-values both equal to 0.93, which is greater than 0.05 ($p > 0.05$), we do not reject the null hypothesis that the overidentifying restriction is valid.

Results are shown in Table 6.36.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	47.46
Prob > F	0.0000
R-Squared	0.6107
Adjusted R-Squared	0.5979
Root MSE	1.4705

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
PV	0.0073571	0.001161	6.34	0.000	[0.0050586, 0.0096557]
GM	-0.4896769	0.9098216	-0.54	0.591	[-2.290909, 1.311555]
CI	6.671151	0.8522959	7.83	0.000	[4.983806, 8.358496]
T	0.7489908	0.3203338	2.34	0.021	[0.1148056, 1.383176]
cons.	-0.3907765	0.6319272	-0.62	0.537	[-1.641843, 0.8602901]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	1940.06
Prob > χ^2	0.0000
R-Squared	0.9594
Root MSE	0.47334

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8236232	0.0372671	22.10	0.000	[0.750581, 0.8966654]
PV	0.0021166	0.0005203	4.07	0.000	[0.0010968, 0.0031363]
GM	2.327524	0.2832693	8.22	0.000	[1.772326, 2.882721]
cons.	-1.572136	0.1817723	-8.65	0.000	[-1.928403, -1.215868]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	7.70533 (p = 0.0055)
Wu-Hausman F(1,121)	7.88154 (p = 0.0058)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 0.007141 (p = 0.9327)
Basmann $\chi^2(1)$	= 0.006858 (p = 0.9340)

Table 6.36: Results of simultaneous equations model 11

Model 12

The twelfth regression model we test using the two stage least square estimation model is;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 PV_i + \beta_2 CI_i + \beta_3 GM_i + \beta_4 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 PV_i + \alpha_2 CI_i + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.17}$$

The instrumented variable is $\ln(COGS)$ and instruments are PV, GM, CI and T. The first stage of the least square estimation is similar to previous models we have discussed.

The two stage least square estimation has a 93.71% R-squared value. The explanatory variables are all significant and are positively correlated with the dependent variable, logarithm of inventory level. Note that the coefficient of determination of capital intensity has a high value. Therefore, has a strong relationship with the dependent variable. Estimated coefficient of cost of goods sold is 0.815 indicating that $\ln(COGS)$ has a 82% effect on inventory level. However, PV and CI are insignificant variables.

As in previous models we have implemented the DWH test. Because the p-values are greater than the critical value of 0.05 we do not reject the null hypothesis that the variables are exogenous.

The overidentification test results specify that we reject the null hypothesis at 0.05 level. Hence, the chosen instruments PV, GM, CI and T are not valid.

The details are shown in Table 6.37.

Results for the first stage of 2SLS estimation

Number of observations	126
F(4,121)	47.46
Prob > F	0.0000
R-Squared	0.6107
Adjusted R-Squared	0.5979
Root MSE	1.4705

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
PV	0.0073571	0.001161	6.34	0.000	[0.0050586, 0.0096557]
CI	6.671151	0.8522959	7.83	0.000	[4.983806, 8.358496]
GM	-0.4896769	0.9098216	-0.54	0.591	[-2.290909, 1.311555]
T	0.7489908	0.3203338	2.34	0.021	[0.1148056, 1.383176]
cons.	-0.3907765	0.6319272	-0.62	0.537	[-1.641843, 0.8602901]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	1211.16
Prob > χ^2	0.0000
R-Squared	0.9371
Root MSE	0.58874

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.8150675	0.1712006	4.76	0.000	[0.4795205, 1.150614]
PV	0.0021683	0.0014249	1.52	0.128	[-0.0006244, 0.004961]
CI	1.074237	1.231239	0.87	0.383	[-1.338947, 3.48742]
cons.	-0.8735604	0.2556847	-3.42	0.001	[-1.374693, -0.3724275]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	0.653472 (p = 0.4189)
Wu-Hausman F(1,121)	.630813 (p = 0.4286)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 42.8834 (p = 0.0000)
Basmann $\chi^2(1)$	= 62.4292 (p = 0.0000)

Table 6.37: Results of simultaneous equations model 12

Model 13

In model 13, we removed GM and T variables from the $\ln(COGS)$ equation in Model 8 (6.13) and the simultaneous equation becomes;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 CI_i + \beta_2 GM_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 GM_i + \alpha_3 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.18}$$

The first stage of the model yields an F-value of 82.76 with probability 0 and an R^2 value of 67.05%. All exogeneous variables except gross margin is statistically significant. Most effective variable on cost of goods sold is capital intensity with a coefficient value of approximately 6.26.

Second stage diagnostics provide an R-squared value of 96.09%. All of the variables in the second stage are significant and positively correlated with logarithm of inventory level.

Judging by the p-values of the endogeneity test, we reject the null hypothesis that variables are exogenous.

Since model 13 does not have any overidentified instruments we cannot perform the overidentification test. The details of the results are presented in 6.38.

Model 14

Model 14 is given as;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 CI_i + \beta_3 T_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.19}$$

In the first stage of the equation, we have $\ln(COGS)$ as the dependent endogenous regressors and $\ln(PV)$, CI and T as the exogenous variables. The exogenous variables are all statistically significant with p-values less than 0.05.

The 2SLS results indicate an R^2 value of 94.57%. All of the exogenous variables are positively correlated with the inventory level. Note that all but $\ln(PV)$ is statistically significant. $\ln(PV)$ has p-value approxiamtely equal to the critical value of 0.05, thus, insignificant.

The endogeneity test results allow us to accept the null hypothesis since p-values are highly significant. Therefore, our chosen instruments are exogenous hence, uncorrelated with the error term.

Results for the first stage of 2SLS estimation

Number of observations	126
F(3,122)	82.76
Prob > F	0.0000
R-Squared	0.6705
Adjusted R-Squared	0.6624
Root MSE	1.3474

ln(COGS)	Coefficient	Std.Err.	z	P> $\|z\|$	[95 % Conf. Interval]
ln(PV)	1.031168	0.1113585	9.26	0.000	[0.8107226, 1.251613]
GM	-0.73554	0.8176064	-0.90	0.370	[-2.354074, 0.8829936]
CI	6.258468	0.7827584	8.00	0.000	[4.708919, 7.808016]
cons.	-2.757795	0.5836883	-4.72	0.000	[-3.913265, -1.602326]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(3)$	2243.59
Prob > χ^2	0.0000
R-Squared	0.9609
Root MSE	0.46396

ln(IL)	Coefficient	Std.Err.	z	P> $\|z\|$	[95 % Conf. Interval]
ln(COGS)	0.7967237	0.0430682	18.50	0.000	[0.7123116, 0.8811359]
ln(PV)	0.2879135	0.0667453	4.31	0.000	[0.1570952, 0.4187318]
GM	2.133824	0.2695548	7.92	0.000	[1.605507, 2.662142]
cons.	-2.161588	0.1848758	-11.69	0.000	[-2.523938, -1.799238]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	5.94129 (p = 0.0148)
Wu-Hausman F(1,121)	5.98787 (p = 0.0158)

Table 6.38: Results of simultaneous equations model 13

By analyzing the overidentification test results we, again, observe that $\ln(PV)$, CI and T are valid instruments since p-values are bigger than the critical value of 0.05.

The results of simultaneous equations model 14 is presented in Table 6.39.

Results for the first stage of 2SLS estimation

Number of observations	126
F(3,122)	88.57
Prob > F	0.0000
R-Squared	0.6853
Adjusted R-Squared	0.6776
Root MSE	1.3167

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	0.9800595	0.1096572	8.94	0.000	[0.7629821, 1.197137]
CI	5.472716	0.7172414	7.63	0.000	[4.052865, 6.892567]
T	0.7146002	0.2783696	2.57	0.011	[0.1635398, 1.265661]
cons.	-2.662872	0.5371785	-4.96	0.000	[-3.726271, -1.599474]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(2)$	1635.72
Prob > χ^2	0.0000
R-Squared	0.9457
Root MSE	0.5470

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.9535979	0.0462302	20.63	0.000	[0.8629884, 1.044207]
ln(PV)	0.1492705	0.0758567	1.97	0.049	[0.0005941, 0.2979469]
cons.	-1.129563	0.1527444	-7.40	0.000	[-1.428937, -0.83019]

Results of the endogeneity test

Durbin (score)$\chi^2(1)$	1.09567 (p = 0.2952)
Wu-Hausman F(1,122)	1.0702 (p = 0.3029)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	= 0.593063 (p = 0.4412)
Basmann $\chi^2(1)$	= 0.576951 (p = 0.4475)

Table 6.39: Results of simultaneous equations model 14

Model 15

Model 15 is given as;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 CI_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(\widehat{COGS_i}) + \epsilon_i \end{aligned} \tag{6.20}$$

By examining the results of Table 6.40 we observe that type of the firm does not have any significant effect on the equation, since R^2 changes approximately

1% from model 14 to model 15. The first stage still indicates a positive correlation with the endogenous regressors, especially capital intensity. The two stage least square estimation has a similar R^2 value of 94.61%. All of the independent variables are statistically significant.

We do not find any evidence of endogeneity, the results indicate that the variables are exogenous hence, we do not reject the null hypothesis.

Since $\ln(COGS)$ is exactly identified. As we have previously mentioned, under exact identification one cannot test instruments for exogeneity. Detailed results of simultaneous equations model 15 can be find in Table 6.40.

Results for the first stage of 2SLS estimation

Number of observations	126
F(2,123)	123.93
Prob > F	0.0000
R-Squared	0.6683
Adjusted R-Squared	0.6629
Root MSE	1.3463

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	1.022404	0.1108455	9.22	0.000	[0.802992, 1.241816]
CI	5.958079	0.7074194	8.42	0.000	[4.557786, 7.358373]
cons.	-2.965009	0.5359039	-5.53	0.000	[-4.025798, -1.90422]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(2)$	1602.49
Prob > χ^2	0.0000
R-Squared	0.9461
Root MSE	0.54527

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.9429849	0.0480876	19.61	0.000	[0.8487349, 1.037235]
ln(PV)	0.1637998	0.0779199	2.10	0.036	[0.0110796, 0.3165199]
cons.	-1.12679	0.1523025	-7.40	0.000	[-1.425297, -0.8282828]

Results of the endogenity test

Durbin (score)$\chi^2(1)$	.488634 (p = 0.4845)
Wu-Hausman F(1,122)	.474964 (p = 0.4920)

Table 6.40: Results of model 15

Model 16

Model 16 is given as;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 GM_i + \beta_3 CI_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.21}$$

In model 16, we add both gross margin and capital intensity to the cost of goods sold equation to analyze the joint effect of the variables. In the first stage of equation (6.21) the R^2 value is 67.05% same as the previous model. In this model, gross margin is still negatively correlated with cost of goods sold and statistically insignificant. In the second stage, the R^2 value is 94.64% and all of

the variables are statistically significant. Product variety is positively correlated with level of inventory.

Just like in the previous simultaneous equations model, this model, also, cannot pass the endogeneity test and overidentification test.

The detailed results are given in Table 6.41

Results for the first stage of 2SLS estimation

Number of observations	126
F(3,122)	82.76
Prob > F	0.0000
R-Squared	0.6705
Adjusted R-Squared	0.6624
Root MSE	1.3474

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	1.031168	0.1113585	9.26	0.000	[0.8107226, 1.251613]
GM	-0.73554	0.8176064	-0.90	0.370	[-2.354074, 0.8829936]
CI	6.258468	0.7827584	8.00	0.000	[4.708919, 7.808016]
cons.	-2.757795	0.5836883	-4.72	0.000	[-3.913265, -1.602326]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(2)$	1588.72
Prob > χ^2	0.0000
R-Squared	0.9464
Root MSE	0.5436

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	0.9085564	0.0476687	19.06	0.000	[0.8151274, 1.001985]
ln(PV)	0.2109328	0.0773673	2.73	0.006	[0.0592955, 0.36257]
cons.	-1.117793	0.1518301	-7.36	0.000	[-1.415375, -0.8202118]

Results of the endogenity test

Durbin (score)$\chi^2(1)$	0.041955 (p = 0.8377)
Wu-Hausman F(1,122)	0.040636 (p = 0.8406)

Results of the overidentification test

Sargan (score)$\chi^2(1)$	45.6489 (p = 0.0000)
Basmann $\chi^2(1)$	69.3104 (p = 0.0000)

Table 6.41: Results of simultaneous equations model 16

Model 17

Simultaneous equations model 17 is written as;

$$\begin{aligned} \ln(COGS_i) &= \beta_0 + \beta_1 \ln(PV_i) + \beta_2 GM_i + \nu_i \\ \ln(IL_i) &= \alpha_0 + \alpha_1 \ln(PV_i) + \alpha_2 \ln(\widehat{COGS}_i) + \epsilon_i \end{aligned} \tag{6.22}$$

In the first stage, the value of R^2 we obtain is 49.79%, which is infact relatively low compared to other simultaneous equations models. Gross margin and logarithm of product variety is positively correlated with cost of goods sold and they are significant.

In the second stage, R^2 value, again, is relatively low compared to other models with a value of 51.82%. $\ln(COGS)$ is positively correlated with the inventory level ,whereas $\ln(PV)$ is negatively correlated but we observe that the effect of $\ln(PV)$ is insignificant.

We have applied the DWH statistics in order to test for endogeneity. The test leads to strong rejection of the null hypothesis that the instrumented variable is exogenous.

Since there are as many instruments as regressors, the model is just identified.

The detailed results are given in Table 6.42.

Results for the first stage of 2SLS estimation

Number of observations	126
F(2,123)	60.98
Prob > F	0.0000
R-Squared	0.4979
Adjusted R-Squared	0.4897
Root MSE	1.6565

ln(COGS)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(PV)	1.295774	0.130726	9.91	0.000	[1.03701, 1.554538]
GM	2.053007	0.9091769	2.26	0.026	[0.253347, 3.852667]
cons.	-0.7710629	0.6493506	-1.19	0.237	[-2.056413, 0.5142869]

Results for the second stage of 2SLS estimation

Number of observations	126
Wald $\chi^2(2)$	154.13
Prob > χ^2	0.0000
R-Squared	0.5182
Root MSE	1.6296

ln(IL)	Coefficient	Std.Err.	z	P> |z|	[95 % Conf. Interval]
ln(COGS)	1.836089	0.4356451	4.21	0.000	[0.9822403, 2.689938]
ln(PV)	-1.058869	0.6092747	-1.74	0.082	[-2.253025, 0.1352878]
cons.	-1.360172	0.4676852	-2.91	0.004	[-2.276818, -0.4435255]

Results of the endogenity test

Durbin (score)$\chi^2(1)$	41.7438 (p = 0.0000)
Wu-Hausman F(1,122)	60.4436 (p = 0.0000)

Table 6.42: Results of simultaneous equations model 17

Tables 6.43 and 6.44 summarizes the results of the simultaneous equation models we have tested using two stage least squares. We identify the significance of each variable using two stage least squares. We display the dependent variable's p-value along with the p-values of each independent variable for every model in Tables 6.43 and 6.44. Note that, a p-value smaller than 0.05 indicates that the exogenous variable used in the model is a statistically significant variable for that equation. As we have previously mentioned, we apply DWH test, to check whether the related model has an endogeneity bias and overidentification test in order to understand whether the model uses instruments that are valid. Note that, "N/A" in an overidentification test means that the SEM is just identified. That is number of endogenous variables equal to number of instrumental variables.

When we interpret our findings we observe that Model 4 is the best model among the 17 simultaneous equations models. It has an R^2 value of 96% and it passes the overidentification test and the endogeneity test. The coefficients of Model 4 are all significantly different from zero with p-values of 0, thus, highly statistictically significant. $\ln(COGS)$, $\ln(PV)$ and $\ln(GM)$ are positively correlated with $\ln(IL)$. It follows that the endogeneous variables are not correlated with the error term, so the simultaneous equations model passes the DWH test. Moreover, Model 4 passes the test for instrument validity, meaning that instruments are exogenous. Our estimation results show that Model 4 has positive correlation between inventory level and product variety, hence, supports Hypothesis 4. Furthermore, Model 4 supports the relationships in Figure 6.25.

Unlike Kekre and Srinivasan's [9] findings, our results of simultaneous equations models indicate that health care firms can increase product variety only at the expense of increasing inventory level. Kekre and Srinivasan, also, point out that total inventory would increase. However, the authors indicated that product variety will have an impact on firm's market share, thus, reducing the level of inventory. In model 4, the first equation indicates that there is a positive correlation between product variety and cost of goods sold, which indirectly indulges an increase in inventory level. However, the increase in inventory cannot only be explained by the indirect impact of cost of goods sold. Product variety has a direct positive correlation with inventory level. This may be due to increased setup times in the product line and other risk pooling. We manage to capture exactly this effect in Model 4. The increase in inventory level, due to an increase in product variety, has relatively a larger effect than its impact on cost of goods sold.

	First Stage					Second Stage						
Models	**Dependent Variable**	**Independent Variables**				**Dependent Variable**	**Independent Variables**				**OI test**	**Endo-geneity test**
Model 1	ln(COGS)	ln(PV)	ln(GM)	ln(CI)	ln(T)	ln(IL)	ln(COGS)	ln(PV)	ln(GM)	ln(CI)		
	(0.000)	(0.000)	(0.364)	(0.000)	(0.002)	(0.160)	(0.000)	(0.122)	(0.000)	(0.516)	N/A	Pass
Model 2	ln(COGS)	ln(GM)	ln(CI)	ln(PV)	T	ln(IL)	ln(COGS)	ln(GM)	ln(CI)			
	(0.000)	(0.364)	(0.000)	(0.000)	(0.002)	(0.005)	(0.000)	(0.000)	(0.001)		Pass	Fail
Model 3	ln(COGS)	ln(PV)	ln(CI)	ln(GM)	T	ln(IL)	ln(COGS)	ln(PV)	ln(CI)			
	(0.000)	(0.000)	(0.000)	(0.364)	(0.002)	(0.035)	(0.000)	(0.033)	(0.328)		Fail	Pass
Model 4	ln(COGS)	ln(PV)	ln(GM)	ln(CI)	T	ln(IL)	ln(COGS)	ln(PV)	ln (GM)			
	(0.000)	(0.000)	(0.364)	(0.000)	(0.002)	(0.127)	(0.000)	(0.000)	(0.000)		Pass	Pass
Model 5	ln(COGS)	ln(PV)	GM	CI	T	ln(IL)	ln(COGS)	ln(PV)	GM	CI		
	(0.000)	(0.000)	(0.146)	(0.000)	(0.006)	(0.000)	(0.000)	(0.086)	(0.000)	(0.602)	N/A	Pass
Model 6	ln(COGS)	GM	CI	ln(PV)	T	ln(IL)	ln(COGS)	GM	CI			
	(0.000)	(0.146)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)		Pass	Fail
Model 7	ln(COGS)	ln(PV)	GM	CI	T	ln(IL)	ln(COGS)	ln(PV)	GM			
	(0.000)	(0.000)	(0.146)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)		Pass	Fail
Model 8	ln(COGS)	ln(PV)	GM	CI	T	ln(IL)	ln(COGS)	ln(PV)	CI			
	(0.000)	(0.000)	(0.146)	(0.000)	(0.006)	(0.000)	(0.001)	(0.005)	(0.051)		Fail	Fail
Model 9	ln(COGS)	PV	GM	CI	T	ln(IL)	ln(COGS)	PV	GM	CI		
	(0.537)	(0.000)	(0.591)	(0.000)	(0.021)	(0.000)	(0.000)	(0.073)	(0.000)	(0.932)	N/A	Pass

Table 6.43: Summary results

	First Stage					Second Stage						
Models	**Dependent Variable**	**Independent Variables**				**Dependent Variable**	**Independent Variables**				**OI test**	**Endo-geneity test**
Model 10	ln(COGS)	GM	CI	PV	T	ln(IL)	ln(COGS)	GM	CI			
	(0.537)	(0.591)	(0.000)	(0.000)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)		Pass	Fail
Model 11	ln(COGS)	PV	GM	CI	T	ln(IL)	ln(COGS)	PV	GM			
	(0.537)	(0.000)	(0.591)	(0.000)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)		Pass	Fail
Model 12	ln(COGS)	PV	CI	GM	T	ln(IL)	ln(COGS)	PV	CI			
	(0.537)	(0.000)	(0.000)	(0.591)	(0.021)	(0.001)	(0.000)	(0.128)	(0.383)		Fail	Pass
Model 13	ln(COGS)	ln(PV)	GM	CI		ln(IL)	ln(COGS)	ln(PV)	GM			
	(0.000)	(0.000)	(0.370)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)		N/A	Fail
Model 14	ln(COGS)	ln(PV)	CI	T		ln(IL)	ln(COGS)	ln(PV)				
	(0.000)	(0.000)	(0.000)	(0.011)		(0.000)	(0.000)	(0.049)			Pass	Pass
Model 15	ln(COGS)	ln(PV)	CI			ln(IL)	ln(COGS)	ln(PV)				
	(0.000)	(0.000)	(0.000)			(0.000)	(0.000)	(0.036)			N/A	Pass
Model 16	ln(COGS)	ln(PV)	GM	CI		ln(IL)	ln(COGS)	ln(PV)				
	(0.000)	(0.000)	(0.370)	(0.000)		(0.000)	(0.000)	(0.006)			N/A	Pass
Model 17	ln(COGS)	ln(PV)	GM			ln(IL)	ln(COGS)	ln(PV)				
	(0.237)	(0.000)	(0.026)			(0.004)	(0.000)	(0.082)			N/A	Fail

Table 6.44: Summary results

Chapter 7

Conclusion

In the thesis, we study the effect of gross margin, capital intensity and product variety on inventory turnover. Moreover, we analyze the relationship between product variety and inventory level. The analysis are done using empirical data obtained from Wharton Research Data Services. The dataset contains a total of 126 firms, which consist of 36 pharmaceutical and 90 medical device firms for the fiscal year 2010. We test two hypotheses of Gaur et al. [4] on this dataset, namely the effect of capital intensity to inventory turnover and the effect of gross margin to inventory turnover. However, we extend Gaur et al. [4] study and incorporate product variety's effect into the models, as suggested by the authors. We use multiple regression models and simultaneous equations models to test the hypotheses. In order to test hypotheses 1, 2 and 3 we benefit from multiple regression models, to test hypothesis 4 we use the simultaneous equation models. Our primary interest is in the effect of product variety to inventory turnover and inventory levels.

The first multiple regression model uses GM, CI, PV and T to identify the independent effects on the dependent variable, inventory turnover. Multiple regression Model 2 analyzes the effect of logarithm of product variety to inventory turnover. Model 3 incorporates the effect of GM, CI, $\ln(PV)$ and T to $\ln(IT)$. We observe that, by taking the logarithm of inventory turnover the R^2 has increased significantly. In multiple regression model 4 we have eliminated the

independent variable T, since we do not think that type of a firm has a major effect to $\ln(IT)$. The R^2 value drops slightly, just as we have predicted. Moreover, stepwise regression confirms that Model 4 is the best multiple regression model that can be built given the variables.

The last multiple regression model eliminates $\ln(PV)$ from Model 4. We observe that R^2 has dropped by 3%. We conclude that $\ln(PV)$ has a significant effect on inventory turnover. Multiple regression Model 4 supports Hyptohesis 3 with an R^2 of approximately 39%. Therefore, we choose multiple regression model 4 as the best multiple regression model.

We test Hypothesis 4 using simultaneous equations models because product variety has an influence both on inventory level and cost of goods sold, which may lead to an endogeneity problem. We construct a rich variety of simultaneous equation models to understand the correlation between product variety and inventory level. We apply instrumental variable estimator to solve simultaneous equations. More specifically, we use two stage least squares interpretation of the instrumental variables. Furthermore, we test for endogeneity bias by using Durbin-Wu-Hausman test and test for overidentifying restrictions using Sargan and Basmann tests.

Simultaneous equations model 4 stands apart from the rest of the models. The model has an R^2 value of 96.22%, which means that approximately 96.22% of the variation can be explained by the variables used. The two stage least square has all the variables statistically significant and it passes the overidentification and the endogeneity tests.

Kekre and Srinivasan [9] indicate that as product variety increases it produces an increase in market share. Simultaneously, an increase in product variety increases inventory level. However, the authors indicate that the firm will achieve higher market share due to higher product variety, the inventory level would be lower, thus, an increase in product variety will result in a reduction on the total inventory level. However, our study shows that in health care industry this would not be the case. We find that product variety triggers an increase in inventory level. We argue that, in health care industry, an increase in product line will

increase cost of goods sold and cost of goods sold will have an indirect effect on inventory level. But from Model 4, we know that product variety has a direct positive correlation with inventory level and this effect is bigger than the effect of product variety on cost of goods sold. Therefore, a company with a boarder product line as compared to a company with lower product line will on the average have more inventory held. Remember that inventory turnover is the ratio of cost of goods sold divided by the inventory level. As inventory level increases the inventory turnover ratio declines. Hence, an increase in product variety decreases inventory turnover, which was stated in Hypothesis 3.

In our multiple regression analysis, Model 4 properly explains the mechanics between product variety and inventory turnover. That is; as product variety increases inventory turnover decreases. Furthermore, we have confirmed the hypotheses of Gaur et. al [4] that gross margin is negatively correlated with inventory turnover (Hypothesis 1) and capital intensity is positively correlated with inventory turnover (Hypothesis 2). We have shown that these hypotheses, also, stand in the health care industry. Simultaneous equations model 4 explains the dynamic between product variety and inventory level. Unlike Kekre and Srinivasan's [9] paper, we find that as product variety increases so does inventory level. Kekre and Srinivasan use self-reported data, which causes bias in the estimates of the explanatory variables in the OLS estimation. Additionally, the product variety data, obtained from PIMS database, is subjective in nature, whereas we use real product variety data obtained from FDA. Moreover, the authors have performed their analysis on four subgroupings, this causes heterogeneity problem in their data. We manage to focus on a single industry so heterogeneity problem does not exist in our dataset.

We should emphasize that product variety's effect on inventory level strongly depends on the industry analyzed. Although our results only cover the health care sector, the tests can be used over wide range of industries and applications. However, in other industries measuring product variety may not be easy. Detailed analysis should be made of the effect of product variety in different sectors to capture the magnitude of its effect to inventory level and cost of goods sold. Other

limitations of this study are that, it does not incorporate operational characteristics such as price and length of product life cycle. Moreover, one can investigate the effect of product variety on a panel data to clearly see the effects of product variety.

Bibliography

[1] Standard&Poor's, *Industry Surveys Healthcare: Pharmaceuticals*. Standard & Poor's, 2010.

[2] "Wharton Research Data Services. (2011)," *University of Pennsylvania*, Retrieved on August 2011.

[3] V. Gaur, S. Kesavan, and A. Raman, "Do inventory and gross margin data improve sales forecasts for U.S. public retailers?," *Management Science*, vol. 56, no. 9, pp. 1519–1533, 2010.

[4] V. Gaur, M. L. Fisher, and A. Raman, "An econometric analysis of inventory turnover performance in retail services," *Management Science*, vol. 51, no. 2, pp. 181–194, 2005.

[5] S. Rajagopalan and A. Malhotra, "Have U.S. manufacturing inventories really decreased? An empirical study," *Manufacturing & Service Operations Management*, vol. 3, no. 1, pp. 14–24, 2001.

[6] G. P. Cachon and M. Olivares, "Drivers of finished goods-inventory in the U.S. automobile industry," *Management Science*, vol. 56, no. 1, pp. 202–216, 2010.

[7] S. Roumiantsev and S. Netessine, "What can be learned from classical inventory models: a cross-industry exploratory investigation," *Manufacturing and Service Operations Management*, vol. 9, no. 4, pp. 409–429, 2007.

[8] G. P. Cachon, T. Randall, and G. M. Schmidt, "In search of the bullwhip effect," *Manufacturing & Service Operations Management*, vol. 9, no. 4, pp. 457–479, 2007.

[9] S. Kekre and K. Srinivasan, "Broader product line: A necessity to achieve success?," *Management Science*, vol. 36, no. 10, pp. 1216–1231, 1990.

[10] Z. Ton and A. Raman, "The effect of product variety and inventory levels on retail store sales: A longitudinal study," *Production and Operations Management Society*, vol. 19, no. 5, pp. 546–560, 2010.

[11] G. van Ryzin and S. Mahajan, "On the relationship between inventory costs and variety benefits in retail assortments," *Management Science*, vol. 45, no. 11, pp. 1496–1509, 1999.

[12] H. P. Marvel and J. Peck, "Inventory turnover and product variety," *Journal of Law and Economics*, vol. 51, no. 3, pp. 461–478, 2008.

[13] A. Balakrishnan, M. S. Pangburn, and E. Stavrulaki, "Stack them high, let 'em fly: Lot-sizing policies when inventories stimulate demand," *Management Science*, vol. 50, no. 5, pp. 630–644, 2004.

[14] G. P. Cachon, "Managing supply chain demand variability with scheduled ordering policies," *Management Science*, vol. 45, no. 6, pp. 843–856, 1999.

[15] StataCorp.2009, *Stata Statistical Software: Release 11.* College Station, TX: StataCorp LP.

[16] W. H. Greene, *Econometric Analysis.* Prentice Hall, 2007.

[17] R. C. Hill, *Undergraduate Econometrics.* John Wiley & Sons Inc., 2001.

[18] T. Yamano, "Lecture notes on advanced econometrics," Updated on October 2009.

[19] J. Durbin, "Errors in variables," *Revue de l'Institut International de Statistique/Review of the International Statistical Institute*, vol. 22, no. 1/3, pp. 23–32, 1954.

[20] D. Wu, "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica*, vol. 41, no. 4, pp. 733–750, 1973.

[21] J. A. Hausman, "Specification tests in econometrics," *Econometrica*, vol. 46, no. 6, pp. 1251–1271, 1978.

[22] G. D. Eppen and L. Schrage, "Centralized ordering policies in a multi-warehouse system with lead times and random demand," *TIMS Studies in Management Sciences*, vol. 16, pp. 51–67, 1981.

[23] G. D. Eppen, "Effects of centralization on expected costs in a multi-location newsboy problem," *Management Science*, vol. 25, no. 5, pp. 498–501, 1979.

[24] A. C. Cameron and P. K. Trivedi, *Microeconometrics Using Stata*. College Station, TX: StataCorp LP, 2009.

[25] S. Benjaafar, J. S. Kim, and N. Vishwanadham, "On the effect of product variety in production - inventory systems," *Annals of Operations Research*, vol. 126, no. 1-4, pp. 71–101, 2004.

Appendix A

Data

Firm	**GM**	**CI**	**PV**	**COGS**	**IL**
Teva Pharmaceutical Industries Ltd	0.6275	0.8710	721	6,004	3,866
Mylan Inc	0.4863	0.8654	686	2,799.837	1,240.271
Roche Holding AG	0.7841	0.8705	402	11,327.784	5,306.863
Glaxosmithkline PLC	0.8018	0.8722	395	8,813.459	5,905.910
Baxter International Inc	0.5629	0.8003	377	5,707.000	2,371.000
Boston Scientific Corp	0.6884	0.9539	368	2,432.000	894.000
Smith & Nephew PLC	0.8082	0.7364	359	760.000	923.000
Abbott Laboratories	0.6763	0.9209	337	11,382.887	3,188.734
Pfizer Inc	0.8167	0.9412	317	12,423.000	8,405.000
Watson Pharmaceuticals Inc.	0.4768	0.8646	271	1,866.300	631.000
Johnson & Johnson	0.7426	0.9118	255	15,853.000	5,378.000
Novartis AG	0.7707	0.9407	243	11,609.000	6,093.000
Zimmer Holdings Inc.	0.8270	0.8420	218	730.000	936.400
Wright Medical Group Inc	0.7632	0.6077	169	122.897	166.339
Perrigo Co	0.3672	0.7922	148	1,435.817	448.871
Par Pharmaceutical Companies Inc	0.3994	0.8083	131	605.954	72.580
Becton, Dickinson and Co	0.5774	0.8179	113	3,115.201	1,145.337
ArthroCare Corp	0.7393	0.8489	105	92.651	34.087
Merck & Co Inc.	0.8941	0.9289	96	4,868.000	5,868.000
DENTSPLY International Inc.	0.5397	0.8629	93	1,022.302	308.738
Astrazeneca PLC	0.8717	0.9485	89	4,359.000	1,682.000
Eli Lilly and Co	0.8600	0.8652	86	3,231.400	2,517.700
Exactech Inc.	0.6724	0.6356	86	62.405	61.602
St. Jude Medical Inc.	0.7813	0.8944	86	1,129.423	667.545

Table A.1: Data

Firm	GM	CI	PV	COGS	IL
Varian Medical Systems Inc	0.4557	0.6384	70	1,282.650	363.933
Vascular Solutions Inc.	0.6821	0.7020	69	24.939	12.601
Cynosure Inc.	0.6321	0.5325	66	30.088	18.684
Lannett Co Inc.	0.3693	0.7327	66	78.951	19.057
Allergan Inc.	0.8774	0.9495	65	602.900	229.400
Edwards Lifesciences Corp	0.7569	0.7830	61	351.800	203.600
NuVasive Inc.	0.8824	0.8000	60	56.239	107.577
Bristol-Myers Squibb Co.	0.7731	0.9367	59	4,420.000	1,204.000
Masimo Corp.	0.7207	0.6729	59	113.241	45.028
Colgate-Palmolive Co.	0.6141	0.8590	55	6,006.000	1,222.000
Tornier NV	0.7892	0.8040	55	47.937	77.525
CONMED Corp.	0.5517	0.7862	53	319.934	172.796
Amgen Inc.	0.9006	0.9096	52	1,497.000	2,022.000
Invacare Corp.	0.3173	0.8122	50	1,175.636	174.375
Zoll Medical Corp	0.5887	0.7137	49	182.604	69.958
Covidien Plc.	0.6067	0.9138	48	4,102.000	1,381.000
Stryker Corp.	0.7102	0.7554	48	2,121.100	1,056.800
Taro Pharmaceutical Industries Ltd	0.6425	0.7570	46	140.330	83.709
Mindray Medical International Ltd.	0.6078	0.8520	43	276.262	79.185
Integra LifeSciences Holdings Corp	0.6764	0.7958	42	236.926	146.928
Merit Medical Systems Inc	0.4831	0.8093	42	153.401	60.597
Elan Corp PLC	0.5554	0.9657	39	520.000	39.000
Intuitive Surgical Inc.	0.7575	0.9278	39	342.600	86.800
Warner Chilcott Plc.	0.8635	0.9740	38	405.878	119.497
Hologic Inc.	0.6549	0.9592	35	579.654	192.482
L'Oreal S.A.	0.7472	0.9040	34	6,540.025	2,401.821
Shire PLC.	0.8862	0.9310	34	394.900	260.000

Table A.2: Data

Firm	GM	CI	PV	COGS	IL
ProPhase Labs Inc	0.6339	0.7793	32	5.309	1.682
Alphatec Holdings Inc	0.7550	0.8312	31	42.047	51.635
ICU Medical Inc.	0.5198	0.7032	31	136.644	44.056
Sirona Dental	0.6254	0.9388	30	288.542	74.027
American Bio Medica Corp.	0.4363	0.2953	29	5.874	3.604
OraSure Technologies Inc	0.6715	0.7731	29	24.645	7.346
Hemagen Diagnostics Inc.	0.4828	0.3025	28	2.700	1.386
Hospira Inc.	0.4702	0.7888	27	2,075.200	955.500
NxStage Medical Inc.	0.4036	0.7889	27	106.891	34.950
BIOLASE Technology Inc.	0.3773	0.3757	26	16.330	6.987
Cephalon Inc.	0.8655	0.9039	26	377.963	291.360
Quidel Corp.	0.5916	0.8995	25	46.289	17.707
CCA Industries Inc.	0.5727	0.3272	23	21.580	9.076
Corgenix Medical Corp.	0.6155	0.4012	23	3.175	2.500
Endo Pharmaceuticals Holdings Inc.	0.7487	0.9345	23	431.353	178.805
Syneron Medical Ltd.	0.5161	0.8362	23	91.718	22.720
Akorn Inc.	0.5323	0.6647	22	40.411	18.917
Given Imaging Ltd.	0.8191	0.8088	22	28.540	19.076
Palomar Medical Technologies Inc.	0.6614	0.7974	22	21.578	13.021
SonoSite Inc.	0.7413	0.7503	22	71.240	37.126
Kensey Nash Corp.	0.7922	0.8770	21	16.755	8.886
Spectranetics Corp.	0.7935	0.8111	21	24.345	8.054
Teleflex Inc.	0.4919	0.8894	21	915.447	338.598
Bard (C.R.) Inc.	0.6648	0.8417	19	911.900	308.900
Orthovita Inc.	0.6759	0.5522	18	30.681	22.035
Volcano Corp.	0.6848	0.7246	18	92.720	40.499

Table A.3: Data

Firm	**GM**	**CI**	**PV**	**COGS**	**IL**
Cambridge Heart Inc.	0.3193	0.4409	17	1.919	0.686
Stereotaxis Inc.	0.7477	0.5469	17	13.637	5.441
Accuray Inc.	0.5026	0.4646	16	110.227	28.186
Allied Healthcare Products Inc.	0.2705	0.4725	16	33.580	11.156
AtriCure Inc.	0.8093	0.3507	16	11.255	5.680
Estee Lauder Companies Inc.	0.8002	0.7282	16	1,560.600	826.600
CAS Medical Systems Inc.	0.4461	0.3383	15	13.341	4.813
Gen-Probe Inc.	0.7549	0.9213	15	133.178	66.416
Kinetic Concepts Inc.	0.6205	0.9244	15	765.656	172.552
LeMaitre Vascular Inc.	0.7687	0.7086	15	12.965	8.375
Mako Surgical Corp.	0.6004	0.7929	15	17.701	10.504
Meridian Bioscience Inc.	0.6515	0.6848	15	49.840	27.965
Salix Pharmaceuticals Ltd.	0.8049	0.9121	15	65.732	16.021
Trimedyne Inc.	0.3923	0.3729	15	3.890	2.613
Gilead Sciences Inc.	0.7808	0.8302	14	1,742.736	1,203.809
CareFusion Corp.	0.5190	0.9279	13	1,890.000	422.000
IRIDEX Corp.	0.4902	0.2362	13	22.276	9.212
Rochester Medical Corp.	0.5262	0.6978	13	19.637	9.240
Utah Medical Products Inc.	0.5500	0.8386	13	11.305	3.097
Digirad Corp.	0.2816	0.6009	12	40.364	5.432
Natus Medical Inc.	0.6412	0.8430	12	78.461	37.627
Cardica Inc.	0.2838	0.5688	11	2.851	1.131
Cutera Inc.	0.5806	0.5641	11	22.341	6.448
Dynatronics Corp.	0.3950	0.4372	11	19.942	5.767

Table A.4: Data

Firm	**GM**	**CI**	**PV**	**COGS**	**IL**
Medicis Pharmaceutical Corp.	0.9000	0.9040	11	69.981	39.777
Orthofix International NV	0.7963	0.7755	11	114.958	84.589
Regeneration Technologies Inc.	0.5025	0.4331	11	82.670	87.278
Bovie Medical Corp.	0.4537	0.6286	10	13.236	7.605
Cornerstone Therapeutics Inc.	0.6408	0.8994	10	45.015	15.174
Derma Sciences Inc.	0.3085	0.5904	10	39.051	12.499
ATRION Corp.	0.5338	0.8261	9	50.614	17.400
Celgene Corp.	0.9308	0.9573	9	251.018	260.130
Paradigm Medical Industries Inc.	0.5771	0.4019	9	0.307	0.509
Synergetics USA Inc.	0.6114	0.7293	9	20.234	11.891
Cardiovascular Systems Inc.	0.7778	0.4717	8	14.403	4.319
Columbia Laboratories Inc.	0.6935	0.2795	8	8.769	2.586
Retractable Technologies Inc.	0.3878	0.5990	8	22.173	8.682
Span-America Medical Systems Inc.	0.3863	0.7149	8	32.131	4.135
Hansen Medical Inc.	0.3484	0.6793	7	10.838	6.232
Rockwell Medical Technologies Inc.	0.1889	0.5942	6	48.304	2.937
United Therapeutics Corp.	0.9101	0.9494	6	54.265	35.520
Amylin Pharmaceuticals Inc.	0.6097	0.8782	5	261.014	118.629
Chembio Diagnostics Inc.	0.5019	0.5179	5	8.320	1.349
Electromed Inc.	0.7654	0.7617	5	3.356	1.471
Fonar Corp.	0.4601	0.6915	5	17.176	3.103
SeraCare Life Sciences Inc.	0.4337	0.5442	5	28.53	9.029
Trinity Biotech PLC	0.5302	0.7721	5	42.871	17.576

Table A.5: Data