Examining the Role of Personal Problems in Determining Income

by

Sara Elizabeth Greene

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 10, 2015

Approved by

John D. Jackson, Chair, Professor Emeritus of Economics
Michael L. Stern, Associate Professor of Economics
Thomas R. Beard, Professor of Economics
Henry S. Thompson, Professor of Economics

Abstract

The goal of this dissertation is to both theoretically and empirically examine how various personal problems affect income. The analysis will specifically examine the effects of divorce, alcohol use, misuse of legal drugs, and illegal drug use on income. We look at the factors affecting each of the personal problems individually and how these personal problems also affect income. By using a simultaneous equations approach, we find that, although these personal problems do not have the largest effect on income, they do have a substantial impact and therefore should not be left out of wage and income determination models.

Acknowledgments

I would first and foremost like to thank John Jackson for his outstanding guidance, support, patience, and friendship throughout my time at Auburn University. I would also like to thank Mike Stern, Randy Beard, Henry Thompson, and Mitchell Brown for their help with my dissertation. Lastly, I would like to thank my family and friends for their love and encouragement to pursue my academic goals.

Table of Contents

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

In 2013, the Substance Abuse and Mental Health Services Administration estimated that 24.6 million Americans, or 9.4% of the population aged 12 or older, were current illicit drug users. While marijuana was the most common illicit drug (used by 7.5% of those aged 12 or older), 2.5% of the population were categorized as illicit drug users for abusing prescription-type psychotherapeutics. Another common substance, alcohol, was currently used by 136.9 million Americans (or 52.2% of the population) age 12 or older[1].

The effect of substance use on an individual's physical and mental health can be extreme. For example, the Centers for Disease Control and Prevention reported that 46,471 deaths in 2013 were drug-induced, and 29,001 deaths were alcohol-induced. Stated differently, the death rate[2] for drug induced causes was 14.7 and the death rate for alcohol-induced causes was 9.2.

Also of great importance is the harmful effect of stressful events on an individual's mental health. The Centers for Disease Control and Prevention reported that 3.7% of noninstitutionalized adults incurred serious psychological distress in the past 30 days[3]. A divorce, for example, can be physically, mentally and emotionally exhausting. With the divorce

[1] Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health (NSDUH), 2013.
[2] The death rate is the rate per 100,000 population. The populations used are based on the estimates from the 2010 census.
[3] Schiller JS, Ward BW, Freeman G. Early release of selected estimates based on data from the 2013 National Health Interview Survey. National Center for Health Statistics. June 2014. Available from: http://www.cdc.gov/nchs/nhis.htm.

rate being so high in the present day (3.6 per 1,000 population in 2011[4]), the effects of this potentially stressful event cannot be overlooked.

The consequences of a decline in an individual's physical and/or mental health can be catastrophic. While worst case scenarios involve death or serious injury, costs are also imposed on an individual's personal life, family life, work life, and society at large.

Productivity costs associated with substance use or a psychologically stressful event can be extremely high. For example, substance use may result in being sick more often, inadequate sleep, increased stress and fatigue, and decreased motivation. All of these potential changes to an individual's physical and mental health would likely lead to a decrease in time spent at work, and a decrease in productivity for the time spent working. Likewise, a stressful event such as a divorce may increase stress that in turn would result in a lack of sleep, a decrease in motivation, and a decline in physical heath. A decrease in hours spent at work and a decrease in productivity while at work could yield a decrease in income.

This dissertation aims to analyze both theoretically and empirically how various personal problems may affect income. The analysis will specifically examine the effects of divorce, illegal drug use, misuse of legal drugs, and alcohol use on income.

Chapter II will analyze the effect of these personal problems on income from a theoretical perspective by applying the theory of wage determination, which involves looking at both the demand for and the supply of labor. Because labor income equals hourly wage multiplied by hours worked, the theory of wage determination will lead to predictions about how these personal problems may affect income. Because a personal problem is likely to affect both the demand for and the supply of labor, the effect on income will be theoretically ambiguous. An

[4] CDC/NCHS National Vital Statistics System

empirical investigation of how these personal problems affect a person's income will be needed to resolve this ambiguity. A deeper analysis of how these personal problems can affect a worker is needed before undertaking such an empirical investigation.

Chapter III will analyze how a divorce can affect productivity, summarize the extensive literature on divorce, and form and develop a model to observe what factors determine the probability of divorce. Specifically, it will analyze how substance use, personal characteristics, family characteristics, and economic characteristics affect the probability of divorce.

Chapter IV will delve more deeply into the issue of substance use: specifically, alcohol use, illegal drug use, and misuse of legal drugs. It will address how substance use can affect productivity, analyze literature on substance use, and develop a model to predict the probability of using these substances. It will investigate how being divorced, and other personal, family, and economic factors, affect the probability of substance use.

Chapter V will briefly summarize some of the literature on income determination, then present and discuss an income determination model to address how divorce, substance use, and other personal, family, and economic characteristics affect income.

Several relationships to be analyzed may be simultaneously determined. Therefore, Chapter VI will look at simultaneous equation models in the literature, and control for simultaneity bias in the models discussed in Chapters III, IV, and V. By statistically exploiting their joint determination, this chapter presents the most reliable estimates of the effects of the various determinants on personal problems, and more importantly, the effect of these personal problems on income. Chapter VII will conclude by discussing the major findings from this dissertation, the implications of the results as well as potential policy implications, and future research ideas.

CHAPTER II

THEORY AND DATA

We are interested in how personal problems, such as substance use or undergoing a divorce, affect an individual's income. In this chapter, we aim to use economic theory to address this question. First, we will discuss the theory of wage determination by considering both the demand for and supply of labor. Then, because labor income[5] is equal to hourly wage multiplied by hours worked, we will be able to discuss the effect of these personal problems on both wages and income.

Wage determination involves looking at both the demand for and the supply of labor. Assuming a perfectly competitive labor market, a firm's goal is to maximize profit in the production of output by optimally using various inputs such as labor (L) and capital (K). Maximum profit in the production of output will occur when the firm's marginal revenue (MR) is equal to its marginal cost (MC). This condition can be translated into a profit-maximizing condition for the purchase of inputs.

While all inputs are variable in the long run, we can assume that the only variable input in the short run is the volume of labor services. Therefore, because we are assuming that the firm is operating in the short run, we will represent the rigidity of the capital input as "$\overline{K}$". Designating the price of a unit of labor by w (the wage rate) and the price of a unit of capital by r (the rental rate of capital), the firm's total cost (TC) equal to:

[5] Henceforth, the term "income" will mean "labor income". In other words, income means income accumulated from working and is calculated as hourly wage times hours worked.

$$TC = FC + r\bar{K} + wL \tag{2.1}$$

where total cost (TC) equals fixed cost (FC), plus payments to the fixed capital input ($r\bar{K}$), plus variable cost (VC), payments to the variable input labor (wL).

In other words, the firm's total variable cost (TVC) is equal to the wage rate times the amount of labor:

$$TVC = wL \tag{2.2}$$

This makes the firm's marginal cost equal to the following:

$$MC = \frac{\Delta TC}{\Delta Q} = \frac{\Delta TVC}{\Delta Q} = \frac{\Delta wL}{\Delta Q} = \frac{w\Delta L + L\Delta w}{\Delta Q} \tag{2.3}$$

Because $\Delta w = 0$ in a perfectly competitive labor market, this makes:

$$MC = w\frac{\Delta L}{\Delta Q} \tag{2.4}$$

Because $\Delta Q/\Delta L$ is the marginal product of labor, or MP_L, Equation 2.4 can be restated as:

$$MC = w\frac{\Delta L}{\Delta Q} = w\frac{1}{MP_L} = \frac{w}{MP_L} \tag{2.4}$$

Accordingly, because marginal cost equals marginal revenue in a perfectly competitive market, maximizing profit in the hiring of inputs means:

$$MC = MR = \frac{w}{MP_L} \tag{2.5}$$

Rearranging this equation to solve for the wage rate, we get

$$w = MR * MP_L = MRP \tag{2.6}$$

meaning that the firm maximizes profit by hiring labor input up to the point where the cost of the last unit of labor input employed (the nominal wage rate) is equal to the amount of revenue that unit adds to the firm (its marginal revenue product, defined as $MRP = MR * MP_L$). Assuming perfect competition in the product market, a firm's marginal revenue is the product price (P), meaning

$$w = P * MP_L = VMP_L \tag{2.7}$$

This relationship implies that the firm will maximize profit by hiring up to the point at which the marginal product of labor ($VMP_L = P * MP_L$) is equal to the nominal wage.

This rule holds because of the firm's profit-maximizing objective. A firm will not hire an additional worker if the cost of hiring the additional worker is greater than the additional revenue the firm will receive from hiring him. If hiring an additional worker would cost $10 an hour (his wage), but that worker can only bring in an extra $4 an hour (the value of his marginal product), the firm will not hire him. If, however, that worker could bring in an extra $10 an hour or more, the firm will hire him.

In Figure 1, Panel A we depict a firm's total product curve, which reflects one aspect of a firm's production function ($Q = Q(L, K_P, K_H, T, R)$). This curve shows how total product (Q) changes as we vary one of the firm's inputs, keeping all other inputs constant. In this case, we are varying the amount of labor (L) while holding all other inputs such as human capital (K_H), physical capital (K_P), technology (T), and resources (R) constant. This production function is upward sloping because increasing labor increases total product. The additional total product that is produced by hiring one more worker is known as the marginal product (MP) and more specifically the marginal product of labor (MP_L). Marginal product is the slope of the total product curve and is given by

$$MP_L = \frac{\partial Q}{\partial L} \approx \frac{\Delta Q}{\Delta L} \tag{2.8}$$

The slope of the total product curve is increasing at a decreasing rate (getting flatter and flatter as L increases) because each additional worker hired is contributing less and less to this total product. In other words, the MP_L decreases as L increases meaning there is a diminishing marginal product of labor. This inverse relationship is shown in Figure 1, Panel B with the MP_L curve.

The total product curve can rotate if any input (besides labor) changes. For instance, an increase in human capital (K_H) would rotate the total product curve upwards as seen in Figure 1, Panel A. This is because for any given level of labor, each worker is more productive and thus the MP_L will be higher. This increase in the MP_L is shown not only with a steeper slope of the total product curve in Figure 1, Panel A, but also with the upward/rightward shift of the marginal product of labor curve in Figure 1, Panel B.

The value of a worker's marginal product (VMP) can be found by multiplying a constant price of the output (P) by his marginal product. Given that the firm will hire up to the point where wage equals the VMP, if the wage is equal to w_1, buyers will buy the amount given by $P * MP_1$ (as seen in the original demand curve in Figure 1, Panel C).

Varying the price of labor will produce the VMP curve which is the demand curve for labor. This labor demand curve is downward sloping because (as stated earlier) as the firm continues to hire workers, the marginal product of each worker is decreasing (as is seen on the original MP_L curve in Figure 1, Panel B). This is known as the law of diminishing marginal returns.

If there is a change in price of the output, or a change in the marginal product of labor, there will be a shift in the demand for labor. For example, an increase in education or experience (*i.e.* human capital) would increase worker productivity (the marginal product of labor), thereby increasing the demand for labor. This increase in the marginal product of labor is demonstrated in Figure 1, Panel B and the resulting increase in the demand for labor is shown in Figure 1, Panel C. Any event or change that would decrease a worker's productivity would decrease the demand for his labor.

The theory behind labor supply stems from the idea that every individual faces a trade-off between the consumption of goods and the consumption of leisure. This tradeoff is shown best with the following utility function, where the individual is assumed to choose between real income derivable from work, Y, and leisure, S, so as to maximize this utility subject to an income constraint:

$$U = U(Y, S) \tag{2.9}$$

We assume $\frac{\partial U}{\partial Y} > 0$ and $\frac{\partial U}{\partial S} > 0$. We also assume that the set of pairs (Y, S) that yield the same utility for an individual will lie on the same indifference curve, *i.e.* an individual is indifferent between all of these combinations of Y and S yielding him equal satisfaction. For example, a consumer may be equally as happy with a combination that gives him much leisure and low income as a combination that gives him little leisure and a high income.

It is also worth noting that higher indifference curves (*i.e.*, more of both Y and S) yield combinations that give the consumer more utility, that indifference curves (by definition) cannot cross, and that indifference curves are negatively sloped. An individual's marginal rate of substitution between income and leisure is given by the slope of his indifference curve at a given point. Because:

$$dU = \frac{\partial U}{\partial Y} dY + \frac{\partial U}{\partial S} dS \tag{2.10}$$

and because $dU = 0$ along an indifference curve, then:

$$0 = \frac{\partial U}{\partial Y} dY + \frac{\partial U}{\partial S} dS \tag{2.11}$$

Because the slope of an indifference curve is its rise (dY) over its run (dS), it follows that

$$-\left(\frac{dY}{dS}\right) = \left(\frac{\frac{\partial U}{\partial S}}{\frac{\partial U}{\partial Y}}\right) \tag{2.12}$$

which is the marginal rate of substitution mentioned above. This diminishing marginal rate of substitution implies that an individual is willing to sacrifice less and less income for an extra hour of leisure as his leisure time increases.

We assume that a consumer maximizes the above utility function subject to his income constraint given by

$$Y = \frac{w}{p}(T - S) = w(T - S) \tag{2.13}$$

where T represents the total amount of time allotted to an individual (say, 24 hours) and S represents his leisure hours. Thus, $T - S$ gives the number of hours worked (which will be denoted L), p represents the price level, w represents the nominal wage, $\frac{w}{p}$ represents the real wage and Y represents real income. Therefore, the income constraint implies that an individual's real income will be equal to the number of hours he works (T - S) multiplied by his real wage ($\frac{w}{p}$=w) per hour.

The slope of the budget line is the rise ($dY = \frac{w}{p}T$) divided by the run ($-dS = T$), or the real wage ($-\frac{dY}{dS} = \frac{\frac{w}{p}T}{T} = \frac{w}{p}$= w). The slope of an indifference curve is $-(\partial U/\partial S)/(\partial U/\partial Y)$. Equilibrium occurs at tangency, as seen in Figure 2, Panel A, which establishes equilibrium income (Y), leisure (S), and hence, hours worked ($T - S$). Increasing $\frac{w}{p}$ (the real wage) from, say, $\frac{w_1}{p}$ to $\frac{w_2}{p}$ increases the slope of the budget line, decreasing the equilibrium amount of leisure hours (from S_1 to S_2) and increasing the equilibrium hours worked (from L_1 to L_2). Note that $L_1 = T - S_1$ and $L_2 = T - S_2$. We can map the real wages ($\frac{w}{p}$) and hours worked, (L) to form a supply of labor curve. The labor supply curve tells us the amount of labor that workers are willing and able to supply at different real wages. The S_L function tells us that

$$S_L = S(\frac{w}{p}) \tag{2.14}$$

or linearly

$$L_S = a + b\left(\frac{w}{p}\right) \tag{2.15}$$

The equation of the line representing this relationship is

$$\frac{w}{p} = -\frac{a}{b} + \frac{1}{b} L_S \tag{2.16}$$

or

$$\frac{w}{p} = g(L) \tag{2.17}$$

where $g = S^{-1}$ So, for nominal wages (as seen in Figure 2, Panel B):

$$w = p * g(L) \tag{2.18}$$

Several factors may shift the labor supply curve. For example, if we were observing real after-tax wages, then changes in the marginal tax rate would shift the labor supply curve. Also, any change in an indifference curve map, shown in Figure 2, Panel A, would also shift the labor supply curve.

For example, if an event were to occur that results in an individual preferring more leisure now than before for any given level of real income, the entire indifference curve would shift outward or rightward – from U_1 to U_1' or U_2 to U_2' in Figure 2, Panel A. This shift in the indifference curve would further result in an inward (or leftward) shift of the supply of labor curve. This example of a decrease in supply is shown in Figure 2, Panel B.

Plotting the demand and supply curves for labor together, we see that they will intersect at one point (as shown by the intersection of D_L and S_L in Figure 3). This point represents the equilibrium amount of labor supplied (hours worked) and the equilibrium wage.

If the demand for labor were to increase (due to, say, an increase the value of marginal product), the labor demand curve would shift to the right, thereby increasing the equilibrium number of hours worked and increasing the equilibrium wage. If the demand for labor were to decrease (as shown by the leftward shift from D_L to D_L' in Figure 3), the equilibrium number of hours worked would decrease as would the equilibrium wage. Because income (Y) is equal to $w * L$, a decrease in labor demand would decrease income (from $w_1 * L_1$ to $w_2 * L_2$). Similarly, an increase in the demand for labor would increase the equilibrium wage and employment level, causing income to increase.

If the supply of labor were to decrease, the labor supply curve would shift to the left, causing the equilibrium wage to increase and the equilibrium number of hours worked to decrease (as seen in Figure 4). This could either increase or decrease income, depending on the elasticities of supply and demand for labor and the size of the shift. Similarly, an increase in the supply of labor would cause the equilibrium wage to decrease and the equilibrium number of hours worked to increase. This, too, would either increase or decrease income, depending on the elasticities of supply and demand for labor and the size of the shift.

If both the demand and supply of labor were to change, both curves would shift and, depending on the direction and size of the shift, could increase or decrease the equilibrium wage and employment level. In Figure 5, which we will discuss in detail below, we have illustrated two potential scenarios with a decrease in both the demand and supply of labor. It can be seen that, depending on the elasticities of demand and supply and the size of their respective shifts, income in that event could decrease, increase, or stay the same.

As stated in Chapter I, we are interested in investigating how various personal problems (such as divorce or substance abuse) affect a person's income. These personal problems will most likely affect the worker's productivity (thereby affecting the demand for labor), and/or the worker's labor/leisure choice (thereby affecting the supply of labor).

Because personal problems (such as divorce and substance abuse) might make a worker less productive, we might suspect that these personal problems would decrease the demand for labor (the scenario presented in Figure 3). This would shift the demand curve for labor to the left. *Ceteris paribus*, this would decrease w^* and L^* (the wage rate and the amount of labor hours), thereby decreasing income, because income is equal to hours worked multiplied by the wage rate.

On the other hand, these personal problems could alter the individual's labor/leisure choice and make them prefer more leisure than previously desired. *Ceteris paribus*, this would shift the labor supply curve to the left (the scenario depicted in Figure 4), causing wages to increase but the amount of labor hours to decrease. Therefore, depending on the elasticity of demand for labor, income could either increase or decrease.

Putting together a decrease in labor demand and a decrease in labor supply (as we did in Figure 5), we see that there will be an anticipated decrease in the amount of labor hours, but an ambiguous change in the wage rate. If the decrease in labor demand is relatively larger than the decrease in labor supply, the wage rate and hours worked would decrease, thereby yielding a decrease in income. But if the decrease in labor supply is relatively larger than the decrease in labor demand, the wage rate would increase, thereby yielding an ambiguous effect on income even though hours worked decreases.

As shown in Figure 5, the original income level (which we will denote as Y_1) is represented by area $w_1, D, L_1, 0$. The second income level (which we will denote as Y_2) is represented by the area $w_1, C, L_2, 0$. The third income level (which we will denote as Y_3) is represented by the area $w_2, A, L_3, 0$. We can see that $Y_2 < Y_1$ by the amount C, D, L_1, L_2. However, Y_3 could be $\leq$ or $\geq$ Y_1 depending on whether the increase in income due to higher wages (w_2, A, B, w_1) is $\geq or \leq$ the decrease in income due to the decrease in hours worked (B, D, L_1, L_3). Because this change in income is theoretically ambiguous, it is necessary to look at this problem empirically.

As we will discuss in later chapters, the income variables in our data set are coarsely defined. Because of the way an individual's income is measured (*i.e.* without knowing if they are the head of the household or not), the income we will be observing is actually the survey respondent's "family income". Therefore, when thinking about the effect of these personal problems on income, we need to think about the effect on family income of the personal problems affecting an individual.

It may be the case that, when one working member of a family has a personal problem, other members of the family increase their workload to compensate. For example, if a husband has a substance problem that would decrease family income, his wife may increase her time at work to compensate. If, for example, the husband has a substance abuse problem would cause his income to decrease by $10,000, but his wife increases her workload and earns an additional $5,000, then we would observe only a $5,000 decrease in family income. As a result, the effects we observe on family income will most likely be conservative estimates of the true effect on individual income .

The personal problems we will explore are divorce, illegal drug use, legal drug misuse, and alcohol use. A divorce causes stress which, in turn, may make an individual more likely to miss work and, or at least alternatively, less productive while at work. Therefore, we predict that a divorce would both decrease productivity at work (thereby decreasing labor demand) and alter the worker's indifference curve map to weigh more heavily in favor of leisure (thereby decreasing labor supply). As a result, the effect of a divorce on income is theoretically ambiguous.

Similarly, alcohol and drug use could also be expected to decrease productivity at work, perhaps due to hangovers and other physical effects, thereby decreasing labor demand. At the same time, alcohol and drug use may also make an individual prefer more leisure to labor, thereby shifting his indifference curve map and decreasing labor supply. As a result, the effect of these personal problems on family income is also theoretically ambiguous because we don't know the magnitude of changes involved, *a priori*.

Because of these ambiguities, the question of how personal problems affect income begs empirical investigation. Typical data sources used for economic analysis have little or no information on these personal problems. For example, while the Integrated Public Use Microdata Series (IPUMS) and the Panel Study of Income Dynamics (PSID) do have information on divorce, they have no information whatsoever on substance use. The National Longitudinal Survey of Youth (the NLSY) does include information on divorce and substance use, but the substance use categories can only be broken down into alcohol use, marijuana use, and “hard” drug use. Most importantly, no information is provided on the misuse of legal substances – a common personal problem in the United States in the modern era.

We utilize a survey sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services, that incorporates more questions on personal matters: the 2006 National Survey on Drug Use and Health (NSDUH). The data were collected in face-to-face interviews at the respondent’s place of residence. Respondents included residents of households, noninstitutional group quarters such as shelters and dormitories, and civilians living on military bases. Residents of institutional group quarters (like jails and hospitals), homeless persons who do not use shelters, and military personnel on active duty were not included. In all, 137,057 addresses were screened nationally for this 2006 survey, 67,802 completed interviews were obtained. Due to privacy reasons, 55,279 of those records were used. Although this sample has excellent personal problems data, its economic data leaves something to be desired.

In this Chapter II, we examined theoretically how the demand for and supply of labor would change due to the impact of personal problems. The demand for labor, in a perfectly competitive market, is determined by the value of the marginal product of labor. Therefore, any factors that shift the marginal product of labor will shift the demand for the labor. The supply of labor shows willingness to work at various wages, reflecting the worker’s labor-leisure tradeoff. Therefore, any factors that alter the worker’s labor-leisure tradeoff will shift the supply of labor.

Taking into account both the demand for and supply of labor to find the equilibrium wage and amount of labor hours, we saw that any shift in the demand for or supply of labor would change this equilibrium. Because income is equal to the hourly wage times hours worked, any shifts in demand and supply clearly affect income as well. Therefore, because personal problems may decrease a worker's productivity (shifting the demand for labor) and alter their labor-leisure tradeoff (shifting the supply of labor), we concluded that personal problems should affect income but that the direction of the effect is theoretically ambiguous, *a priori*.

We can answer the question of how these personal problems affect a person's income empirically, but first we need to know more about these personal problems. In the following chapters, we will delve more deeply into the personal problems of interest: divorce, alcohol use, illegal drug use, and misuse of legal drugs. We will analyze both theoretically and empirically what factors predict an individual having these personal problems, so that we can understand these topics more clearly. Once we have a firm grasp on these personal problems, we will then be able to empirically analyze how these personal problems affect income.

CHAPTER III

DIVORCE

The previous chapter discussed the effects of personal problems, such as divorce and drug use, on the labor market through both the supply and demand of labor. This chapter will first look more deeply into how a divorce can affect productivity, then discuss some literature on the topic. To get an insight into this relationship between divorce and productivity, it is important to understand what factors affect divorce. Therefore, we will present and discuss our divorce model based on models from the literature, our predictions of how certain personal, family, and economic factors will affect the probability of divorce, and then discuss our results.

It is obvious that divorce negatively affects home life. If children are present, there is the conflict of custody battles and the disruption in the growth of at least one of the parent-child relationships. No matter the type of custody awarded, at least one parent will spend significantly less time with their children. Even if no children are present, the stress of a divorce is emotionally overwhelming for a couple. Any division of labor and specialization that once existed for the couple (such as the wife taking care of the house and the husband providing the income, or them sharing both duties) is now gone and all tasks must now be executed individually. This added stress in an individual’s life typically leads to unhealthy lifestyle patterns such as less sleep (or lower quality sleep), poorer eating habits, and indulgence in drugs and alcohol. These poor habits, the added stress, and a potentially unstable psychological state can negatively impact a person's work life as well. Absences from work may increase and the time that is spent at work may be less productive.

Gary Becker, one of the most influential economists in this area, has done extensive research on the microeconomic theory behind divorce [Becker (1974)]. Becker's use of utility theory to explain marriage and divorce is straightforward: a person will marry when the utility of being married is greater than the utility of being single. They will divorce when the utility of staying married is less than the utility of splitting up.

Becker also analyzes "optimal sorting" between mates using comparative advantages and specialization. A successful marriage should occur when the husband and wife each specialize in different skills. One mate should specialize in market skills by advancing their education and earning an income to support the family. The other mate should therefore have a comparative advantage in non-market skills, such as child-rearing and taking care of the home. This specialization makes the division of labor within the marriage equal.

When this equal division of labor does not occur, Becker asserts that an unhappy marriage will ensue and a divorce may occur. Another possible source of divorce, according to Becker, happens when one settles for a less than optimal mate. This can occur because searching for a compatible mate has costs, including both time and money.

One very important aspect of Becker's theory is that of marital-specific capital. Marital-specific capital is a term that represents investments made by married persons that would be significantly less valuable if those persons were single. Examples of this include specialized market/nonmarket skills that are more useful when married, "sexual adjustment with one's spouse", and most importantly – children [Becker (1974)]. Children are capital because parents invest their time, money, and energy into developing these children into individuals who can help with the household, eventually run a household of their own, and take care of their parents in their old age. They are marital-specific capital because the children belong to the couple rather than to either of the individuals, so that "if the household dissolves, then the returns on this investment may diminish due to child custody restrictions" [Stevenson (2007)]. Consequently, an investment in marital-specific capital (such as children) decreases the probability of divorce.

A significant amount of the literature analyzes the increase in divorce rates due to changes in divorce laws, specifically the no-fault unilateral divorce laws that were enacted in the late 1960's. These no-fault laws changed the earlier requirements of divorce by allowing one member of the marriage to seek divorce on grounds of irreconcilable differences. In other words, divorce could be obtained without any showing of fault. Naokezny, Shull, and Rodgers (1995) found that the divorce rate across fifty states was positively and significantly affected by the enactment of no-fault divorce laws as well as median family income. In other words, they found that, as median family income increases, the divorce rate in no-fault states increases. The effects of education and religiosity on the divorce rate were also assessed, but with no significant results.

While the results of Naokezny et al. results on religion and education disagree with other findings in the literature that find positive relationships between these variables and divorce (such as Bramlett and Mosher (2002)), the effects of other exogenous variables (such as spouse's labor force participation) are even more controversial. An early study by Becker, Landes, and Michael (1977) does not even include a spouse's earnings or employment when modeling divorce, but other studies show significant yet contrasting effects. Dronkers, Kalmijn, and Wagner (2006) state "An economic tradition attributes the rise in divorce rates to changes in the balance between the cost and benefits of marriage for both husband and wife. If this is true, there should be a higher divorce rate among women with high-income jobs, because a high income facilitates [the ability] to bear the costs of divorce, and women with a high income are economically more independent from their spouse" (p. 479).

Sander (1985) agrees that "the divorce rate is significantly and substantially affected by the earning ability of women in market work" (p. 519). He states "the most important aspect of the [Becker's] theory, though, is that high-wage women gain less from marriage relative to other women because the gains from specialization within marriage (the wife in household work and the husband in market work) are less" (p. 519). He notes another commonly used reason for divorce among working women is that "an increase in the earning ability of women enables them to leave an unhappy marriage and either remain divorced or remarry" (p. 519).

Sander's state-level analysis models the effects of the rural nonfarm divorce rate for men, farm assets, population density, and the labor force participation rate for farm women on the farm divorce rate. He finds a significant and positive relationship between the female labor force participation rate and the divorce rate. Amato, Booth, Johnson, and Rogers (2007) found that although wives' hours of employment increased perceptions of marital problems, they decreased marital problems due to "alleviating perceived economic hardship" thereby offsetting that effect.

From the reading of Becker and the literature mentioned above, microeconomic theories about divorce and possible predictors of divorce were developed. The sample for this model is cross-sectional data from the Center of Disease Control's 2006 National Survey of Drug Use and Health. Of the 55,279 individuals that were given this survey, the sample consists of 36,963 individuals. Those under the age of 18 were not included in the sample and any individuals who did not answer pertinent survey questions were skipped.

On the basis of those data, I propose the following model:

$$P(D) = f(S, P, F, E) \tag{3.1}$$

implying that $P(D)$ (the probability of being divorced) is a function of S (substance use and/or abuse), P (personal characteristics), F (family characteristics), and E (economic characteristics).

This model is estimated using probit rather than ordinary least squares for the following reason. Suppose we wish to analyze the propensity for, or preference for, divorce. If this preference (D^*) was observable, it would presumably be normally distributed about some mean $E(D^*)$ and be determined by several exogenous variables X_j $(j = 1, \dots, k)$. Then if we wished to model the behavior, we could set up a regression model $D^* = x\beta + \epsilon$ and estimate it with ordinary least squares. Unfortunately, D^* is not observable. All that we can observe is whether an individual is divorced ($D = 1$) or not divorced ($D = 0$).

If D^* conforms to the usual preference axioms, there is some critical value of D^*, say μ_1, such that if the preference for divorce exceeds that value, people will divorce ($D^* > \mu_1$ implies $D = 1$) and if the preference is less than that value ($D^* < \mu_1$ then $D = 0$) people will not divorce. Thus, we have a normally distributed (but unobservable) divorce preference centered on its mean, $X\beta$, that is divided into two observable categories at μ_1, such that:

$$D_i{}^* \sim N\left(X_i\underline{\beta}, \sigma^2\right) \forall\, i \tag{3.2}$$

It follows that the probability that the i^th^ individual is not divorced is

$$P(D^*{}_i < \mu_1) = F\left[\frac{\mu_1 - X_i\underline{\beta}}{\sigma}\right] - F\left[\frac{\mu_0 - X_i\underline{\beta}}{\sigma}\right] \tag{3.3}$$

where $F[\cdot]$ is the standard normal cumulative distribution function evaluated at $\cdot$ and f is the standard normal density function. It is conventional to set μ_0 = -∞, μ_1 = 0, μ_2 = +∞, and σ =1, so that this probability of being not divorced is:

$$P(D^*{}_i < 0) = P(D_i = 0) = F\left(-X_i\underline{\beta}\right) - F(-\infty) = 1 - F\left(X_i\underline{\beta}\right) \tag{3.4}$$

and, similarly, the probability that the i^th individual is divorced is

$$P(D^*_i > 0) = P(D_i = 1) = F(+\infty) - F\left(-X_i\underline{\beta}\right) = 1 - \left[1 - F\left(X_i\underline{\beta}\right)\right] = F(X_i\underline{\beta}) \quad (3.5)$$

Using these results we can write the likelihood function of the sample of n observations as

$$L^* = \prod_{i=1}^{n}(1 - F(X_i\underline{\beta})^{1-D_i}\,(F(X_i\underline{\beta})^{D_i} \quad (3.6)$$

where D_i = 1 if divorced and = 0 not divorced. This, in turn, yields the log likelihood function:

$$L = \ln L^* = \sum_{i=1}^{n}(1 - D_i)\ln(1 - F(X_i\underline{\beta})) + D_i\ln(F\left(X_i\underline{\beta}\right)) \quad (3.7)$$

This function can be differentiated with respect to the β_j's ($j = 1, \ldots, k$) to obtain the first order conditions for maximization, which consists of k equations in k unknowns (the β_j) that are nonlinear in the β_j's. These can be solved by iterative nonlinear techniques to obtain the b_j's (the maximum likelihood estimates of the β_j's). Because they are maximum likelihood estimates, the b_j's are known to be asymptotically normally distributed with mean vector $\underline{\beta}$ and a covariance matrix given by the information matrix (the negative of the inverse of the matrix of second partials of the likelihood function). This result allows us to conduct all of the typical hypothesis tests on the β_j's with asymptotic validity.

However, the b_j's do not have the same interpretation as their ordinary least squares counterparts. To see this, note that the behavioral content of our model lies in determining the change in the probability of divorce caused by a change in, say, X_j. From above, the probability of divorce is:

$$P(D^* > 0) = P(D = 1) = F(X\underline{\beta}) \quad (3.8)$$

Thus:

$$\frac{\partial P(D=1)}{\partial X_j} = \frac{\partial F(X\beta)}{\partial X_j} = f(X\beta)\frac{\partial X\beta}{\partial X_j} = \beta_j f(X\beta) \quad (3.9)$$

which is consistently estimated by $b_j f(X\underline{b})$. Because $f(X\underline{b})$ is an ordinate of the standard normal, it is always positive. Thus the sign and significance of the partial derivatives is determined by the sign and significance of b_j. The magnitude of the partial requires taking the sum of the product of, say, the sample means of the X_j's with their corresponding by probit coefficients (the b_j's), multiplying that sum by the relevant b_j, and referring the result to the standard normal ordinate table.

Accordingly, our econometric model is as follows:

$$P(Divorce) = \alpha + \beta_1 EverMisused\ LegalDrugs + \beta_2 EverUsedIllegalDrugs + \beta_3 RegularDrinker + \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White + \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize + \beta_{11} FamilyIncome + \beta_{12} FamilyIncome^2 + \beta_{13} WorkingWoman + \epsilon \tag{3.10}$$

The dependent variable in our model specified above, (ex post) $P(Divorce)$, is a dummy variable equal to one if the individual's marital status was "divorced or separated" in 2006, and equal to 0 if they were married, widowed, or had never been married. When this model is estimated with probit (as explained above) it will represent the probability of being divorced.

As the summary statistics show, 3,230 individuals are divorced in our sample. Only one other marital status question ("Number of Times Married") was asked, therefore no additional information on the timing of the divorce, the age of the spouses at the time of divorce, or the age of the spouses at the time of marriage (a common predictor of divorce) is provided.

$EverMisusedLegalDrugs$ is a dummy variable equal to one if the respondent has ever used painkillers, tranquilizers, stimulants, or sedatives non-medically. Specifically, respondents were asked, “Have you ever, even once, used any type of prescription [drug type] that was not prescribed for you or that you took only for the experience or feeling it caused?”.

$EverUsedIllegalDrugs$ is a dummy variable equal to one if respondent ever used marijuana, cocaine, crack, heroine, or hallucinogens. Hallucinogens include LSD, PCP, peyote, mescaline, mushrooms, ecstasy, or any other listed hallucinogens.

$RegularDrinker$ is a dummy variable equal to one if the answer to “When is the last time you used alcohol?” was answered with “the past 30 days”. While many studies examine the substance use of children whose parents got a divorce, we are interested in the substance use of the parents themselves. It is reasonable to conjecture that all three substance variables are positively correlated with the probability of divorce. The question, however, is whether substance use predicts divorce. For example, it may be that being divorced causes you to drink more, but does being a regular drinker increase your probability of divorce? The estimates from this model will provide inference for this prediction.

It is reasonable to assume that individuals who have used and/or abused substances are typically more volatile and risky in nature (as opposed to individuals who have not), and therefore area less suitable mate. Particularly, if a mate is addicted to one of the aforementioned substances, a marriage could be expected to suffer as well. Therefore, the predicted signs on these substance variables are positive. It is important to note, however, that the chosen alcohol variable ($RegularDrinker$) is not comparable to an alcohol abuse variable.

$CollegeGraduate$, a dummy variable equal to one if the individual has completed college, is expected to be negative. It could be argued that this coefficient should be positive because if both the husband and wife are increasing their education, there is a high probability that there is less specialization between them because, as Becker et al. point out, typically “more educated women participate more in the labor force" (p. 1147). However, an increase in the education of both spouses increases their levels of market and nonmarket skills, increasing gains from marriage. Also, the additional education for the couple increases their likely information about themselves and their spouse prior to marriage, thereby enabling them to better screen for and identify an optimal mate, make a more educated decision about getting married, and work through any marital problems more effectively. Stated more eloquently by Nakonezny et. al: "persons who are well educated may be at a lower risk for dissolution because such persons may embody greater interpersonal skills, maturity, and resources that benefit a marital relationship" (p. 478).

The data on the individual's age is categorical: an individual selects "1" if they are 12-17 years old, "2" for 18-25 years old, "3" for 26-34 years old, "4" for 35-49 years old, "5" for 50-64 years old, and "6" if the individual is 65 years or older. Midpoints for each age category were created, as well as a midpoint for the open ended interval (65 years plus)[6] which is equal to 72.

---

[6] The midpoint for the open ended interval is calculated by running a grouped data regression where the left-hand side variable is the categorical age variable, the explanatory variable is the constant term only, and the interior limits are 18, 26, 35, 60, and 65. Estimating this model yields a mean (23.98) and variance (19.51) of the distribution. A z value that corresponds to the lower limit of the highest ranked (open-ended category) is estimated with the following equation:

$$Z = \frac{OEI_{LL} - Mean}{Standard\ Deviation}$$

Where OEI_{LL} represents the open-ended interval lower limit. Using our computed values, we find a z value equal to 2.5.We then use the censored normal results to find the mean of the category in terms of z using the following equation:

$$E[z] = \frac{\theta(z)}{1 - \Phi(z)} = x$$

where θ represents the standard normal density and Φ represents the standard normal distribution. Finally, we back out the midpoint value on the original data's scale by "unnormalizing" the result above,

$$\frac{M - Mean}{Standard\ Deviation} = x \rightarrow M = Mean + (x * Standard\ Deviation)$$

where M is the desired estimate of the category midpoint and is equal to 72.

Age^2 are the midpoints squared. It is typical to expect that, after a certain age, the probability of divorce decreases. The probability of remarriage or finding another mate at a later age decreases, so the costs of a divorce might outweigh the benefits after a certain age.

$White$ is a dummy variable equal to one if the individual's race is non-Hispanic white, equal to zero if their race is non-Hispanic black/African American, non-Hispanic Native American/Alaskan Native, non-Hispanic Native Hawaiian/Other Pacific Island, non-Hispanic Asian, or Hispanic. Solely comparing the divorce rates between races is misleading because of "a complex set of historical, economic, structural, and cultural factors that have yet to be disentangled" (Amato 2010). Including certain exogenous variables in the model (such as family income) controls for some of these entangled factors, thereby allowing this race indicator to help answer this empirical question of ethnic differences in the probability of divorce. For example, we assume most Hispanics to be Catholic, and we assume most Catholics will be more hesitant to divorce than non-Catholics. Therefore, it would be reasonable to assume that Hispanics are less likely to divorce (especially when not controlling for religion).

The sign of $Fulltime$, a dummy variable equal to one if the respondent's employment status is "employed full time", could be argued either way. On the one hand, in comparison to individuals who are employed part-time, individuals who are unemployed, and individuals who are not in the labor force, we predict individuals who work full-time are more likely to get divorced. Individuals who work more are obviously at home less, which may cause conflict with a spouse if there is not an equal division of labor within the household. On the other hand, being employed full-time could imply greater specialization in market activity assuming the individual is the breadwinner. Because we control for wives working in another explanatory variable

($WorkingWoman$), $Fulltime$ will pick up this effect. Because both sides of the story can be argued, we resort to solving the sign of $Fulltime$ empirically.

$NonMSA$ is a dummy variable equal to one if the respondent lives in a location which is not part of a "Core Based Statistical Area" according to the 2000 Census data and the June 2003 CBSA classifications provided by the Office of Management and Budget. $NonMSA$ is expected to be negative. Couples living in urban areas would have a higher chance of remarriage (due to the wider pool of future mates in a larger and more concentrated population) and therefore a higher expected benefit of divorce. Couples who live in rural areas, on the other hand, are more isolated and may have a lower chance of remarriage. No information on the individual's state, city, or county of residence is provided due to privacy reasons, therefore this $NonMSA$ variable is our only option for identifying metropolitan versus nonmetropolitan effects.

The expected coefficient on $FamilySize$, a variable representing the number of persons in the respondent's household, is one debated widely in the literature as mentioned earlier. Children bring joy into a family and are the best type of marital-specific capital, which would lead one to assume that this coefficient will be negative. The value of children to a couple is substantially greater when they are together. Parents may want a divorce, but will work through their problems to make a better life for their children. On the other hand, it is possible for parents to have too many children given their current situation (either personal or financial). This could lead to stressed parents, a stressful marriage, and an increase in their probability of divorce. Both cases are plausible, therefore this question needs to be analyzed empirically.

The NSDUH's data on income and employment is very limited. All income variables (whether family income or respondent's income) are categorical in nature. More on the chosen income variable ($FamilyIncome$) will be discussed in a following chapter.

The respondents were asked which of the following categories contained their total family income: Option 1 being less than \$20,000, option 2 being between \$20,000 and \$49,999, option 3 being between \$50,000 and \$74,999, and option 4 being greater than \$75,000. Midpoints of each family income category were calculated (\$10,000, \$34,999.50, and \$62,499.50) as well as the midpoint of the open ended interval (\$106,152) using the method discussed in the footnote 6. The midpoints squared yield the $FamilyIncome^2$ variable.

$FamilyIncome$ is expected to be negative, while $FamilyIncome^2$ is expected to be positive. Parents of families with low income typically experience more stress, and more stress increases the probability of divorce. However, families with high incomes are bound to litigate more because the expected financial outcome of a divorce is much higher. In other words, high income families have more to fight about during a divorce.

The last explanatory variable, $WorkingWoman$, is a dummy variable equal to one if the individual is female and works full time, zero otherwise. The literature has much debate over the significance and sign of this variable. Using economic reasoning, there is reason to believe both sides of the story.

On the one hand, a wife working full time could increase the probability of divorce because, unless the husband takes care of the domestic duties, there is no specialization of work within the marriage. As was established by Becker, marriage may work best when the husband and wife specialize in different skills (one in market and the other in nonmarket skills). When both the husband and wife are in the work force, and no one is left at home to take care of the home and children, this usually stable division of labor is thrown off, leading to stress and increasing the risk of divorce. Also, if the wife is more successful than her husband, this can potentially threaten his position of being the family's provider and may also lead to marital problems.

Another reason, pointed out by Dronkers, Kalmijn, and Wagner (2006), is that a working woman might end up in a divorce because she can actually afford the divorce (lawyers, court fees, etc) due to her having her own income. One reason divorce was not as prevalent in the 1940's and 50's is that women could not get divorced unless the decision was mutual. Women who did not work could not afford the costs of divorce, and therefore were forced to stay in their unhappy marriages.

On the other hand, wives working fulltime could decrease the probability of being divorced. The husband may appreciate the wife's contribution to the family income, because with a higher combined income, more luxuries could be afforded, resulting in increased happiness. This could be particularly true if both the husband and wife have a high income, because they can afford to pay a third party (a housekeeper or babysitter) to take care of the domestic duties. After weighing the potential happiness and unhappiness a marriage may face due to a working wife, the prediction is that the sign of Working Woman will be positive. This hypothesis is supported by the belief that the loss of productivity from the unstable division of labor outweighs the additional income.

Because the model above is estimated using probit, the sign of the coefficients yield a slightly different interpretation from coefficients estimated using linear regression. Because the probit coefficients only provide information on the sign and significance of the variables at hand, the marginal effects and standardized marginal effects of each variable are calculated as well.

As mentioned earlier, the marginal effects are calculated by multiplying $f(\bar{z})$ (the ordinate of the standard normal evaluated at the sum of the product of the probit coefficients (the

b_j's) and their corresponding sample means) by the probit coefficients. The standardized marginal effects are calculated by multiplying the $f(\bar{z})$ by the standardized probit coefficients. The standardized probit coefficients, in turn, are calculated by multiplying the probit coefficients (the b_j's) by the ratio of the standard deviation of probit coefficients (X_j) to the standard deviation of $P(D)$ (which in our case is standardized to equal one).

The marginal effects tell us how the probability of divorce changes given a one unit change in X_j. The standardized marginal effects tell us how the probability of divorce (in terms of its standard deviation) changes given a one standard deviation change in X_j. These standardized marginal effects are useful for seeing which variables have the biggest effects on the probability of divorce[7].

Table 1 provides the probit coefficients, t-statistics, and marginal effects for modeling the probability of divorce. The coefficients on the substance use variables ($EverMisusedLegalDrugs$, $EverUsedIllegalDrugs$, and $RegularDrinker$) are as expected, positive and significant. In other words, individuals who have misused legal drugs, used illegal drugs, and drink alcohol regularly are more likely to be divorced. Specifically, individuals who are regular drinkers are .8%[8] more likely, individuals who have used illegal drugs are 2.3% more likely, and individuals who have misused legal drugs are 1.3% more likely to be divorced.

These results are consistent with Amato and Rogers (1997) who find reported marital problems of drinking and drug use to be two of the most consistent predictors of divorce (as well as infidelity, spending money foolishly, jealousy, moodiness, and irritating habits). Individuals who use and/or abuse substances are typically riskier and less rational than those who do not (as

---

[7] The standardized probit coefficients and standardized marginal effects are reported in the appendix.
[8] From here on out, the "percentages" we refer to are actually percentage points because all probabilities are measured in percentages.

are individuals who spend money foolishly and engage in infidelity). Risky and irrational people are less than optimal mates because they are associated with not only more conflict within the marriage, but a decreased ability to deal with conflicts.

$CollegeGraduate$ is negative and significant, which agrees with our prediction and a majority of the literature (Amato, 2010). In other words, individuals who increase their education by graduating from college are less likely to get divorced. The marginal effects show that college graduates are 1.9% less likely to get divorced. Nakonezny et. al states that "previous research has demonstrated a general inverse relationship between marital dissolution and education attainment (*e.g.*, Bumpass & Sweet, 1972; Glenn & Supancic, 1984; Glick, 1984; Goode, 1962; Heaton, 1991; Kurdek, 1993; Morgan and Rindfuss, 1985; Udry, 1966)" (p. 478).

An increase in an individual's education probably yields better screening skills when searching for compatible mates, as well as better problem-solving skills when dealing with marital problems. According to Amato (2010), this concurs with the divorce trend found by Bramlett and Mosher in 2002 that "With respect to education, individuals with college degrees tend to have more stable marriages than do individuals who [sic] high school diplomas or less education" (Amato, 2010, p.651).

The question of any ethnic differences in the probability of divorce is answered with a negative and significant coefficient on $White$, implying that individuals who are white are less likely to get divorced. In other words, there is some ethnic or racial trait of Caucasians (that is not controlled for by the other exogenous variables) that predicts them to be less likely (.5% in the marginal effects) to get divorced. This agrees with the divorce trends discussed in Amato's paper (2010). He states that African-Americans are associated with the highest rate of divorce (see also, Bramlett & Mosher, 2002), and that Hispanics have a higher divorce rate than non-Hispanic Whites.

$NonMSA$ is negative but not significant, implying that our model does not predict population density to be a major factor of divorce

The coefficients on Age and Age^2 (significantly positive and negative, respectively) yields an upside-down U-shaped relationship between age and the probability of divorce. As an individual's age increases, their probability of divorce increases. Once some age is reached, however, an increase in the individual's age leads to a decrease in their probability of divorce. This age where the probability changes sign is 50 years old[9]. This implies that, as an individual's age increases, their probability of divorce increases until they reach age 50. Once an age of 50 is reached, their probability of divorce decreases as their age continues to increase. In other words, younger couples become more and more likely to divorce, while couples over the age of 50 become less and less likely to divorce.

$FamilySize$ is negative and significant, implying that individuals with larger families are less likely to get divorced. This result concurs with Becker's idea that an increase in the number of children (a type of marital-specific capital) will decrease a couple's probability of divorce. Our marginal effects results yield that increasing family size by one (having an additional child) decreases the probability of divorce by 1.2%.

$WorkingWoman$ is positive and significant while $Fulltime$ is positive but insignificant, implying that females who work fulltime are more likely to get divorced, while men working fulltime doesn't affect the probability of divorce. Specifically, as told by the marginal effects,

[9] We can find this critical value using the fact that $\frac{\partial P(D)}{dA}(\beta_5 + 2\beta_6 A)xf(X\beta) = 0$. This implies that $\beta_5 + 2\beta_6 A = 0$, meaning that $2\beta_6 A = -\beta_5$, and therefore $A = \frac{-\beta_5}{2\beta_6}$. Replacing our coefficients for Age (.1608) and Age^2 (-.002) for β_5 and β_6 respectively, we find that our critical A is equal to 50.

being a working woman increases probability of divorce by 3.6%. Working fulltime means spending less time at home in comparison with working part-time, being unemployed, or not being in the labor force.

Spending less time at home can put stress on a marriage, therefore resulting in an increased probability of divorce. A woman who works full time (unless her husband is staying at home) implies a less-than-optimal sorting between mates as far as comparative advantages and specialization within a marriage are concerned. This increased probability of divorce for working women may also be due to their ability to afford a divorce. Schoen, Astone, Rothert, Standish, and Kim (2002) found that the probability of divorce increased with wives' employment for unhappy marriages, but not for happy marriages.

While $FamilyIncome$ is negative and significant, $FamilyIncome^2$ is positive and significant, indicating that the probability of divorce decreases as a family's income increases, but at some point, as family income continues to increase, the probability of divorce starts to increase. This family income where the probability of divorce turns from negative to positive is \$195,121[10]. Therefore, as far as our sample values go, the probability of divorce decreases at a decreasing rate with income (becomes less and less negative, becoming flatter). In other words, as family income increases, the probability of divorce decreases, but that decrease in the probability of divorce becomes less and less prominent. As family income increases, financial stress decreases, yielding a happier couple, but this effect becomes less and less pronounced.

The standardized marginal effects (shown in Table A in the Appendix) show that age and income are the most important factors in predicting divorce. A one standard deviation change in

---

[10] This critical income value is found using the same method in footnote 9 which dealt with age. We know that $(\beta_{11} + 2\beta_{12}A)xf(X\beta) = 0$, so $2\beta_{12}A = -\beta_{11}$, and thus $A = \frac{-\beta_{11}}{2\beta_{12}}$. Replacing our coefficients for $FamilyIncome$ (.112x10^{-4}) and $FamilyIncome^2$ (.287x10^{-10}) for β_{11} and β_{12} respectively, we find that our critical A is equal to 195,121.

age yields a .263 standard deviation change in the probability of divorce, while a one standard deviation change in income yields a -.042 standard deviation change in the probability of divorce.

Substance abuse, as a group, ranks third in the importance of predicting divorce. A one standard deviation change in each abuse variable, taken together, gives a .021 standard deviation change in the probability of divorce (about one-tenth of the effect of age and one-half of the effect of income, but more than any other variable). Among the substance abuse variables, illegal drugs appears to be the most important in predicting divorce, with legal drugs and alcohol consumption being of roughly equal importance, about half that of illegal drugs.

This chapter analyzed the effects of substance use, personal characteristics, family characteristics, and economic characteristics on the probability of divorce. Using a probit model for the reasons specified earlier, and analyzing these coefficients and their t-statistics, the marginal effects, and the standardized marginal effects, we find that substance abuse, education, race, age, family size, income, and wives' working are factors that predict divorce, with age, income, and substance abuse being the most important.

CHAPTER IV

SUBSTANCE USE

The previous chapter discussed the effects of divorce on productivity, and literature on the topic, as well as what factors predict divorce. This chapter will look at how substance use, another personal problem, can affect productivity. To better understand this relationship, we need to observe what factors affect substance use. Therefore, we will analyze literature on substance use and present and discuss three specific substance use models based on models from the literature: a regular drinking model, an illegal drug use model, and a legal drug misuse model. Predictions of how being divorced affects the probability of substance use, as well as other certain personal, family, and economic factors, will be included, and the results will be examined.

Substance use imposes major costs on society, with arguably the biggest being health costs. “Health costs from alcohol and drug abuse are the result of fatal and non-fatal accidents, especially on the highway and at work, liver cirrhosis, heart disease, various cancers and accidental overdoses” (Saffer and Chaloupka, 1998, p. 1). According to the 2011 CDC data, 43,544 people died from drug-induced deaths and 26,654 people died from alcohol-induced deaths. “In addition, there are costs due to poor birth outcomes and the physical, and emotional damage, caused to children by alcohol and drug abusing parents” (Saffer and Chaloupka, 1998, p. 1).

While the effects of alcohol abuse and illegal drug use may be the first health costs that come to mind, it is important to note that the health costs associated with misusing legal drugs are equally as important. “According to the Center for Disease Control, poisoning is the leading cause of accidental death in the United States and 40 percent of those deaths are caused by misuse of prescription pain killers (opioids). The National Institute on Drug Abuse and National Institute on Health both report that prescription drug abuse affects an estimated 20% of the American population” (Skinner, 2013, p. 2).

While the health costs of substance use are extremely high, other costs such as employment problems and crime also cause considerable concern. “Crime costs from alcohol and drug abuse are due to psychological effects on users, the need to generate income to buy drugs and to [sic] lesser extent to buy alcohol, and result from the extra-legal nature of illicit drugs transactions” (Saffer and Chaloupka 1998, p. 1). “Employment costs from alcohol and drug abuse are lost income due to reduced productivity, increased unemployment and absenteeism, and changes in career choice” (Saffer and Chaloupka 1998, p.1).

“Risks stemming from drug use, such as injury from accidents or fights, family discord, and conflicts with law enforcement can also limit the capability of an individual to maintain work attendance” (Skinner). While the immediate effects of a moderate amount of alcohol consumption do not appear to be as dangerous as other substances, it still impairs judgment and can have lingering effects the next day. These lingering effects can lead to work absences or decreased productivity at work. Also, assuming the consumption of alcohol and other substances is done during leisure time, an increase in this consumption would be increasing leisure time and decreasing the time that could have been productive.

The literature on alcohol and drug use is extensive. Most of the alcohol studies analyze the price elasticity of alcohol and find that alcohol consumption decreases as the price of alcohol rises. Leung and Phelps (1993) reviews 21 of these studies and finds that most estimate the price elasticities for beer, wine, and spirits separately and they usually do so using aggregate (as opposed to individual) data. "Studies using aggregate data find price elasticities for beer from about -.2 to -1.0, for wine from about -.3 to about -1.8 and for spirits from about -.3 to about -1.8. Studies using individual data estimate price elasticities for beer from about -.5 to about -3.0, for wine at about -.5, and for spirits from about -.5 to about -4.0" (Chaloupka & Saffer 1998, p. 3).

While Kendell, de Roumanie, and Ritson (1983) and Kenkel (1993) estimate total alcohol price elasticities for men and women separately, Grossman, Coate, and Arluck (1987) focus on beer and spirits price elasticities for youths. The results of these studies suggest that women and youth may have a more elastic demand for alcohol than the general population. Heinen and Pompelli (1988) and Kenkel (1993) are two studies that include more right hand side demographic variables in analyzing alcohol use, including both race and marital status.

Much of the literature on drug use also analyzes price elasticities. Van Ours (1995) uses data from Indonesia in the Dutch colonial period to find the price elasticity of opium use and participation. DiNardo (1993) uses a state aggregated version of the Monitoring the Future data set and finds no effect of price on cocaine use by high school seniors. Crime studies such as Silverman and Spruill (1977) and Brown and Silverman (1974) suggest an inelastic demand for heroin. While Silverman and Spruill (1977) use pooled cross-sectional time-series data from 41 Detroit neighborhoods from 1970-1973, Brown and Silverman (1974) use data from New York City. Nisbet and Vakil (1972) estimate marijuana price elasticity using data on UCLA students.

Saffer and Chaloupka (1998) look at marijuana decriminalization, heroin prices, and cocaine prices to obtain elasticities for marijuana, heroin, and cocaine. This use of marijuana decriminalization is common in predicting marijuana use, because it has an effect similar to cutting the price. While Pacula (1994) and Thies and Register (1993) use NLSY data, DiNardo and Lemieux (1992), and Johnston, O'Malley, and Bachman (1981) use the Monitoring the Future surveys. None of these studies find an effect on marijuana use.

Other effects of marijuana decriminalization are analyzed in the literature as well, such as Model (1992) who uses state aggregate data from the Uniform Crime Reports to find a positive effect on property crimes and a negative effect on violent crimes. In a later paper, Model (1993) uses Drug Abuse Warning Network data to find that decriminalization increases marijuana related emergency room visits but decreases all other emergency room visits.

Some drug studies include more demographic variables to predict drug use. DiNardo and Lemieux (1992) use the MTF surveys and Pacula (1994) uses the NLSY to analyze the effect of race on illegal drug use. Using data from the NLSY, Theis and Register (1993) include marital status as an extra explanatory variable. Sickles and Taubman (1991), also using NLSY data, include gender, race, age, education, and religious participation. Chaloupka and Saffer (1998) estimate regressions predicting alcohol and drug participation (specifically, marijuana, cocaine, and heroin participation). Alcohol participation is a variable measuring the number of days the respondent consumed alcohol in the past 30 days. Drug participation variables are dummy variables equal to one if the respondent has used the substance in the past year. Independent variables included price variables for each drug, a continuous total personal income variable, age, race, ethnicity, gender, and marital status. The alcohol equation is estimated using OLS, but the drug equations are estimated using probit due to the nature of the drug variables.

The last major section of the drug and alcohol literature focuses on cross price effects between alcohol and illicit drugs. While three of the major studies find alcohol and drugs to be substitutes, one study finds evidence of them being complements. DiNardo and Lemieux (1992) use the MTF surveys to analyze the effects of marijuana decriminalization, the drinking age, alcohol price, and race on alcohol and marijuana participation, as well as the cross price effects between alcohol and marijuana. They find that decriminalization had a significant negative effect on alcohol participation, but had no effect on marijuana participation. They also find that the drinking age had a significant positive effect on marijuana participation and a negative effect on alcohol participation. These results suggest that alcohol and marijuana are substitutes. This concurs with Chaloupka and Laixuthai (1994) who also use the MTF surveys to find alcohol and marijuana to be substitutes using marijuana price, marijuana decriminalization, and beer taxes. Similarly, Pacula (1998) uses NLSY data and includes variables for decriminalization, beer taxes, and the legal drinking age to reach this same substitution conclusion. Thies and Register (1993), on the other hand, use the NLSY to find that decriminalization increases alcohol and cocaine use, but doesn't affect heavy alcohol use or marijuana use, which would suggest that alcohol and cocaine are complements to marijuana.

The sample employed here for these three substance use/abuse models are cross-sectional data from the Center of Disease Control's 2006 National Survey of Drug Use and Health. The survey was given to 55,279 individuals, but the sample for the first two models consists of 36,935 individuals, while the sample for the third model consists of 36,813 individuals. Respondents under the age of 18 were excluded, and any individuals who did not answer pertinent survey questions were skipped.

We propose three equations following this general substance use/abuse model:

$$P(S) = f(D, P, F, E) \tag{4.1}$$

implying that $P(S)$ (the probability of using and or abusing substances) is a function of D (if an individual is divorced), P (personal characteristics), F (family characteristics), and E (economic characteristics).

Our substance use/abuse models are estimated using probit rather than ordinary least squares for the following reason. Suppose we wish to analyze the propensity for or preference for substance use. If this preference (S^*) was observable, it would presumably be normally distributed about some mean $E(S^*)$ and be determined by several exogenous variables X_j $(j = 1, \ldots, k)$. Then if we wished to model the behavior, we could set up a regression model $S^* = X\beta + \epsilon$ and estimate it with ordinary least squares. Unfortunately, S^* is not observable. All we can observe is whether an individual has used or abused a substance ($S = 1$) or not ($S = 0$). If S^* conforms to the usual preference axioms, there is some critical value of S^*, say μ_1, such that if the preference for substance use exceeds that value, people will use substances ($S^* > \mu_1$ implies $S = 1$) and if the preference is less than that value people will not use substances ($S^* < \mu_1$ implies $S = 0$). Thus, we have a normally distributed (but unobservable) substance preference centered on its mean $X\beta$ that is divided into two observable categories at μ_1, such that:

$$S^* \sim N\left(X_i\underline{\beta}, \sigma^2\right) \forall\, i \tag{4.2}$$

It follows that the probability that the ith individual will not engage in substance use is:

$$P(S^*{}_i < \mu_1) = F\left[\frac{\mu_1 - X_i\underline{\beta}}{\sigma}\right] - F\left[\frac{\mu_0 - X_i\underline{\beta}}{\sigma}\right] \tag{4.3}$$

where F(x) is the (less than or equal to) cumulative distribution function for the standard normal distribution evaluated at x. It is conventional to set μ_0 = -∞, μ_1 = 0, μ_2 = +∞, and σ =1, so that the probability of not engaging in substance use is:

$$P(S^*_i < 0) = P(S_i = 0) = F\left(-X_i\underline{\beta}\right) - F(-\infty) = 1 - F\left(X_i\underline{\beta}\right) \quad (4.4)$$

and similarly the probability that the i[th] individual has used substances is:

$$P(S^*_i > 0) = P(S_i = 1) = F(+\infty) - F\left(-X_i\underline{\beta}\right) = 1 - \left[1 - F\left(X_i\underline{\beta}\right)\right] = F(X_i\underline{\beta}) \quad (4.5)$$

Using these results we can write the likelihood function of the sample of n observations as:

$$L^* = \prod_{i=1}^{n}(1 - F(X_i\underline{\beta})^{1-S_i}\,(F(X_i\underline{\beta})^{S_i} \quad (4.6)$$

where S_i = 1 if they have used substances and = 0 otherwise. This, in turn, yields the log likelihood function:

$$L = \ln L^* = \sum_{i=1}^{n}(1 - S_i)\ln(1 - F(X_i\underline{\beta})) + S_i\ln(F\left(X_i\underline{\beta}\right)) \quad (4.7)$$

This function can be differentiated with respect to the β_j's ($j = 1, \ldots, k$) to obtain the first order conditions for maximization which consists of k equations in k unknowns (the β_j) that are nonlinear in the β_j's. These can be solved by iterative nonlinear techniques to obtain the b_j's (the maximum likelihood estimates of the β_j's). Because they are maximum likelihood estimates, the b_j's are known to be asymptotically normally distributed with mean vector $\underline{\beta}$ and a covariance matrix given by the information matrix (the negative of the inverse of the matrix of second partials of the likelihood function). This result allows us to conduct all of the typical hypothesis tests on the β_j's with asymptotic validity.

However, the b_j's do not have the same interpretation as their ordinary least squares counterparts. To see this, note that the behavioral content of our model lies in determining the change in the probability of substance use caused by a change in, say, X_j. From above, the probability of substance use for the i[th] individual is:

$$P(S*_i > 0) = P(S_i = 1) = F(X_i\underline{\beta}) \quad (4.8)$$

As a result:

$$\frac{\partial P(S_i=1)}{\partial X_j} = \frac{\partial F(X_i\beta)}{\partial X_j} = f(X_i\beta)\frac{\partial X_i\beta}{\partial X_j} = \beta_j f(X_i\beta) \qquad (4.9)$$

which is consistently estimated by $b_j f(X_i \underline{b})$. Because $f(X_i \underline{b})$ is an ordinate of the standard normal, it is always positive. Thus the sign and significance of the partial derivatives is determined by the sign and significance of b_j. Determining the magnitude of the partial derivatives requires taking the sum of the product of, say, the sample means of the X_j's with their corresponding by probit coefficients (the b_j's), multiplying that sum by the relevant b_j, referring the result to the standard normal ordinate table, and taking the product of that ordinate with the relevant b_j.

Referring back to our behavioral model, our subsequent analysis will employ the following variables to measure the fundamental variables S, D, P, F, and E. $P(S)$ will be $P(RegularDrinker)$ in our first equation, $P(LegalDrugMisuse)$ in our second equation, and $P(IllegalDrugUse)$ in our third equation. D (whether an individual is divorced) will be represented by $Divorce$ in each of the three equations. P (the personal characteristics we believe will predict substance use) will include $CollegeGraduate$, Age and Age^2, and $White$.

There are also three other personal characteristics that are specific to a particular substance use equation. The other personal characteristic is $Smoker$ in our regular drinking equation, $PsychologicalDistress$ in our legal drug misuse equation, and $Risky$ in our illegal drug use equation.

F, the family characteristic we believe will affect substance use, is $FamilySize$. E, the economic characteristics we believe will predict substance use, include $Fulltime$, $NonMSA$, $FamilyIncome$, and $FamilyIncome^2$.

Therefore, our three individual substance use models are as follows:

$P(RegularDrinker) =$ (4.10)

$$\alpha + \beta_1 EverMisused\ LegalDrugs + \beta_2 EverUsedIllegalDrugs + \beta_3 Divorce$$
$$+ \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White$$
$$+ \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize$$
$$+ \beta_{11} FamilyIncome + \beta_{12} FamilyIncome^2 + \beta_{13} Smoker + \epsilon$$

$$P(LegalDrugMisuse) = \tag{4.11}$$
$$\alpha + \beta_1 RegularDrinker + \beta_2 EverUsedIllegalDrugs + \beta_3 Divorce$$
$$+ \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White$$
$$+ \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize$$
$$+ \beta_{11} FamilyIncome + \beta_{12} FamilyIncome^2$$
$$+ \beta_{13} PsychologicalDistress + \epsilon$$

$$P(IllegalDrugUse) = \tag{4.12}$$
$$\alpha + \beta_1 RegularDrinker + \beta_2 EverMisused\ LegalDrugs + \beta_3 Divorce$$
$$+ \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White$$
$$+ \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize$$
$$+ \beta_{11} FamilyIncome + \beta_{12} FamilyIncome^2 + \beta_{13} Risky + \epsilon$$

The dependent variable in our first substance use model, $P(RegularDrinker)$, is a dummy variable equal to one if the respondent answered the question "When is the last time you used alcohol?" with "within the past 30 days". In other words, we are modeling the probability of being a regular drinker. Our sample includes 21,556 regular drinkers.

The dependent variable in our second substance use model, $P(LegalDrugMisuse)$, is a dummy variable equal to one if the respondent has ever used painkillers, tranquilizers, stimulants, or sedatives non-medically. Specifically, they were asked, "Have you ever, even once, used any type of prescription [drug type] that was not prescribed for you or that you took only for the experience or feeling it caused?" The sample for our second substance use model includes 9,497 respondents who have misused legal drugs.

The dependent variable in our third substance use model, $P(IllegalDrugUse)$, is a dummy variable equal to one if the respondent ever used marijuana, cocaine, crack, heroine, or hallucinogens. Hallucinogens include LSD, PCP, peyote, mescaline, mushrooms, ecstasy, or any other listed hallucinogens. The sample for our third substance use model includes 18,517 respondents who have ever used illegal drugs.

The correlation between these three substance use variables is likely to be positive because substances are frequently complements with other substances. A great portion of the literature observes the relationship between marijuana and alcohol use. A summary of the biomedical literature in NIAAA (1993) suggests that alcohol use and illicit drug use may be mutually reinforcing. That is, the use of two drugs taken together produces an effect greater than the sum of the effects of each drug taken individually" (Chaloupka and Saffer 1998, p. 6). While Dinardo and Lemieux (1992), Chaloupka and Laixuthai (1994), and Pacula (1998) find alcohol and marijuana to be substitutes, Thies and Register (1993) find results that suggest alcohol and cocaine are complements to marijuana. Our models allow us to test this hypothesis. If the three types of substances are complements to one another, then all substance variables should be positive and significant in each of the three models.

$Divorce$, a dummy variable equal to one if the individual's marital status was divorced or separated at the time of the survey, is expected to be positive in all three models but the causality can flow both ways. It could be that individuals who regularly drink (or use and abuse other substances) are more likely to get divorced (which we modeled in the last chapter) but also that individuals who are divorced might start regularly drinking (or using/abuse other substances) after the fact. Getting divorced is a stressful experience, and this stress can lead to increases in the use of substances. In this chapter, we model this second scenario of whether or not being divorced predicts substance use. Because we expect this stressful experience of a divorce to increase substance use, we predict the coefficient on $Divorce$ to be positive and significant in all three models.

$CollegeGraduate$, a dummy variable equal to one if the respondent has completed college, is expected to be negative with respect to all substance variables. Individuals with a higher level of education are expected to understand the caution and risks of substance use, and would therefore be less likely to be use and/or abuse substances than non-graduates. However, because such a huge proportion of the population drink alcohol, say, a glass of wine at dinner, and because the cautions and risks associated with "a glass of wine at dinner" are so low, the coefficients in the regular drinker equation may not behave as expected. The coefficients in the legal drug misuse and illegal drug use equations, on the other hand, are expected to both be negative.

The data on the individual's age is categorical: 12-17 years old, 18-25 years old, 26-34 years old, 35-49 years old, 50-64 years old, 65 years or older. Midpoints for each age category were created, as well as a midpoint for the open ended interval (65 years plus)[11] which is equal to

[11] The midpoint for the open ended interval is calculated by running a grouped data regression where the left-hand side variable is the categorical age variable, the explanatory variable is the constant term only, and the interior limits are 18, 26, 35, 60, and 65. Estimating this model yields a mean (23.98) and variance (19.51) of the distribution. A z value that corresponds to the lower limit of the highest ranked (open-ended category) is estimated with the following equation:

$$Z = \frac{OEI_{LL} - Mean}{Standard\ Deviation}$$

Where OEI_{LL} represents the open-ended interval lower limit. Using our computed values, we find a z value equal to 2.5.We then use the censored normal results to find the mean of the category in terms of z using the following equation:

$$E[z] = \frac{\theta(z)}{1 - \Phi(z)} = x$$

72 (this was calculated by the method used in footnote 6). Age^2 are the midpoints squared. Because the sample consists of individuals 18 and older, we would expect the probability of being a regular drinker to increase as age increases due to the legal minimum drinking age of 21. Many studies, however, show that individuals below the drinking age actually binge drink when they consume alcohol. The probability of misusing legal drugs may increase with age because as individuals age, the number of medications that are prescribed to them typically increases. The probability of using illegal drugs may decrease with age because, as individuals age, they may "grow out of" the excitement of participating in the illegal drug market.

$White$ is a dummy variable equal to one if the individual's race is non-Hispanic white, equal to zero if their race is non-Hispanic black/African American, non-Hispanic Native American/Alaskan Native, non-Hispanic Native Hawaiian/Other Pacific Island, non-Hispanic Asian, or Hispanic. We want to investigate if there is something ethnic or racial in predicting substance use. In other words, are there certain characteristics exhibited in certain races or ethnic groups that would increase their probability of being a regular drinker, their probability of misusing legal drugs, or their probability of using illegal drugs? For example, if we assume that most Hispanics are religious, and to postulate that religious individuals tend not to be regular drinkers (which could be argued), then we would expect Hispanics as a group to have a lower probability of being a regular drinker. Because our model does not control for religion, it may well be that a race indicator picks up a lot of that effect.

---

where θ represents the standard normal density and Φ represents the standard normal distribution. Finally, we back out the midpoint value on the original data's scale by "unnormalizing" the result above,

$$\frac{M - Mean}{Standard\ Deviation} = x \rightarrow M = Mean + (x * Standard\ Deviation)$$

where M is the desired estimate of the category midpoint and is equal to 72.

$Fulltime$ is a dummy variable equal to one if the respondent's employment status is "employed full time". Assuming that substance use is done during leisure time, an increase in work time would mean a reduction in leisure time, therefore a reduction in substance use. On the other hand, if an individual's job is very stressful, they may look to substances to relax, increasing their drinking or drug use. Therefore, the sign on this coefficient is theoretically ambiguous.

$NonMSA$ is a dummy variable equal to one if the respondent lives in a segment not in a "Core Based Statistical Area" (CBSA) according to the 2000 Census data and the June 2003 CBSA classifications provided by the Office of Management and Budget. We predict this coefficient to be negative in all three models due to market availability. Being in a city means the number of bars nearby is greater, and dry counties are mostly rural, so it may be that there is a negative relationship between those living in a non MSA region and the probability of being a regular drinker. The same thought process may be applied to illegal drug use as well. Being in a city may mean an increase in the number of drug dealers, as well as the number of drug options available. In other words, there is a heavier illegal drug market in the city implying a negative relationship between non MSA regions are the probability of using illegal drugs. The number of doctors and pharmacies are greater in a city and less available in rural areas, and the cost of obtaining legal drugs is higher in rural areas. This again leads us to predict that legal drug misuse may be less prominent in non MSA regions.

The coefficient on $FamilySize$, a variable representing the number of persons in the respondent's household, is expected to be negative for all three models. As the number of children increases, the responsibilities associated with raising those children increase. Assuming the parents are rational and do increase the time spent child-rearing, their leisure time must therefore decrease. Because substance use ought to be restricted to leisure time, their probability of using substances may decrease.

The NSDUH's data on income and employment is very limited. All income variables (whether family income or respondent's income) are categorical in nature. More on the chosen income variable ($FamilyIncome$) will be discussed in the following chapter. The respondents were asked which of the following categories contained their total family income: Option 1 being less than $20,000, option 2 being between $20,000 and $49,999, option 3 being between $50,000 and $74,999, and option 4 being greater than $75,000. Midpoints of each family income category were calculated ($10,000, $34,999.50, and $62,499.50) as well as the midpoint of the open ended interval ($106,152) using the method discussed in the sixth footnote. The midpoints squared yield the $FamilyIncome^2$ variable.

How $FamilyIncome$ predicts substance use is a question of whether alcohol, illegal drugs, and legal drugs are normal goods (goods for which, as income increases, demand for that good increases). We believe all three categories of substances to be normal goods, therefore we expect individuals with higher incomes are more likely to use these substances. In other words, we predict $FamilyIncome$ to be positive in all three models. The coefficient on $FamilyIncome^2$ will reveal if there is a change in the probability of using substances as income continues to increase.

$Smoker$, a dummy variable only included in the regular drinking model, is equal to one if the respondent smokes more than 6 to 15 cigarettes (about ½ a pack) a day on average. Cigarettes and alcohol are considered to be complements, so this coefficient is expected to be positive.

$PsychologicalDistress$, a dummy variable only included in the legal drug misuse model, was created from a "serious psychological distress indicator" coded by the CDC. Survey respondent's aged 18 or older were asked a series of questions about their psychological distress during the one month in the past year when they were at their worst emotionally (including feeling nervous, hopeless, restless or fidgety, sad or depressed, etc.) and given a score from 0 to 24. The respondents were then given a 0 (no serious psychological distress indicator) if they scored below a 13, and given a 1 (yes to serious psychological distress indicator) if they scored a 13 or above. This coefficient is expected to be positive, because individuals who are psychologically distressed are more likely to have prescription drugs that are commonly abused.

$Risky$, a dummy variable only included in the illegal drug use model, was created from the survey question, "How often do you get a real kick out of doing things that are a little dangerous?" If the respondent answered "sometimes" or "always", they were given a 1 indicating they are risk takers. If the respondent answered "never" or "seldom", they were given a 0 indicating they are risk averse. This coefficient is expected to be positive, because risky individuals are less afraid of the consequences of buying and using illegal drugs.

Because the model is estimated using probit, the sign of the coefficients yield a slightly different interpretation from coefficients estimated using linear regression. The probit coefficients only provide us information on the sign and significance of the variables at hand. Therefore, the marginal effects and standardized marginal effects of each variable are calculated as well. As mentioned earlier, the marginal effects are calculated by multiplying $f(\bar{z})$ (the ordinate of the standard normal evaluated at the sum of the product of the probit coefficients (the b_j's) and their corresponding sample means) by the appropriate probit coefficients. The standardized marginal effects are calculated by multiplying the $f(\bar{z})$ by the standardized probit coefficients. The standardized probit coefficients are calculated by multiplying the probit coefficients (the b_j's) by the ratio of the standard deviation of the relevant explanatory variable (X_j) to the standard deviation of Y (which in the probit model is standardized to one). The marginal effects tell us how the probability of substance use changes given a one unit change in X_j. The standardized marginal effects tell us how many standard deviations in the probability of substance use change given a one standard deviation change in X_j. These standardized marginal effects are useful for seeing which variables have the biggest effects on the probability of substance use[12].

Table 2, Table 3, and Table 4 provide the probit coefficients, t-statistics, and marginal effects for modeling the probability of being a regular drinker, the probability of misusing legal drugs, and the probability of using illegal drugs, respectively. In all three probits, the coefficients on the other substance use variables are positive and significant, showing that substances are complements. This result that alcohol, illegal drugs, and legal drugs are used as complements to each other concurs with Thies and Register's (1993) finding that alcohol and cocaine are complements to marijuana.

For example, in our Regular Drinking results, the probit coefficients on $EverMisusedLegalDrugs$ and $EverUsedIllegalDrugs$ are positive and significant as expected. In other words, individuals who have misused legal drugs or used illegal drugs are more likely to be regular drinkers. Specifically, individuals who have misused legal drugs are 7.1% more likely, and those who have used illegal drugs are 27.2% more likely, to be regular drinkers.

---

[12] The standardized probit coefficients and standardized marginal effects are reported in Tables B, C, and D in the Appendix.

The legal drug misuse probit yields the same interpretation: people who have used illegal drugs and are regular drinkers are more likely to misuse legal drugs. Specifically, individuals who have used illegal drugs are 29.3% more likely, and those who are regular drinkers are 5.6% more likely, to misuse legal drugs.

Finally, the illegal drug use probit follows suit in showing that people who have misused legal drugs and are regular drinkers are more likely to use illegal drugs. Specifically, individuals who are regular drinkers are 26.9% more likely, and those who have misused legal drugs are 39% more likely, to use illegal drugs.

In all three substance use models, the coefficient on $Divorce$ is positive, but it is not significant in the regular drinking equation. This insignificance in the regular drinking equation may be due to two countervailing effects of divorce on drinking. People may drink more to cope with the psychological strain of divorce, implying a positive effect of divorce on regular drinking. On the other hand, if drinking contributed to the divorce, it may be the case that they may cut back on alcohol consumption to address this personal problem (implying a negative effect of divorce on regular drinking).

We do find, however, that being divorced is a significant predictor of misusing legal drugs and using illegal drugs. Specifically, individuals who are divorced are 2.9% more likely to misuse legal drugs and 9.6% more likely to use illegal drugs. This somewhat concurs with Theis and Register (1993) who find that marriage has a negative effect on drug use. The lack of an effect of divorce on regular drinking may be due to a gender component that our model does not take into account. Kenkel (1993) finds that being a divorced woman has a negative effect on heavy drinking while being a divorced male increases heavy drinking. Heinen and Pompelli

(1988) find that being single increases alcohol consumption. Chaloupka and Saffer (1998) find that married individuals consume less alcohol, less marijuana, less cocaine, and less heroine.

$CollegeGraduate$ is positive and significant in the Regular Drinker model, and negative and significant in the other two models, implying that individuals with a college education are more likely to be regular drinkers, but less likely to misuse legal drugs and use illegal drugs. Specifically, individuals who are college graduates are 13.3% more likely to be regular drinkers, 3.3% less likely to misuse legal drugs, and 6.8% less likely to use illegal drugs. Sickles and Taubman (1991) also find that use of drugs decreases with education. The positive relationship between college education and regular drinking may be due to some habit formation in college, requirements of business networking, or possibly socialization.

In the Regular Drinker model, the coefficients on Age and Age^2 are significantly negative and positive, respectively, so that they yield a U-shaped relationship between age and the probability of being a regular drinker. In other words, as an individual gets older, the probability that they are a regular drinker decreases, but at some point as they continue to age, they become more likely to be a regular drinker. The estimated age where the probability of being a regular drinker changes sign is 69 years old[13]. This implies that, as far as our sample values go, the probability of being a regular drinker decreases at a decreasing rate with age (becoming less and less negative, or becoming flatter). This agrees with Chaloupka and Saffer (1998) who obtained a significantly negative relationship between their "Youth" variable and

---

[13] We can find this critical value using the fact that $(\beta_5 + 2\beta_6 A)xf(X\beta) = 0$. Since f(Xβ) must be positive, the critical value of A can be found by setting $\beta_5 + 2\beta_6 A = 0$, meaning that $2\beta_6 A = -\beta_5$, and therefore $A = \frac{-\beta_5}{2\beta_6}$. Replacing our coefficients for Age (-.028) and Age^2 (.002x10^{-1}) for β_5 and β_6 respectively, we find that our critical A is equal to 69.25.

alcohol use, and somewhat with Sickles and Taubman's (1991) analysis, which found a negative effect of age on illicit drug use.

The coefficients on Age and Age^2 in the Legal Drug Misuse model are insignificantly positive and significantly negative, respectively. The critical age value we calculate (with the method just discussed in footnote 13) is 9 years old[14], implying that, as far as our sample values go, the probability of misusing legal drugs decreases as age increases.

The coefficients on Age and Age^2 in the Illegal Drug Use model (significantly positive and negative, respectively) yields an upside-down U-shaped relationship between age and the probability of using illegal drugs. As an individual's age increases, their probability of using illegal drugs increases. Once some age is reached, however, an increase in the individual's age leads to a decrease in their probability of using illegal drugs. This age where the probability changes sign is estimated to be 37 years old[15]. This implies that as an individual's age increases, their probability of illegal drug use increases until they reach age 37. Once an age of 37 is reached, as an individual's age increases, their probability of illegal drug use decreases Johnson et al. (1988) also find that the use of some drugs decreases at older ages, say 35 years old. They note that they are unsure whether the users have gone through a rehabilitation program, dropped out of the sample, or died. Kandel (1980), however, finds that individuals in their late 30's don't die, but rather "mature out" of heroin use.

The question of any ethnic differences in the probability of substance use is answered with a positive and significant coefficient on $White$ in all three models. This implies that individuals who are white are more likely to use and abuse substances. In other words, there is

---

[14] We can find this critical value using the same method discussed in footnote 13, but replacing our coefficients for Age (.003) and Age^2 ($-.16x10^{-3}$) for β_5 and β_6 respectively, we find that our critical A is equal to 9.47.

[15] We can find this critical value using the same method discussed in footnote 13, but replacing our coefficients for Age (.074) and Age^2 (-.001) for β_5 and β_6 respectively, we find that our critical A is equal to 37.

some ethnic or racial trait of Caucasians (that is not controlled for by the other exogenous variables) that predicts them to be 9.4% more likely to be a regular drinker, 8% more likely to misuse legal drugs, and 6.7% more likely to use illegal drugs (with all percentages being told by the marginal effects). This disagrees with the result of several papers including Sickles and Taubman (1991) who find that blacks and Hispanics are more likely to consume illegal drugs, and Saffer and Chaloupka (1998) who find that find that racial and ethnic minorities are more likely to consume cocaine than the total population. Our results do, however, reflect Saffer and Chaloupka's (1998) result that racial and ethnic minorities are less likely or equally likely to consume alcohol, marijuana and heroin. DiNardo and Lemieux (1992) and Pacula (1998) also agree in concluding that minorities are less likely or equally likely to consume illegal drugs as whites.

$Fulltime$ is positive and significant in all three models. In other words, individuals who work full time are 8% more likely to drink regularly, 0.7% more likely to misuse legal drugs, and 3% more likely to use illegal drugs than individuals who work part time, or are unemployed or not in the labor force. Perhaps working a fulltime job yields a larger desire to unwind and relax, and this unwinding may tend to involve substance use.

The coefficient on our urban/rural indicator ($NonMSA$) is negative in all three models, but insignificant in the Illegal Drug Use model. This implies that individuals who live in a city are 5.9% more likely to be regular drinkers and 3% more likely to misuse legal drugs than individuals who live in a rural area. This may be due to the proximity of bars and restaurants, as well as pharmacies and doctors, as discussed earlier. This variable may also serve as a proxy for the role of religion in their lives.

$FamilySize$ is negative and significant in all three models, as predicted, yet insignificant in the legal drug misuse model. Individuals with larger families are 3.8% less likely to be regular drinkers and 0.9% less likely to use illegal drugs. Assuming "regular" drinking and drug using is done during leisure time, an increase in the number of children would lead to a reduction in leisure time.

Our regular drinking results show $FamilyIncome$ is positive and significant while $FamilyIncome^2$ is negative but insignificant, indicating that the probability of being a regular drinker increases as a family's income increases up until some point. This family income where the probability of being a regular drinker stops increasing is estimated to be $318,659[16]. Therefore, as far as our sample values go, the probability of being a regular drinker increases with income. This may be because alcohol is a normal good, meaning an increase in income yields an increase in the demand for this good (alcohol). Because the linear term dominates, marginal effects show that as family income increases by $10,000, the probability of being a regular drinker increases by 1.88%. This results concurs with Chaloupka and Saffer's (1998) finding that income has a positive and significant effect on alcohol use for all groups except blacks.

Our legal drug misuse results show $FamilyIncome$ is negative and significant while $FamilyIncome^2$ is positive but insignificant, indicating that the probability of misusing legal drugs decreases as a family's income increases. Therefore, as far as our sample values go, the probability of misusing legal drugs decreases with income throughout the range of our sample. It

[16] We can find this critical value using the fact that $(\beta_{11} + 2\beta_{12}A)xf(X\beta) = 0$. Since f(Xβ) must be positive, the critical value of A can be found by setting $\beta_{11} + 2\beta_{12}A$ =0, meaning that $2\beta_{12}A = -\beta_{11}$, and therefore $A = \frac{-\beta_{11}}{2\beta_{12}}$. Replacing our coefficients for $FamilyIncome$ (.485x10^{-5}) and $FamilyIncome^2$ (-761x10^{-11}) for β_{11} and β_{12} respectively, we find that our critical A is equal to 318,659.66.

may be the case the legal drugs are substitutes for the more expensive illegal drugs among the less well off. Marginal effects show that as family income increases by $10,000, the probability of misusing legal drugs decreases by 0.54%.

Our illegal drug use results show $FamilyIncome$ and $FamilyIncome^2$ are significantly negative and positive, respectively. This indicates that the probability of using illegal drugs decreases as a family's income increases up until some point, and then increases as family income continues to increase. This family income where the probability of using illegal drugs switches signs is estimated to be $53,152[17]. Therefore, the probability of using illegal drugs decreases as family income increases, until family income reaches $53,152. Then, as family income increases, the probability of using illegal drugs increases.

$Smoker$ is positive and significant as expected, implying that smokers are more likely to be regular drinkers. Specifically, individuals who smoke are 4.4% more likely to be regular drinkers. This result supports the assumption that alcohol and cigarettes are complements.

$PsychologicalDistress$ is positive and significant, implying that individuals who have psychological distress are more likely to misuse legal drugs. Specifically, psychologically distressed individuals are 11.9% more likely to misuse legal drugs.

$Risky$ is positive and significant, indicating that risk taking individuals are more likely to use illegal drugs. Specifically, risk taking individuals are 12.6% more likely to use illegal drugs.

The standardized marginal effects (see Appendix Tables B, C, and D) show that other substance use and age are the most important factors in predicting regular drinking and illegal

---

[17] We can find this critical value using the same method discussed in footnote 13, but replacing our coefficients for $FamilyIncome$(-.289x10^{-5}) and $FamilyIncome^2$ (.271x10^{-10}) for β_{11} and β_{12} respectively, we find that our critical A is equal to 53,152.44.

drug use. While other substance use is the most important factor in predicting legal drug misuse, age is not as important.

Among these two substance abuse variables, illegal drug use appears to be the most important in predicting regular drinking. A one standard deviation change in each abuse variable, taken together, gives a .17 standard deviation change in the probability of being a regular drinker. A one standard deviation change in age yields a .163 standard deviation change in the probability of being a regular drinker. While other substance use is the most important factor in predicting legal drug misuse, age is not as important.

Age is the most important factor in predicting illegal drug use. A one standard deviation change in age yields a .436 standard deviation change in the probability of using illegal drugs. Other substance use, as a group, ranks second in the important of predicting illegal drug use. A one standard deviation change in each substance use variable, taken together, gives a .31 standard deviation in the probability of using illegal drugs. Among the two other substance variables, legal drug misuse appears to be more important in predicting illegal drug use, with regular drinking being almost as important.

Other substance use, as a group, ranks first in importance of predicting legal drug misuse. A one standard deviation change in each substance use variable, taken together, gives a .17 standard deviation change in the probability of misusing legal drugs. Among the two other substance variables, illegal drug use appears to be more important in predicting legal drug misuse, and regular drinking less important. While other substance use is the most important factor in predicting legal drug misuse, age is not as important.

This chapter analyzed the effects of being divorced, using other substances, personal characteristics, family characteristics, and economic characteristics on the probability of substance use. Specifically, we looked at how these characteristics affect the probability of being a regular drinker, the probability of misusing legal drugs, and the probability of using illegal drugs. Using probit models for the reasons specified earlier, and analyzing these coefficients and their t-statistics, the marginal effects, and the standardized marginal effects, we find that using other substances is the most important factor in predicting substance use.

CHAPTER V

INCOME

The previous chapter discussed the productivity costs of substance use, as well as what factors predict substance use. This chapter will dive into one of the most frequented topics in economics: income. We will briefly summarize some of the literature on income, then present and discuss an income model based on models from the literature. Predictions of what personal, family, and economic characteristics determine income will be discussed, as well as our results.

The literature on the determinants of differences in wages dates all the way back to Adam Smith's *The Wealth of Nations*. Smith determined that competitive factors, differences in individual abilities, and institutional factors constituted the major differences in wages.

The early quantitative analysis on differences in wages focused on differences in wages by occupation and industry. Douglas (1930) analyzed the change in wages of both white collar and blue collar workers in the United States from 1890 to 1926. Slichter (1950) emphasized the difference in wages between industries, and believed it to be attributable to skill differences and "company wage policies".

The work of pioneers such as Becker (1964, 1975) and Mincer (1974) about investment in human capital came later in the 1960s and 1970s. This shifted the focus to the effect of education, age, and experience on earnings.

Mincer (1974) developed the standard human capital earnings function which is of the form:

$$\ln y = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + u \tag{5.1}$$

where s represents schooling level and x represents work experience (Willis 1986). The rate of return on schooling, measured by β_1, is assumed to be constant for all schooling levels. The coefficients on the quadratic experience terms, β_2 and β_3, are respectively positive and negative. In other words, earnings increase as experience increases, and then start to decrease. Because actual labor force experience wasn't recorded by early data sources, a transformation of the worker's age (age minus schooling minus 6) was used as a proxy. The term u is a mean zero stochastic disturbance.

Blundell et al. (1999) dive deeper into investment in human capital, specifically, how much education and training really affects the individual, the firm, and the economy. They note that human capital has three main components: the individual's early ability (whether acquired or innate), the knowledge and skills acquired through education, and the skills and expertise acquired through on-the-job training. The literature shows a significantly positive relationship between education and training and wages.

Blundell et al. point out that the most discussed issue is whether "the higher earnings that are observed for better-educated or more-trained workers are caused by their education or training, or whether individuals with greater earning capacity and ability choose to acquire more education or training" (Blundell et al. 1999, p. 3). This dilemma has come to be known as the human capital versus self-selection or screening controversy.

While Wise (1975) concludes that education contributes to productive ability, Lazear (1977) argues the screening hypothesis. This screening hypothesis "in its most basic form asserts that schooling acquisition costs differ across individuals according to their ability levels. If high ability individuals face lower marginal cost of schooling schedules than do low ability individuals, the former group will for a given return obtain more education. Employers will pay higher wages to the more educated because they recognize that ability and attained level of education are positively correlated as the result of differential costs" [Lazer (1977, p. 252)].

Jones and Jackson (1990) investigate this human capital versus self-selection topic using data from 811 individuals who received an undergraduate degree from the college of business at a large Southern university. Their dependent variable, annual earnings, is the natural logarithm of the individual's salary estimate for their salary class interval. Explanatory variables include cumulative grade point average, gender, sample characteristics, and human capital and job characteristics. These human capital characteristics include job tenure, experience, and whether the individual is a degree holder. Job characteristics included a SMSA dummy, whether residence is in the state of their university, firm size (greater than 100 employees = 1), and occupation. Regression analysis "provide[d] little support for screening and no strong evidence for rejecting a human capital interpretation" [Jackson and Jones (1990, p. 254)].

Schmidt and Zimmerman (1989) use data from West Germany to analyze the relationship between firm size and employee compensation, using the human capital theory of earnings as their framework. The hypothesis scrutinized in their paper is that "unmeasured labor quality contributes to the strong positive relationship between firm size and wages reported in the literature" [Schmidt & Zimmerman (1989, p. 8)]. Only full-time male workers between the ages of 18 and 65 were included. Besides firm size, the additional explanatory variables were personal and work characteristics. Personal characteristics believed to influence earnings include education, experience, number of jobs, whether they are unskilled or white-collar workers, job duration, marital status, and number of children living in the household. Work characteristics included in the model are firm size, the innovative activity of the firm, whether the firm allows time at home only on the weekends, whether they work on Sundays, the weather conditions, the pollution conditions, the noise level at work, and other work condition variables (dirty work, fast paced, monotonous, etc.). The dependent variable used is the natural logarithm of monthly earnings. Schmidt and Zimmerman found that, contrary to popular belief, controlling for other wage determinants still yields a persistent firm size premium.

Wage discrimination, another large portion of the literature, focuses on many different areas: discrimination between male and female, black and white, and union and nonunion workers. Blinder (1973) famously estimates reduced form and structural wage equations to analyze the difference in white-black and male-female wages. He notes that while it was well known that white wages are higher than black wages, and that male wages are higher than female wages, he aimed to answer how much of those differences are due to other factors such as education levels, higher paying occupations, etc. Blinder estimates two wage equations as follows:

$$Y_i^H = \beta_0^H + \sum_{i=1}^{n} \beta_j^H X_{ji}^H + u_i^H \tag{5.2}$$

$$Y_i^L = \beta_0^L + \sum_{i=1}^{n} \beta_j^L X_{ji}^L + u_i^L \tag{5.3}$$

where the H superscript denotes the high-wage group and the L superscript denotes the low-wage group. Computations using these equations yield a portion of the differential explained by the regression and a portion typically attributed to discrimination. The portion attributed to discrimination can be broken down further into the differential due to the differences in coefficients and the differences in average characteristics. Blinder is then able to pinpoint discrimination by seeing how the market evaluates different demographic groups who have the same traits.

Blinder drops individuals whose household heads are younger than 25, to try and eliminate household heads who were still acquiring education. He uses the natural logarithm of the wage rage as the dependent variable. He notes, "because income variables include many diverse sources of nonlabor income and because earnings variables confound differences in wage rates with differences in labor supply (which are themselves functions of wages), it is imperative that studies such as the present one use actual wage rates as the dependent variable" [Blinder (1973, p. 439)]. He also notes that "use of earnings data instead of wage rates can seriously bias estimates of the real rates of return to education. Suppose, for example, that wages depend on education and labor supply depends on wages. Then, estimating the effect of education on *earnings* will over- or underestimate the impact on *wage rates* according as the labor supply curve is normal or backward bending" [Blinder (1973, p. 439)].

Blinder estimates the following reduced-form wage equation using ordinary least squares:

$$\log w = F(B, Z) + v_1 \tag{5.4}$$

where B is a set of 13 family-background variables and Z is a set of other exogenous variables. He then estimates the following structural wage equation using ordinary least squares

$$\log w = f(Ed, Occ, J, M, V, T, Z) + u_1 \tag{5.5}$$

where Ed represents six educational dummy variables, Occ represents eight occupational dummy variables, J is a dummy variable for vocational training, M is a dummy variable for being a member of a union, V is a veteran dummy variable, T is a set of six dummy variables for job tenure, and Z (as mentioned earlier) is a set of other exogenous variables. This method of two wage-equations (a reduced-form and structural equation) is used because the education and occupational variables are endogenous and simultaneously determined. Blinder notes that ideally he would use two-stage least squares estimation, but his wage equation is underidentified. We will examine this simultaneity issue in depth in the following chapter.

Lee (1978) analyzes the difference in wage determinants in both union and nonunion sectors using a simultaneous equations model with a binary qualitative variable and limited dependent variables. He uses socioeconomic, personal and individual characteristics as exogenous variables in his model. For both the union and nonunion equations, Lee includes regional location (North-Eastern, Northern-Central, and Southern Region dummies), city size (dummies for inside or outside an SMSA), educational level (highest grade completed dummies), market experience, and weeks worked per year as the socioeconomic variables. Race (a dummy for a white worker), sex, and health limitations are the personal characteristics that are included. Instead of an age variable, Lee uses market experience (defined as age minus years of highest grade completed minus 6) as well as a second order term to account for the nonlinearity of the earnings profile. To account for the worker's job type, dummy variables are included for industries such as mining, construction, manufacturing (durable and nondurable goods), transportation, communication, utilities, and sanitary services.

Incorporating the spirit of these earlier studies just discussed, we now propose our own model of income determination. The sample employed here for this income model is cross-sectional data from the Center of Disease Control's 2006 National Survey of Drug Use and Health. The survey was given to 55,279 individuals, but the sample for this model consists of 26,729 individuals. Those under the age of 18 were excluded, along with individuals who did not answer pertinent survey questions. We propose an equation following this general income model

$$y = f(S, D, P, F, E, W) \tag{5.6}$$

implying that y (income) is a function of S (if an individual has used and/or abused substances), D (if an individual is divorced), P (personal characteristics), F (family characteristics), E (economic characteristics), and W (workplace characteristics).

All income variables in this data set are categorical with increasing strength, and known category boundaries. As Caudill and Jackson (1993) note, analysts typically assign the category midpoint to each observation in a particular category in order to remedy this grouped data dependent variable problem. This midpoint value is usually found after the category end points have been transformed to, say, a Pareto distribution, a lognormal distribution, etc. Analysts would then use ordinary least squares (OLS) to estimate the parameters of the model. An alternative method used is to use maximum likelihood estimation (such as probit or logit) to account for the categorical nature of the dependent variable. Unfortunately, as stated by Caudill and Jackson, "both of these approaches result in unsatisfactory parameter estimates. OLS on the category midpoints produces inconsistent estimates. Even the qualitative dependent variable maximum likelihood techniques such as n-chotomous probit product inefficient estimates, since they ignore the information provided by the known values of the category boundaries" [Caudill and Jackson (1993, p. 120)].

We wish to analyze what exogenous variables determine family income. If y^* (an $(n \text{ x } 1)$ vector of implicit observations on family income) were observable, we could set up the following model:

$$y^* = \mathrm{X}\beta + \varepsilon \tag{5.7}$$

and estimate it with ordinary least squares. X is an $(n \text{ x } k)$ matrix of observations on the k independent variables in the model, β is a $(k \text{ x } 1)$ vector of unknown coefficients to be estimated, and ε is an $(n$ x $1)$ vector of stochastic disturbances, each element ε_i of which is assumed i.i.d. $N(0, \sigma^2)$.

We say that y^* is a vector of "implicit observations" because, in this conceptual framework, family income (y^*) is not directly observable. If it were observable, then each (cardinally measurable) $y_i{}^*$ would be independently normally distributed with mean $x_i\beta$ and constant variance σ^2, as implied by our assumptions on ε. Instead, all we are able to observe is the family income category (y_i) – with known end point values – within which $y_i{}^*$ falls. More precisely, if the real number line were partitioned into j mutually exclusive and exhaustive categories with boundaries A_j $(j = 0, \dots, J)$, then we observe $y_i = j$ if:

$$A_{j-1} < y_i^* < A_j \tag{5.8}$$

It is important to emphasize that the observed y_i are only of ordinal strength, but that the category boundaries $\{A_j\}$ are known cardinal numbers. Our problem within this framework is to obtain consistent and asymptotically efficient estimates of the unknown parameters, β and σ^2, of the model. One approach to obtaining such estimates is the method of maximum likelihood.

Based on the assumptions stated above, the probability that $y_i = j$, *i.e.*, the probability that y_i^* falls in the jth category, is given by:

$$P(y_i = j) = \tag{5.9}$$

$$P\left(A_{j-1} < y_i^* < A_j\right) = P\left\{\left[\frac{A_{j-1} - x_i\beta}{\sigma}\right] < \left[\frac{y_i^* - x_i\beta}{\sigma}\right] < \left[\frac{A_j - x_i\beta}{\sigma}\right]\right\}$$

$$= F\left[\frac{A_j - x_i\beta}{\sigma}\right] - F[\frac{A_{j-1} - x_i\beta}{\sigma}]$$

where $F(x)$ is the standard normal cumulative distribution function evaluated at x. For an independent random sample of n observations, the likelihood function is the product of these probabilities taken across the j categories and over the n observations, *i.e.*,

$$L = \prod_{i=1}^{n} \prod_{j=0}^{J} \{F[\frac{A_j - x_i\beta}{\sigma}] - F[\left(\frac{A_{j-1} - x_i\beta}{\sigma}\right)]\}^{\delta_{ij}} \tag{5.10}$$

where $\delta_{ij} = 1$ if the ith observation falls in the jth category, otherwise $\delta_{ij} = 0$. Therefore, the log likelihood function is:

$$L = \sum_{i-1}^{n} \sum_{j=1}^{j} \delta_{ij} \, ln\{F[\frac{A_j - x_i\beta}{\sigma}] - F[\left(\frac{A_{j-1} - x_i\beta}{\sigma}\right)]\} \tag{5.11}$$

Partially differentiating this equation with respect to the unknown parameters (β, σ) and setting the derivatives equal to zero yields $K + 1$ nonlinear equations which can be solved by iterative techniques (e.g., Davidon-Fletcher-Powell) to find consistent and asymptotically efficient estimates of the $\beta_m(m = 1, \ldots, k)$ and σ. Asymptotic standard errors of these estimates can be read from the diagonal of the negative inverse of the Hessian matrix of the above equation.

Referring back to our general model, our subsequent analysis will employ $FamilyIncome$ as our dependent variable, y, and the following measures of the fundamental variables $S, D, P, F, E,$ and W. S (whether an individual has used and/or misused substances) will include $RegularDrinker$, $EverMisusedLegalDrugs$, and $EverUsedIllegalDrugs$. D (whether an individual is divorced) will be represented by $Divorce$ in our equation. P (the personal characteristics we believe will affect income) will include $CollegeGraduate$, Age and Age^2, $White$, $Female$, and $SickRecently$. F (the family characteristic) will be $FamilySize$. E (the economic characteristics we believe will affect income) will include $Fulltime$ and $NonMSA$. W (the workplace characteristics we believe will affect income) will include $WhiteCollar1$, $BlueCollar$, $OneEmployer$, $SmallFirm$, and $LargeFirm$. Therefore, our econometric model is as follows:

$$Income = \tag{5.12}$$

$$\alpha + \beta_1 EverMisusedLegalDrugs + \beta_2 EverUsedIllegalDrugs + \beta_3 RegularDrinker + \beta_4 Divorce + \beta_5 CollegeGraduate + \beta_6 Age + \beta_7 Age^2 + \beta_8 White + \beta_9 Fulltime + \beta_{10} NonMSA + \beta_{11} FamilySize + \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar + \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm + \beta_{18} SickRecently + \epsilon$$

The NSDUH's data on income and employment are very limited. All income variables (whether family income or respondent's income) are categorical in nature. Our dependent variable, $Income$, asked the respondents which of the following categories contained their total family income: Option 1 (less than $20,000), option 2 (between $20,000 and $49,999), option 3 (between $50,000 and $74,999), and option 4 (greater than $75,000). For the reasons mentioned earlier regarding dependent variables of this kind, grouped data regression will be used with the interior limits of the intervals being $20,000, $50,000, and $75,000.

$EverMisusedLegalDrugs$ is a dummy variable equal to one if the respondent has ever used painkillers, tranquilizers, stimulants, or sedatives non-medically. Specifically, respondents were asked, "Have you ever, even once, used any type of prescription [drug type] that was not prescribed for you or that you took only for the experience or feeling it caused?"

$EverUsedIllegalDrugs$ is a dummy variable equal to one if respondent ever used marijuana, cocaine, crack, heroine, or hallucinogens. Hallucinogens include LSD, PCP, peyote, mescaline, mushrooms, ecstasy, or other listed hallucinogens.

$RegularDrinker$ is a dummy variable equal to one if the answer to "When is the last time you used alcohol?" was "within the past 30 days".

Misusing legal drugs, using illegal drugs, and regular drinking are activities that inhibit an individual's judgment, behavior, and productivity. This decrease in productivity would likely decrease an individual's income, making all three predicted coefficients negative.

However, the nature of the relationship between regular drinking and income could be argued. Although this variable is intended to pick up how regular drinking affects income, there may be a simultaneous relationship between alcohol use and income. A higher income clearly yields the opportunity to be able to buy more alcohol. Regular drinking, even a glass of wine with dinner, for example, can be a fairly expensive habit that requires income. On the other hand, it may be that jobs benefiting from taking clients out for dinner and drinks yield a higher income. We will fully exploit this idea of simultaneity bias in the following chapter using our simultaneous equation model.

The same problem may also occur with illegal drug use and legal drug misuse. Using illegal drugs recreationally rather than habitually may be affected by income, meaning an increase in income may lead to an increase illegal drug use. Likewise, obtaining legal drugs requires income or adequate health insurance, and if an individual does not have enough income to purchase drugs or acquire insurance, they are less likely to abuse legal drugs. Ignoring this potential for simultaneity bias, theory would suggest that all three types of substance use would decrease income because of a decrease in productivity.

$Divorce$, a dummy variable equal to one if the individual's marital status was divorced or separated at the time of the survey, is expected to be negative. Because we are dealing with family income, rather than individual income, a divorce would decrease this income (if the spouse works) because the spouse's income is no longer contributing to family income. Also, any division of labor that may have been present during a marriage is removed after a divorced, meaning time may be taken away from work. For example, if a couple alternated taking the kids to and from school, a divorce may mean that the husband or wife has to go to work late and leave work early in order to pick up the kids. This decrease in time at work would therefore lead to a decrease in income.

$CollegeGraduate$, a dummy variable equal to one if the individual has completed college, is expected to be positive. The opportunity cost of additional education includes tuition, school expenses, and foregone income (Willis 1986). This implies that an individual who obtains additional education will most likely require higher pay in compensation. "Individuals will only undergo additional schooling or training (i.e. invest in their human capital) if the costs (tuition and training course fees, forgone earnings while at school and reduced wages during the training period) are compensated by sufficiently higher future earnings" (Blundell p. 3). Therefore, we expect that completing college (extra education with a high opportunity cost) will yield jobs with higher income in compensation.

The literature typically includes either age or experience, in addition to education, to account for more human capital investment. The NSDUH includes no information on an individual's work experience. The data on the individual's age is categorical: an individual selects "1" if they are 12-17 years old, "2" for 18-25 years old, "3" for 26-34 years old, "4" for 35-49 years old, "5" for 50-64 years old, and "6" if the individual is 65 years or older. Midpoints for each age category were created, as well as a midpoint for the open ended interval (65 years plus)[18] which is equal to 72. Age^2 are the midpoints squared. We expect income to increase

[18] The midpoint for the open ended interval is calculated by running a grouped data regression where the left-hand side variable is the categorical age variable, the explanatory variable is the constant term only, and the interior limits are 18, 26, 35, 60, and 65. Estimating this model yields a mean (23.98) and variance (19.51) of the distribution. A z value that corresponds to the lower limit of the highest ranked (open-ended category) is estimated with the following equation:

$$Z = \frac{OEI_{LL} - Mean}{Standard\ Deviation}$$

with age, but perhaps at a decreasing rate after a certain age is reached. This expectation is based off of Becker's (1964) conclusion that age earnings profiles have strictly concave shapes. In other words, we can expect Age to be positive and Age^2 to be negative.

$White$ is a dummy variable equal to one if the individual's race is non-Hispanic white, equal to zero if their race is non-Hispanic black/African American, non-Hispanic Native American/Alaskan Native, non-Hispanic Native Hawaiian/Other Pacific Island, non-Hispanic Asian, or Hispanic. We expect this coefficient to be positive due to the extensive literature on racial discrimination in occupation and earnings. "Despite the gains made by blacks in overcoming occupational segregation, however, black men's earnings continued to fall far short of the earnings of their white peers at all levels of economic attainment (Harrison and Bennett 1995). Something worth mentioning, however, is that many of these studies on racial discrimination in earnings don't analyze this difference *ceteris paribus*. These studies typically look at who makes more money on average, rather than factoring out covariates. We wish to see, after controlling for different levels of education, age, etc., if there is some racial or ethnic difference in income.

$Fulltime$ is a dummy variable equal to one if the respondent's employment status is "employed full time". Working fulltime reveals an attachment to the labor force and has its

---

Where OEILL represents the open-ended interval lower limit. Using our computed values, we find a z value equal to 2.5.We then use the censored normal results to find the mean of the category in terms of z using the following equation:

$$E[z] = \frac{\theta(z)}{1 - \Phi(z)} = x$$

where θ represents the standard normal density and Φ represents the standard normal distribution. Finally, we back out the midpoint value on the original data's scale by "unnormalizing" the result above,

$$\frac{M - Mean}{Standard\ Deviation} = x \rightarrow M = Mean + (x * Standard\ Deviation)$$

where M is the desired estimate of the category midpoint and is equal to 72.

benefits, typically including health insurance for the individual and their family. Contract laborers don't receive any benefits. These benefits would increase the family's real income, implying a positive relationship between working fulltime and family income level. While a full time janitor may make less than a part time consultant, we would generally expect full time workers to work more and hence make more income, *ceteris paribus*. Therefore, we expect the coefficient of $Fulltime$ to be positive.

$NonMSA$ is a dummy variable equal to one if the respondent lives in an area not in a "Core Based Statistical Area" (CBSA) according to the 2000 Census data and the June 2003 CBSA classifications provided by the Office of Management and Budget. Theory suggests that in a frictionless world, i.e., one with no moving costs and no pure locational preference, urban-rural migration will occur until real incomes equalize. Therefore, if the cost of living is higher in the city, nominal incomes would have to be higher in the city so that real incomes would be equal. This means that individuals not living in a city (living in a Non-MSA) would have lower incomes, so we predict this coefficient to be negative.

The coefficient of $FamilySize$, a variable representing the number of persons in the respondent's household, is expected to be negative. There are many opportunity costs associated with raising additional children. The monetary outlays and the time that must be spent raising the children as well must both be considered. That time could have been spent working, leading us to believe that an increase in the number of children would decrease family income. On the other hand, if family members are in the work force, a greater number of family members would yield a greater family income, *ceteris paribus*. An increase in family size may also put pressure on parents to work more to give each additional child the same support that was given to the older children. Therefore, the sign of the $FamilySize$ coefficient is ambiguous.

$Female$, a dummy variable equal to one if the respondent is female, is expected to be negative. As stated earlier, there are many studies that focus on gender differences in income alone. Most of these studies show that, *ceteris paribus*, a female with the exact same credentials, education, and background will make less than their male counterpart. Jones and Jackson (1992) posit that this may be because women do not feel as tied to the labor force.

$WhiteCollar1$ is a variable equal to one if the respondent answered that their occupation is either a) Executive/Administrative/Managerial/Financial, b) Professional (not Education/Entertainment/Media), c) Education and Related Occupations, or d) Entertainers, Sports, Media, and Communications. $BlueCollar$ is a created dummy variable equal to one for respondents whose occupation is either a) Farming, Fishing, and Forestry Occupations, b) Installation, Maintenance, and Repair Workers, c) Construction Trades and Extraction Workers, d) Production, Machinery Setters/Operators/Tenders, or e) Transportation and Material Moving Workers. The omitted dummy variable, WhiteCollar2, is equal to one for individuals who work in fields other than those included in *WhiteCollar1* or *BlueCollar*, consisting of a) Technicians and Related Support Occupations, b) Sales Occupations, c) Office and Administrative Support Workers, d) Protective Service Occupations or e) Service Occupations, except Protective.

It is reasonable to assume that the specified white collar jobs require more skilled workers, or workers who have undergone extra education or training to obtain their jobs. The blue collar jobs, on the other hand, generally require less education and training. If employers are compensating their workers according to their skill level, or additional education or training the workers have undergone, we would expect the white collar workers to make more money than their blue collar counterparts. In other words, $WhiteCollar1$ is expected to be positive while $BlueCollar$ is expected to be negative.

$OneEmployer$ is a dummy variable equal to one if the respondent had only one employer in the past twelve months. The predicted sign of this coefficient is ambiguous because *OneEmployer* is an indicator of mobility. Some individuals benefit from being mobile, because they can move from job to job to earn more money. On the other hand, there are payoffs to having a stable job. It may be that the age of the employee makes a difference. Young employees may hop from job to job in hopes of earning higher pay, but at some point[19] they may see a larger payoff for stability over mobility.

$SmallFirm$ is a dummy variable equal to one if, at the location where the individual works, less than 24 people work for the employer. $LargeFirm$ is equal to one if the respondent answered that more than 100 people work for the employer. The omitted dummy variable, MediumFirm, is equal to one if 25-99 people work for the respondent's employer. We predict that individuals in large firms tend to make more money than their small firm counterparts for several reasons. Schmidt and Zimmerman (1989) cite the various explanations in the literature for why individuals in large firms tend to make more money than individuals in smaller firms. For example, Masters (1969) argues that this is the case because larger firms have more unpleasant work environments, while Oi (1983) argues that workers in large firms are more qualified and engaged in highly specialized work. Another possibility is that efficiency wages (wages above the market-clearing level to incentivize workers) may increase with firm size, as tested by Schmidt and Zimmerman (1989). Bigger firms also sell in larger market areas, hence these bigger firms face greater demand and consequently higher prices for their products. These higher product prices would lead to higher wages under the marginal productivity theory,

[19]This thought might suggest an interaction variable between our age variable and the one employer variable, to account for the fact that mobility might result in higher income for young workers and lower income for older workers.

therefore yielding another reason why we predict that individuals in large firms tend to have higher incomes.

$SickRecently$ is a created dummy variable equal to one if the respondent missed work for being sick or injured at least once in the past 30 days. We expect this coefficient to be negative because individuals are typically allotted a limited number of sick days a year before pay is reduced. Also, the opportunity cost of missing work is much higher for individuals with higher paying jobs. Whereas a factory worker may be able to get someone to cover their shift, a lawyer may personally have to catch up on all of the work missed while sick or injured. Therefore, we would assume that individuals in higher paying jobs who have higher incomes are less likely to miss work.

Table 5 provides the GDR coefficients and asymptotic t-statistics (which can appropriately be viewed as standard normal deviates) for modeling family income. As expected, the coefficient on $Divorce$ is negative and significant. In other words, being divorced is associated with a decrease in income (specifically, about a $12,000 decrease in annual income). Schmidt and Zimmerman's (1989) results concur in that they find a significantly positive impact of being married on earnings.

While $EverMisusedLegalDrugs$ is negative and significant, $RegularDrinker$ and $EverUsedIllegalDrugs$ are positive and significant. This implies that while misusing legal drugs is associated with a decrease in income (which was expected), being a regular drinker and using illegal drugs are associated with an increase in income.

Mullahy and Sindelar (1994) found that problem drinking is associated with reduced employment and increased unemployment. They state, "alcoholism may affect income more by restricting labor market participation than by affecting the wages of [sic] worker" [Mullahy and

Sindelar (1994, p. 494)]. It is important to point out, however, that regular drinking is not the same as problem drinking or alcoholism, which may be why this coefficient is not as predicted. Also, job-related networking and drinks after work with coworkers may explain the positive relationship between regular drinking and income.

The positive coefficient associated with regular drinking could also be due to simultaneity bias (which we mentioned earlier and will be explored in the next chapter) or an incorrectly reversed causal specification (Ekelund et al. 2006). Overall, it is not surprising to find a positive and significant relationship between income and alcohol use. What we may be observing in this coefficient is this significant relationship, but the causality may actually flow backward. In other words, rather than regular drinking causing an increase in income, it may be that an increase in income enables an individual to, say, have a glass of wine with dinner.

This reverse causality may also help explain the positive coefficient of *EverUsedIllegalDrugs*. It may not be the use of illegal drugs that causes an increase in income, but rather the other way around. An increased income enables the purchase of costly illegal drugs, and that is the relationship reflected in the positive coefficient. Gill and Michaels (1992) actually find that, if they allowed for factors that affect both wages and the decision to do drugs, drug users have higher wages than non-drug users and that "a sample of all drug users (which included users of "hard" and "soft" drugs) had lower employment levels than non-drug users, but the smaller sample consisting only of users of hard drugs, surprisingly, did not" [Gill and Michaels (1992, p. 419)].

$CollegeGraduate$ is positive and significant, as expected. Michael Hout (2012) summarizes the literature on the effects of being a college graduate. He notes that DiPrete and Buchmann (2006) and Western et al. (2008) find that "family structure interacts with education in complex ways because each partner's education affects his or her prospect of marrying, divorcing, and remarrying as well as work hours" [Hout (2012, p. 382)]. In his own calculations, using data from 2007 to 2009, Hout finds that both male and female college graduates made more money than their less educated counterparts, and that family income was higher for college graduates as well. Schmidt and Zimmerman (1989) also find that earnings increase with years of schooling.

Age is positive and significant while Age^2 is negative and significant. In other words, an increase in an individual's age is associated with an increase in income up to a point, after which an increase in an individual's age is associated with a decrease in family income. Specifically, a one year increase in an individual's age is associated with an increase in income until the individual reaches age 61[20]. After age 61, a one year increase in an individual's age is associated with a decrease in income. This concurs with Becker's (1964) finding, stated earlier, that age earnings profiles have strictly concave shapes.

$White$ is positive and significant, implying that being white is associated with an increase in income. Lee (1978) also finds that white individuals received higher wages than their nonwhite counterparts. Murnane and Willet (1995) find that black males received lower wages than their white male counterparts, but they find no significant wage gap between Hispanics and whites, nor between black and white females.

---

[20] We can find this critical value using the fact that $(\beta_5 + 2\beta_6 A)xf(X\beta) = 0$. Since f(Xβ) must be positive, the critical value of A can be found by setting $\beta_5 + 2\beta_6 A = 0$, meaning that $2\beta_6 A = -\beta_5$, and therefore $A = \frac{-\beta_5}{2\beta_6}$. Replacing our coefficients for Age (1,693.252) and Age^2 (-13.881) for β_5 and β_6 respectively, we find that our critical A is equal to 61.

As we predicted, $Fulltime$ is positive and significant. Full time workers have more work hours (meaning more pay, *ceteris paribus*) and then to have a greater attachment to the labor force and better benefits, including insurance.

$NonMSA$ is negative and significant implying that individuals living in a major metropolitan area make more money that individuals living in a more rural area. This concurs with Lee (1978) who found higher wages in large SMSA's versus those living outside metropolitan areas. This may be because the cost of living is much higher in metropolitan areas, and therefore individuals living in a city make greater nominal incomes (*ceteris paribus*).

$FamilySize$ is positive and significant, implying that an increase in an individual's family size is associated with an increase in income. This concurs with Schmidt and Zimmerman's (1989) conclusion that the number of children has a statistically significant positive impact on earnings. This may be because the added individuals in the family are contributing to family income, or possibly because parents feel the need to work more as family size increases to continue to support each child equally.

$Female$ is negative and significant, implying that being female is associated with a decrease in income. Lee (1978) reaches the same conclusion that males receive higher wages. Blinder (1973) also reaches this conclusion of higher wages for males, but finds that about two-thirds of this differential is due to outright discrimination in labor markets and one-third of the differential is due to other endogenous variables such as occupational status and job seniority.

$WhiteCollar1$ is positive and significant, while $BlueCollar$ is negative and significant. Schmidt and Zimmerman (1989) included two similar dummy variables in their model: one for whether the worker is unskilled, and one for whether the worker is a "qualified white collar worker." In every specification of their model, they find the unskilled worker variable to be negative and significant and the “qualified white collar worker” variable to be positive and significant. In other words, Schmidt and Zimmerman (1989) found that “qualified” white collar workers make more money while unskilled workers make less money. Because these “qualified” employees either completed additional schooling or training, employers had to compensate those skilled workers with additional income.

$OneEmployer$ is positive and significant. In light of the mobility versus stability argument discussed above, this would imply that our sample of individuals benefits financially from remaining stable in their job. This supports our assertion that age plays a key part in benefiting from mobility or stability, because the average age of our sample is 32 years old. Schmidt and Zimmerman’s (1989) found that their mobility variable, the number of jobs held in different firms, was mostly insignificant. They reasoned that their finding may be attributable to two offsetting factors: a “job shopping effect” and a “lemon effect”. Their “job shopping effect” is the mobility effect that we have identified: “more successful people tend to climb the career ladder by moving faster to better jobs and individuals having had more matches are more likely to have found a good one” [Schmidt & Zimmerman (1989, p. 5)]. Their “lemon effect” was believed to have a negative influence on income, because bad workers are more likely to have a history of losing jobs.

$SmallFirm$ is negative and significant, while $LargeFirm$ is positive and significant. This reflects the positive relationship between firm size and wages found in papers such as Blanchflower (1986), Heywood (1986), Brown and Medoff (1989), and Dunn (1986), to name a few. Schmidt and Zimmerman (1989) found that, even after controlling for personal and work characteristics, wages increase with firm size, and this relationship is significant. They note that Dunn (1986) and Brown and Madoff (1989) reach the same conclusion, and that “there have to exist other reasons why large firms are able to survive the competitive disadvantages of higher labor costs in the long-run. Non-production economies of scale may be one possible explanation" [Schmidt & Zimmerman (1989, p. 8)]. We believe that the explanation may be driven by demand, in that larger firms sell in "bigger markets", implying higher product prices and higher wages assuming wages are equal to price times the marginal product.

$SickRecently$ is negative and significant. This supports our hypothesis that higher-paid workers have a larger opportunity cost of missing work. Skinner et al. (2014) find that individuals in the highest income category ($75,000 or more) miss less work, and cite opportunity costs of work absenteeism as the reason.

This chapter analyzed how being divorced, using substances, personal characteristics, family characteristics, and economic characteristics affect income. Using a grouped data regression for the reasons specified earlier, and analyzing these coefficients and their t-statistics, we find that individuals who are divorced, misuse legal drugs, live in a Non-MSA region, are over the age of 61, are female, have a blue collar job, work for a small firm, and have missed work recently are associated with lower income. At the same time, individuals who are regular drinkers, use illegal drugs, are college graduates, are white, are older (but not over the age of 61), work full time, have larger families, work white collar jobs, worked for only one employer in the past year, and work for large firms are associated with higher income.

All signs were as predicted, except for the signs on our regular drinker and illegal drug use variables. These positive coefficients imply that being a regular drinker or using illegal drugs increases income. While we can argue the rationale behind why the regular drinking sign is positive (it's just a glass of wine at dinner, job networking, etc.), the illegal drug use sign is counterintuitive. It may be that income and the use of illegal drugs are simultaneously determined. For example, an increased income enables the purchase of costly illegal drugs, and that positive relationship between income and illegal drugs may explain why this coefficient is positive. Simultaneity bias, specifically this potentially simultaneous relationship between income and illegal drug use, will be discussed in the following chapter.

CHAPTER VI

SIMULTANEOUS EQUATIONS

The previous three chapters analyzed the productivity costs of divorce and substance use and the factors that predict divorce, substance use, and income. As discussed in the last chapter, there are several relationships we have analyzed that may be simultaneously determined. In this chapter, we will summarize several simultaneous equation models in the literature, present and discuss our simultaneous equation model, and analyze our results.

Lee (1978) uses microeconomic data from the Survey of Economic Opportunity Sample of 1967 to analyze the joint determination of unionism and the effects of unions on wage rates. Because "economic considerations suggest that the propensity to join a union depends on the net wage gains that might results from trade union membership", he uses a simultaneous equation model [Lee (1978, p. 415)]. Lee's model consists of full time workers who have two options: to become a union member or not (a decision made by not only the worker, but the union as well). This means that each worker faces two wage rates, the nonunion wage and the union wage.

Union membership, however, is not free. "Union initiation, fees, dues and other requirements have to be paid or met, and there are certain restrictions on entry (G.S. Becker [1959]). With these factors and some other taste factors, the individual becomes a union member or not, partially by his choice and partially by the selectivity of the labor union. With union or nonunion status determined, the workers' wage rates are set according to their socio-economic status and the job they perform" [Lee (1978, p. 416)].

Lee (1978) proposes that individual i will join the union if the percentage union wage differential is greater than his reservation wage, in other words, he will join the union if:

$$\frac{W_{ui}-W_{ni}}{W_{ni}} > \rho_i \tag{6.1}$$

where W_{ui} and W_{ni} represent the union and nonunion wages, respectively, for individual i, and ρ_i represents the worker's reservation wage based on his preferences. This reservation wage is a function of the worker's individual characteristics and the costs of becoming a union member, and therefore summarizes how receptive he is to the labor union. This reservation wage can therefore be positive or negative. Specifically, Lee models the reservation wage as follows:

$$\rho_i = \alpha X_i + \beta C_i + \varepsilon_{1i} \tag{6.2}$$

where X_i is a vector of personal characteristics, C_i summarizes the cost (both monetary and nonpecuniary costs) of becoming a union member, and ε_{1i} is the error term , which is assumed to be normally distributed with a mean of zero and a variance of σ_i^2.

If the union uses price rationing to accept members, these dues and fees may be a good measure of the cost, C_i. If some other method of rationing is used, however, these dues and fees do not provide a good proxy for C_i. Therefore, Lee (1978) models the cost of union membership as follows:

$$C_i = \gamma_i + \gamma_2 X_i + \gamma_3 Z_i + \varepsilon_{2i} \tag{6.3}$$

where X_i is a vector of personal characteristics of worker i, Z_i is a vector of industry attributes where the worker is employed, and ε_{2i} is the error term , which is assumed to be normally distributed with a mean of zero and a variance of σ_i^2.

Therefore, Lee (1978) writes that worker i will join the union if:

$$\frac{W_{ui}-W_{ni}}{W_{ni}} > (\alpha+\beta\gamma_2)X_i + \beta\gamma_1 + \beta\gamma_3 Z_i + \varepsilon_{1i} + \beta\varepsilon_{2i} \tag{6.4}$$

Rewriting this criterion in the form of a probit model, Lee (1978) states that whether a worker is in the union ($I_i^* > 0$) can be modeled by:

$$I_i^* = \delta_0 + \delta_1 \left(\frac{W_{ui} - W_{ni}}{W_{ni}}\right) + \delta_2 X_i + \delta_3 Z_i - \varepsilon_i \qquad (6.5)$$

Some unionism studies only have one wage regression, where wage is a function of characteristics and a union dummy variable, but this formulation does not allow for interactions. Lee's model allows for interactions by including two wage equations: one for union workers and one for nonunion workers. The model to be estimated has three equations:

$$\log W_{ui} = \theta_{u0} + X_{ui}\theta_{u1} + Z_{ui}\theta_{u2} + \varepsilon_{ui} \qquad (6.7)$$

$$\log W_{ni} = \theta_{n0} + X_{ni}\theta_{n1} + Z_{ni}\theta_{n2} + \varepsilon_{ni} \qquad (6.8)$$

$$I_i^* = \delta_0 + \delta_1(\log W_{ui} - \log W_{ni}) + \delta_2 X_i + \delta_3 Z_i - \varepsilon_i \qquad (6.9)$$

where $\varepsilon_u \sim n(0, \sigma_u^2)$, $\varepsilon_n \sim n(0, \sigma_n^2)$, $\varepsilon \sim n(0, \sigma_\varepsilon^2)$. Because the observed wage rate ($\log W_{ui}$ and $\log W_{ni}$) depends on the worker's union status (I_i^*=1 or 0), this is a simultaneous equations model that includes qualitative and limited dependent variables. The first stage is estimated by probit, and the second stage is estimated by ordinary least squares.

Substituting the wage equations ($\log W_{ui}$ and $\log W_{ni}$) into I_i^*, Lee (1978) estimates the following model using probit:

$$I_i^* = \gamma_0 + \gamma_1 X_i' + \gamma_2 Z_i' - \varepsilon^* \qquad (6.10)$$

where X_i' represents all individual characteristics and Z_i' represents observable union and industrial attributes. He then obtains consistent estimates $\widehat{\gamma_0}$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ for γ_0, γ_1, and γ_2 respectively, using probit after normalizing $\sigma_{\varepsilon *}^2$.

Conditional on union status, the union wage equation is:

$$\log W_{ui} = \theta_{u0} + X_{ui}\theta_{u1} + Z_{ui}\theta_{u2} + \sigma_{1\varepsilon^*}\left(-\frac{f(\Psi_i)}{F(\Psi_i)}\right) + \eta_u \qquad (6.11)$$

where F is the cumulative distribution of a standard normal random variable, and f is its density function, $E(\eta_u | I_i = 0) = 0$, $\Psi_i = \gamma_0 + \gamma_1 X_i' + \gamma_2 Z_i'$.

Conditional on nonunion status, the nonunion wage equation is

$$\log W_{ni} = \theta_{n0} + X_{ni}\theta_{n1} + X_{ni}\theta_{n2} + \sigma_{2\varepsilon^*}\left(-\frac{f(\Psi_i)}{F(\Psi_i)}\right) + \eta_n \tag{6.12}$$

where $E(\eta_n | I_i = 0) = 0$. The parameters (θ_{uj}) can be estimated consistently by regressing the observed union wage $\log W_{ui}$ on X_{ui}, Z_{ui}, and $\left(-\frac{f(\widehat{\Psi}_i)}{F(\widehat{\Psi}_i)}\right)$; where $\widehat{\Psi}_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_i' + \hat{\gamma}_2 Z_i'$; θ_{uj} can be estimated in a similar fashion.

Consider the simultaneous equations model:

$$Y\Gamma + XB + E = 0 \tag{6.13}$$

where 0 is a $(T \text{ x } M)$ matrix of zeros, Y and X are the sample values of the jointly dependent and the predetermined variables, respectively, and E is the matrix of unobservable values of the random error vectors. Γ is a (M x M) matrix of coefficients of the current endogenous variables, where each column refers to the coefficients for a particular equation. B is a (K x M) matrix of unknown coefficients of the exogenous-predetermined variables, and each column contains the coefficients of a particular equation.

In the presence of limited endogenous variables, we consider either maximum likelihood or two-stage techniques. Each of these methods relies on consistent estimation of the reduced form in the first stage. The reduced form of the above equation is:

$$Y = X\Pi + V \tag{6.14}$$

where $\Pi = -B\Gamma^{-1}$ and $V = -\mathrm{E}\Gamma^{-1}$. The reduced form equations for specific endogenous variables can written as:

$$\mathrm{y}_i = X\pi_i + \mathrm{v}_i \tag{6.15}$$

The parameters π_i can be estimated consistently by probit, tobit, etc. depending on the specific form of y_i. The structural parameters are estimated in the second stage. Let the ith structural equation be:

$$y_i = Y_i\gamma_i + X_i\beta_i + e_i \quad (6.16)$$

and partition the reduced form equation above as:

$$[y_i \;\; Y_i \;\; Y_i^*] = X[\pi_i \;\; \Pi_i \;\; \Pi_i^*] + [v_i \;\; V_i \;\; V_i^*] \quad (6.17)$$

where the partitions correspond to the LHS endogenous variable (y_i), the RHS endogenous variables (Y_i), and the excluded endogenous variables (Y_i^*) in the ith equation. Note that $Y_i = X\Pi_i + V_i$. Substituting for Y_i, we obtain:

$$y_i = (X\Pi_i + V_i)\gamma_i + X_i\beta_i + e_i \quad (6.18)$$

$$= X\Pi_i\gamma_i + X_i\beta_i + V_i\gamma_i + e_i \quad (6.19)$$

$$= X\Pi_i\gamma_i + X_i\beta_i + v_i \quad (6.20)$$

If $X\widehat{\Pi}_i\gamma_i$ is added to and subtracted from this equation, we obtain:

$$y_i = X\widehat{\Pi}_i\gamma_i + X_i\beta_i + w_i \quad (6.21)$$

The second stage of estimation consists of estimating this equation by probit, tobit, and so on, depending on the nature of y_i.

We now propose our own simultaneous equation model. The sample employed for this income model is cross-sectional data from the Center of Disease Control’s 2006 National Survey of Drug Use and Health. The survey was given to 55,279 individuals, but the sample for this model consists of 26,615 individuals. Those under the age of 18 were excluded, along with any individuals who did not answer pertinent survey questions.

The model to be estimated consists of five reduced-form equations and five structural equations. The structural form equations are as follows

$$P(Divorce) = \tag{6.22}$$

$$\alpha + \beta_1 \widehat{EverMisused\,LegalDrugs} + \beta_2 \widehat{EverUsedIllegalDrugs} + \beta_3 \widehat{RegularDrinker} + \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White + \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize + \beta_{11} \widehat{FamilyIncome} + \beta_{12} \widehat{FamilyIncome}^2 + \beta_{13} WorkingWoman + \epsilon$$

$$P(RegularDrinker) = \tag{6.23}$$

$$\alpha + \beta_1 \widehat{EverMisused\,LegalDrugs} + \beta_2 \widehat{EverUsedIllegalDrugs} + \beta_3 \widehat{Divorce} + \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White + \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize + \beta_{11} \widehat{FamilyIncome} + \beta_{12} \widehat{FamilyIncome}^2 + \beta_{13} Smoker + \epsilon$$

$$P(LegalDrugMisuse) = \tag{6.24}$$

$$\alpha + \beta_1 \widehat{RegularDrinker} + \beta_2 \widehat{EverUsedIllegalDrugs} + \beta_3 \widehat{Divorce} + \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White + \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize + \beta_{11} \widehat{FamilyIncome} + \beta_{12} \widehat{FamilyIncome}^2 + \beta_{13} PsychologicalDistress + \epsilon$$

$$P(IllegalDrugUse) = \tag{6.25}$$

$$\alpha + \beta_1 \widehat{RegularDrinker} + \beta_2 \widehat{EverMisused\,LegalDrugs} + \beta_3 \widehat{Divorce} + \beta_4 CollegeGraduate + \beta_5 Age + \beta_6 Age^2 + \beta_7 White + \beta_8 Fulltime + \beta_9 NonMSA + \beta_{10} FamilySize + \beta_{11} \widehat{FamilyIncome} + \beta_{12} \widehat{FamilyIncome}^2 + \beta_{13} Risky + \epsilon$$

$$Income = \tag{6.26}$$

$$\alpha + \beta_1 \widehat{EverMisusedLegalDrugs} + \beta_2 \widehat{EverUsedIllegalDrugs} + \beta_3 \widehat{RegularDrinker} + \beta_4 \widehat{Divorce} + \beta_5 CollegeGraduate + \beta_6 Age + \beta_7 Age^2 + \beta_8 White + \beta_9 Fulltime + \beta_{10} NonMSA + \beta_{11} FamilySize + \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar + \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm + \beta_{18} SickRecently + \epsilon$$

The exogenous variable P (the individual's personal characteristics) includes $CollegeGraduate$, Age, Age^2, $White$, $Risky$, $PsychologicalDistress$, $Female$, and $SickRecently$. F (the family characteristic) is $FamilySize$. E (the economic characteristics) includes $Fulltime$ and $NonMSA$. W (the workplace characteristics) includes $WhiteCollar1$, $BlueCollar$, $OneEmployer$, $SmallFirm$, and $LargeFirm$.

The reduced-form equations are of the following forms:

$$P(D) = f(P, F, E, W) \tag{6.27}$$

$$P(S) = f(P, F, E, W) \tag{6.28}$$

$$y = f(P, F, E, W) \tag{6.29}$$

In other words, $P(D)$ (the probability of being divorced), $P(S)$ (the probability of using and/or abusing substances), and y (income) are functions of exogenously determined P (personal characteristics), F (family characteristics), E (economic characteristics), and W (workplace characteristics).

Specifically, the reduced form equations are as follows

$$P(Divorce) = \tag{6.30}$$

$$\alpha + \beta_1 CollegeGraduate + \beta_2 Age + \beta_3 Age^2 + \beta_4 White + \beta_5 Fulltime + \beta_6 NonMSA + \beta_7 FamilySize + \beta_8 WorkingWoman + \beta_9 Smoker + \beta_{10} Risky + \beta_{11} PsychologicalDistress + \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar + \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm + \beta_{18} SickRecently + \epsilon$$

$$P(RegularDrinker) = \tag{6.31}$$

$$\alpha + \beta_1 CollegeGraduate + \beta_2 Age + \beta_3 Age^2 + \beta_4 White + \beta_5 Fulltime + \beta_6 NonMSA + \beta_7 FamilySize + \beta_8 WorkingWoman + \beta_9 Smoker + \beta_{10} Risky + \beta_{11} PsychologicalDistress + \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar + \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm + \beta_{18} SickRecently + \epsilon$$

$$P(LegalDrugMisuse) = \tag{6.32}$$

$$\alpha + \beta_1 CollegeGraduate + \beta_2 Age + \beta_3 Age^2 + \beta_4 White + \beta_5 Fulltime + \beta_6 NonMSA + \beta_7 FamilySize + \beta_8 WorkingWoman + \beta_9 Smoker + \beta_{10} Risky + \beta_{11} PsychologicalDistress + \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar + \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm + \beta_{18} SickRecently + \epsilon$$

$$P(IllegalDrugUse) = \tag{6.33}$$

$$\alpha + \beta_1 CollegeGraduate + \beta_2 Age + \beta_3 Age^2 + \beta_4 White + \beta_5 Fulltime$$
$$+ \beta_6 NonMSA + \beta_7 FamilySize + \beta_8 WorkingWoman$$
$$+ \beta_9 Smoker + \beta_{10} Risky + \beta_{11} PsychologicalDistress$$
$$+ \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar$$
$$+ \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm$$
$$+ \beta_{18} SickRecently + \epsilon$$

$$Income = \tag{6.34}$$

$$\alpha + \beta_1 CollegeGraduate + \beta_2 Age + \beta_3 Age^2 + \beta_4 White + \beta_5 Fulltime$$
$$+ \beta_6 NonMSA + \beta_7 FamilySize + \beta_8 WorkingWoman$$
$$+ \beta_9 Smoker + \beta_{10} Risky + \beta_{11} PsychologicalDistress$$
$$+ \beta_{12} Female + \beta_{13} WhiteCollar1 + \beta_{14} BlueCollar$$
$$+ \beta_{15} OneEmployer + \beta_{16} SmallFirm + \beta_{17} LargeFirm$$
$$+ \beta_{18} SickRecently + \epsilon$$

Using probit analysis to estimate the first four reduced-form equations and grouped data regression to estimate the fifth reduced-form equation, we are able to obtain consistent estimates for the reduced form parameters associated with our exogenous variables and hence the predicted probabilities or values for our endogenous variables. The first four reduced-form equations give us predicted probabilities of either zero or one for $p(Divorce)$, $p(RegularDrinker)$, $p(LegalDrugMisuse)$ and $p(IllegalDrugUse)$. In other words, we are able to predict whether, given an individual's exogenously determined characteristics, they are divorced, are a regular drinker, misuse legal drugs, or use illegal drugs. The fifth reduced-form equation gives us predicted values for $Income$. In other words, we are able to predict an individual's income based on his or her exogenously determined characteristics. These predicted values are then used in the five structural equations specified above.

The exogenous variables are the same variables discussed in the past three chapters. $CollegeGraduate$ is a dummy variable equal to one if the individual has completed college. Age is categorical: an individual selects "1" if he is 12-17 years old, "2" for 18-25 years old, "3" for 26-34 years old, "4" for 35-49 years old, "5" for 50-64 years old, and "6" for 65 years or older. Midpoints for each age category were created, as well as a midpoint for the open ended interval (65 years plus)[21], which is equal to 72. Age^2 are the midpoints squared. $White$ is a dummy variable equal to one if the individual's race is non-Hispanic white, equal to zero if their race is non-Hispanic black/African American, non-Hispanic Native American/Alaskan Native, non-Hispanic Native Hawaiian/Other Pacific Island, non-Hispanic Asian, or Hispanic.

$Fulltime$ is a dummy variable equal to one if the respondent's employment status is "employed full time". $NonMSA$ is a dummy variable equal to one if the respondent lives in a segment not in a "Core Based Statistical Area" (CBSA) according to the 2000 Census data and

---

[21] The midpoint for the open ended interval is calculated by running a grouped data regression where the left-hand side variable is the categorical age variable, the explanatory variable is the constant term only, and the interior limits are 18, 26, 35, 60, and 65. Estimating this model yields a mean (23.98) and variance (19.51) of the distribution. A z value that corresponds to the lower limit of the highest ranked (open-ended category) is estimated with the following equation:

$$Z = \frac{OEI_{LL} - Mean}{Standard\ Deviation}$$

Where OEILL represents the open-ended interval lower limit. Using our computed values, we find a z value equal to 2.5.We then use the censored normal results to find the mean of the category in terms of z using the following equation:

$$E[z] = \frac{\theta(z)}{1 - \Phi(z)} = x$$

where θ represents the standard normal density and Φ represents the standard normal distribution. Finally, we back out the midpoint value on the original data's scale by "unnormalizing" the result above,

$$\frac{M - Mean}{Standard\ Deviation} = x \rightarrow M = Mean + (x * Standard\ Deviation)$$

where M is the desired estimate of the category midpoint and is equal to 72.

the June 2003 CBSA classifications provided by the Office of Management and Budget. $FamilySize$ is a variable indicating the number of persons in the respondent's household.

$WorkingWoman$, is a dummy variable equal to one if the individual is female and works full time, zero otherwise. $Smoker$, a dummy variable only included in the regular drinking model, is equal to one if the respondent smokes more than 6 to 15 cigarettes (about ½ a pack) a day on average.

$PsychologicalDistress$, a dummy variable included only in the legal drug misuse model, was created from a "serious psychological distress indicator" coded by the CDC. Survey respondent's aged 18 or older were asked a series of questions about their psychological distress during the one month in the past year when they were at their worst emotionally (including feeling nervous, hopeless, restless or fidgety, sad or depressed, etc.) and given a score from 0 to 24. The respondents were then given a 0 (no serious psychological distress indicator) if they scored below a 13, and given a 1 (yes to serious psychological distress indicator) if they scored a 13 or above. $Risky$, a dummy variable only included in the illegal drug use model, was created from the survey question, "How often do you get a real kick out of doing things that are a little dangerous?" If the respondent answered "sometimes" or "always", they were given a 1 indicating they are risk takers. If the respondent answered "never" or "seldom", they were given a 0 indicating they are risk averse.

$Female$ is a dummy variable equal to one if the respondent is female. $WhiteCollar1$ is a variable equal to one if the respondent answered that their occupation is either a) Executive/Administrative/Managerial/Financial, b) Professional (not Education/Entertainment/Media), c) Education and Related Occupations, or d) Entertainers, Sports, Media, and Communications. $BlueCollar$ is a created dummy variable equal to one for respondents whose occupation is either a) Farming, Fishing, and Forestry Occupations, b) Installation, Maintenance, and Repair Workers, c) Construction Trades and Extraction Workers, d) Production, Machinery Setters/Operators/Tenders, or e) Transportation and Material Moving Workers. The omitted dummy variable, WhiteCollar2, is equal to one for individuals who work in either a) Technicians and Related Support Occupations, b) Sales Occupations, c) Office and Administrative Support Workers, d) Protective Service Occupations or e) Service Occupations, except Protective.

$OneEmployer$ is a dummy variable equal to one if the respondent only had one employer in the past 12 months. $SmallFirm$ is a dummy variable equal to one if, at the location where the individual works, less than 24 people work for the employer. $LargeFirm$ is equal to one if the respondent answered that more than 100 people work for the employer. The omitted dummy variable, MediumFirm, is equal to one if 25-99 people work for the respondent's employer. $SickRecently$ is a dummy variable equal to one if the respondent missed work for being sick or injured at least once in the past 30 days.

We are ultimately concerned with the results of our structural equations. Our first stage or reduced-form coefficients represent the effect of these factors on the equilibrium values of the dependent variables. The maximum likelihood probit estimates of the coefficients in the reduced form for the divorce, regular drinker, illegal drug use, and legal drug misuse equations, as well as the grouped data regression estimates of the income equation are presented in Table 7. While the behavioral implications of the reduced form estimates are interesting, they are not fundamental to the question at hand. The behavioral implications are more relevant, for our purposes, in terms of the structural estimates.

Our first stage divorce equation correctly predicts 91.2%, the regular drinker equation correctly predicts 66.3%, the illegal drug use equation correctly predicts 64.0%, and the legal drug use equation correctly predicts 74.6%.

The structural form estimates for the divorce, regular drinker, illegal drug use, legal drug misuse, and income equations are presented in Table 8.

$CollegeGraduate$ is positive and significant in the regular drinking and income models, while negative and significant in the other three models. In other words, college graduates are less likely to get divorced, use illegal drugs, and misuse legal drugs, but more likely to be regular drinkers and earn higher income. Looking at the marginal effects, we see that college graduates are 1.6% less likely to be divorced, 7.1% less likely to use illegal drugs and 4.2% less likely to misuse legal drugs, but 9.6% more likely to be regular drinkers. Graduating from college is also associated with a $12.6K increase in income.

Age and Age^2 are positive and negative, respectively, for the divorce, illegal drug use, legal drug misuse and income equations, but Age is insignificant in the legal drug misuse model. These coefficients yield an inverted U-shaped relationship between age and the probability of divorce, the probability of using illegal drugs, and income. As an individual's age increases, their income increases, as well as their probability of divorce and using illegal drugs. Once some critical age is reached, however, an increase in the individual's age leads to a decrease in their income, as well as their probabilities of divorce and using illegal drugs. The age coefficients for the legal drug misuse model imply that the probability of misusing legal drugs decreases as age increases. In the regular drinking model, *Age* is negative and significant, while $\text{Age}^2$is positive but insignificant. This implies that, as age increases, the probability of being a regular drinker decreases.

$White$ is positive and significant in all models except for the divorce model, in which it is negative and significant. This implies that individuals who are white are less likely to get divorced and more likely to use and abuse substances. Specifically, white individuals are 7.9% less likely to get divorced, 8.7% more likely to regularly drink, 8.9% more likely to misuse legal drugs, and 7.7% more likely to use illegal drugs. Being white is also associated with a $13.3K increase in income.

$Fulltime$ is positive in all five models, but it is insignificant in both the divorce and legal drug misuse models. This implies that individuals who work full-time are more likely to be regular drinkers and to have used illegal drugs. Specifically, individuals who work fulltime are 2% more likely to regularly drink and 3.9% more likely to have used illegal drugs. Working full time is also associated with a $7.3K increase in income.

$NonMSA$ is negative and significant in all models except for the divorce model, in which it is positive but insignificant. This implies that individuals who live in NonMSA (rural) areas are less likely to be regular drinkers, to use illegal drugs, and to misuse legal drugs. Specifically, individuals living in a NonMSA area are 4.3% less likely to be regular drinkers, 2.9% less likely to have used illegal drugs, and 2.7% less likely to have misused legal drugs. Living in an NonMSA area is also associated with a $7.8K decrease in income.

$FamilySize$ is negative and significant in the divorce, regular drinker, and illegal drug use models. It is positive and significant in the income model, but positive and insignificant in the legal drug misuse model. This implies that individuals with larger families are less likely to get divorced, to be regular drinkers, and to use illegal drugs. Specifically, individuals with larger families are 1.2% less likely to get divorced, 3.6% less likely to be regular drinkers, and 1.1%

less likely to have used illegal drugs. Individuals with larger families are associated with a $7.3K increase in income.

$WorkingWoman$ is positive and significant in the divorce model. Looking at the marginal effects, we see that working women are 3.8% more likely to be divorced. $Smoker$ is positive and significant in the regular drinker model. Specifically, smokers are 5.1% more likely to be regular drinkers. $Risky$ is positive and significant in the illegal drug use model, while $PsychologicalDistress$ is positive and significant in the legal drug misuse model. The marginal effects show that risky individuals are 10% more likely to have used illegal drugs, and psychologically distressed individuals are 9.4% more likely to have misused legal drugs.

All of the exogenous variables specific to the income model are significant. $Female$ is negative, $WhiteCollar1$ is positive, $BlueCollar$ is negative, $OneEmployer$ is positive, $SmallFirm$ is negative, $LargeFirm$ is positive, and $SickRecently$ is negative. In other words, being a white collar worker, having stable employment, and working at a large firm is associated with higher income, while being female, a blue collar worker, working for a small firm, and recently being sick or injured is associated with lower income.

$Divorce$ is negative and significant in the regular drinker and income model, and positive and significant in the illegal drug use model. As our results from the single equations showed, divorced individuals are less likely to be regular drinkers, more likely to use illegal drugs, and are associated with a decrease in income. The only coefficient for divorce that changes sign from the single equation model to the structural equation model is in the legal drug misuse model. While $Divorce$ was positive and significant in the single equation model for legal drug misuse, it is negative in the structural equation model. This coefficient, however, is only significant at the 10% level. The marginal effects show that divorced individuals are 10.3%

more likely to be regular drinkers, 8.6% less likely to have misused legal drugs, and 16% more likely to have used illegal drugs.

$RegularDrinker$ is positive and significant in all four models. In other words, regular drinkers are more likely to get divorced, use illegal drugs, and misuse legal drugs. They also are associated with a higher income. These signs are consistent with all of the single equation coefficients. Specifically, we see that regular drinkers are 1% more likely to get divorced, 5.7% more likely to have used legal drugs, and 13.5% more likely to have misused legal drugs.

$IllegalDrugUse$ is positive and significant in the divorce, regular drinker, and legal drug misuse models. In other words, individuals who use illegal drugs are more likely to divorce, to be regular drinkers, and to use misuse legal drugs. The marginal effects show that individuals who have used illegal drugs are 2.7% more likely to be divorced, 11.3% more likely to regularly drink, and 14.4% more likely to have misused legal drugs. Unlike the single equation results for the income model, we find a negative and significant coefficient in the structural form of the income model. In other words, using illegal drugs is associated with a decrease in income. This answers the questions we had about this positive coefficient for illegal drug use in our income chapter. Now, after accounting for simultaneity bias, we see that this positive relationship between illegal drug use and income no longer holds.

$LegalDrugMisuse$ is positive and significant in the divorce, regular drinker, and illegal drug use models. It is negative and significant in the income model. In other words, individuals who misuse legal drugs are more likely to divorce, to be regular drinkers, and to use illegal drugs. Specifically, individuals who have misused legal drugs are 2.7% more likely to get divorced, 8.6% more likely to be a regular drinker, and 27.6% more likely to use illegal drugs. Individuals who misuse legal drugs are also associated with a decrease in income.

$Income$ is negative and significant in the divorce, illegal drug use, and legal drug misuse models. It is positive, but insignificant in the regular drinker model. $Income^2$ is positive in the regular drinker, illegal drug use, and legal drug misuse models, although it is insignificant in the legal drug misuse model. It is negative and significant in the divorce model. These signs imply that as income increases the probability of being a regular drinker increases. As income increases, the probability of divorce decreases, and as income continues to increase, the probability of divorce decreases even more. This result is different from our single equation divorce result, which showed that as income continues to increase, the probability of divorce starts to increase. The illegal drug use results show that as income increases, the probability of using illegal drugs decreases to a point, but as income continues to increase, the probability of using illegal drugs increases. The legal drug misuse results show that as income increases the probability of misusing legal drugs decreases. The relationship between higher income and lower abuse of legal drugs is even more significant than it was in the single equation results.

Using beta-coefficient analysis (see Table 9), we are able to see which variables are the most important in determining divorce, regular drinking, legal drug misuse, illegal drug use and income.

We find that age and substance use (as a group) are the most important factors in determining divorce, followed by being a working woman, family size and income. For the effect of substance use on divorce, using illegal drugs is the most important factor, followed by legal drug misuse, then regular drinking.

For being a regular drinker, the most important factor is other substance use, followed closely by age, family size, and income. We see that illegal drug use predicts regular drinking more than legal drug misuse predicts regular drinking.

For predicting illegal drug use, we see that age, other substance use, and income are the most important factors.

In predicting legal drug misuse, we find that other substance use is the most important factor, with illegal drug use being the most important factor of all, followed by age, income, and psychological distress.

We find that age is the most important factor in determining income, followed by family size and then the personal problems discussed in chapters III and IV. Among those personal problems, regular drinking has the biggest effect, then legal drug misuse, illegal drug use, and divorce. While age and family size have the largest effect on income, these personal issues do have a substantial impact.

CHAPTER VII

CONCLUSION

The ultimate goal of this dissertation was to look at the determinants of income. Specifically, we wanted to determine if personal problems (such as a divorce or substance use) have a substantial impact on income. We suspected that these personal problems would affect both the demand for and supply of labor, by affecting an individual's productivity and labor/leisure preference, respectively. While we posited that the labor demand and labor supply would decrease, the effect on wages (and thereby income) was ambiguous. Therefore, in order to observe this effect on income, we needed to look at this problem empirically. We first, however, needed to look at these personal problems more closely to see what determined them.

The first personal problem we investigated was divorce. Looking at the literature on this subject and creating a model of our own, we found that substance use, education, race, age, family size, income, and wives' working are factors that predict divorce, with age, income, and substance use being the most important.

We next looked at factors that predict substance use, which we broke down into alcohol use (being a regular drinker), illegal drug use, and legal drug misuse. We found that using other substances is the most important factor in predicting substance use.

Using a grouped data regression, we found that individuals who are divorced, that misuse legal drugs, who live in a Non-MSA region, are over the age of 61, are female, have a blue collar job, work for a small firm, and have missed work recently are associated with lower income.

Also, individuals who are regular drinkers, that use illegal drugs, are college graduates, are white, who are older (but not over the age of 61), who work full time, have larger families, work white collar jobs, who have only worked for one employer in the past year, and who work for large firms are associated with higher income.

The only result from our income model that we could not rationalize was the positive effect of illegal drug use on income. Figuring that income and using illegal drugs may be simultaneously determined, we modeled a simultaneous equation system. This system also allowed for joint determination of other personal problems with themselves and with income, based on our earlier single equation results. Indeed, after accounting for simultaneity bias, we see that this positive relationship between illegal drug use and income no longer holds.

The use of standardized (beta) coefficients allows us to evaluate the relative importance of the various factors affecting personal problems and income. After accounting for simultaneity bias, we find that age and substance use (as a group) are the most important factors in determining divorce, followed by being a working woman, family size and income. For the effect of substance use on divorce, using illegal drugs is the most important factor, followed by legal drug misuse, then regularly drinking. For being a regular drinker, the most important factor is other substance use, followed closely by age, family size, and income. We see that illegal drug use affects regularly drinking more than legal drug misuse. For predicting illegal drug use, we see that age, other substance use, and income are the most important factors. In predicting legal drug misuse, we find that other substance use is the most important factor (with illegal drug use being the most important factor of all) followed by age, income, and psychological distress.

We find that age is the most important factor in determining income, followed by family size then these personal problems. With personal problems, regular drinking has the biggest effect, then legal drug misuse, illegal drug use, and divorce. It is important to note that while alcohol use might be a personal problem, our measure of alcohol use in this study (regularly drinking) is not “problematic”. Although these personal issues do not have the largest effect on income, they do have a substantial impact and should not be left out of wage or income determination models. Doing so may result in biased estimates of the usual determinants.

To investigate the extent of this potential bias, we ran our income model (using Grouped Data Regression as before) without including any of our personal problems. The results (shown in Table 10) yield all of the same signs and all variables are significant, however we see several changes in the magnitude of the coefficients. The magnitude of the bias resulting from excluding variables is determined by the correlation between the personal problem variables and the included variables. In other words, variables that are highly correlated with the personal problem variables will have a big difference in estimates.

While two variables (Age^2 and $LargeFirm$) have a decreased effect on income, most variables have an increased effect when personal problem variables are omitted. The change in the variables’ effect on income varies from small percentage change to moderate percentage changes. For example, $White$ has a coefficient of \$14,019.79 in our personal problem model and a coefficient of \$14,774.788 in the omitted model. In other words, the positive effect of being white on income increases by only 5% when personal problems are omitted. On the other hand, $Female$ has a coefficient of -\$4,451.125 in our model with personal problems but an increased coefficient -\$5,556.825 in our model that does not include the personal problems. In other words, omitting the personal problems yields around a 25% increase in the effect of being a woman on income. This example is an illustration of how omitting these personal problem variables may lead to imprecise estimates. Specifically with this example, omitting these personal problem variables may lead to a small overestimate of racial bias, but a potentially large overestimate of gender bias, in determining income.

These results strongly suggest that future microeconomic datasets should consider asking more questions on personal problems that an individual may have. While our data set used for this dissertation (which is typically used for analyzing drug use and health problems) has great information about personal problems, there are several areas of this dataset that are less than ideal. Due to privacy issues, there was particularly rough income data, location data, and age data.

One method of dealing with the lack of information in one particular data set is to use coefficients estimated from one data set and use them to create instrumental variables in another data set. This method, called Two Sample Two Stage Least Squares, could not be used for our problem at hand because there were not enough similar variables between data sets to construct appropriate instrumental variables.

Therefore, if microeconomic datasets that have great income data could include more information on personal problems, the effects of these personal problems (as well as the effects of typical determinants) on income could be analyzed more appropriately.

Because of the lack of important information in our data set, this work should be viewed as preliminary. However, the results from this dissertation are important enough to include these types of questions. There is a possibility that excluding these personal problems from wage determination models may bias the usual estimates of economic factors. If microeconomic surveys would start including information on personal problems, as well as good information about income, we could dive even further into this issue of what factors affect an individual's income.

REFERENCES

Amato, P. R. (2010, June). Research on Divorce: Continuing Trends and New Developments. *Journal of Marriage and Family, 72*, 650-666.

Amato, P. R., Booth, A., Johnson, D. R., & Rogers, S. J. (2007). Alone Together. *Age*, 232.

Amato, P. R., & Rogers, S. J. (1997). A Longitudinal Study of Marital Problems and Subsequent Divorce. *Journal of Marriage and the Family*, 612-624.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, by Gary S. Becker.* London.

Becker, G. S., Landes, E. M., & Michael, R. T. (1977). An Economic Analysis of Marital Instability. *Journal of Political Economy, 85*(6), 1141-1187.

Blanchflower, D. G. (1986). Wages and Concentration in Great Britain. *Applied Economics, 18*, 1025-1038.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 436-455.

Blundell, R. (1999). Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy. *Fiscal Studies, 20*(1), 1-23.

Bramlett, M. D., & Mosher, W. D. (2002). Cohabitation, Marriage, Divorce, and Remarriage in the United States. *National Center for Health Statistics, 23*(22), 1-32.

Brown, C. C., & Medoff, J. L. (1989). The Employer Size-Wage Effect.

Brown, G., & Silverman, L. P. (1980). The Retail Price of Heroin: Estimation and Applications. *Quantitative Explorations in Drug Abuse Policy*.

Chaloupka, F., & Laixuthai, A. (1994). Do Youths Substitute Alcohol and Marijuana? Some Econometric Evidence. *National Bureau of Economic Research Working Paper no. 4662*.

Caudill, S. B., & Jackson, J. D. (1993). Heteroscedasticity and Grouped Data Regression. *Southern Economic Journal*, 128-135.

DiNardo, J. (1993). Law Enforcement the Price of Cocaine and Cocaine Use. *Mathematical and Computer Modeling*, *17*(2).

DiNardo, J., & Lemieux, T. (1992). Alcohol, Marijuana, and American Youth: The Unintended Effects of Government Regulation. *National Bureau of Economic Research Working Paper no. 4212*.

Douglas, P. H. (1930). Real Wages in the United States, 1890-1926.

Dronkers, J., Kalmijn, M., & Wagner, M. (2006). Causes and Consequences of Divorce: Cross-national and Cohort Differences, an Introduction to this Special Issue. *European Sociological Review*, *22*(5), 479-481.

Dunn, L. F. (1986). Work Disutility and Compensating Differentials: Estimation of Factors in the Link between Wages and Firm Size. *The Review of Economics and Statistics, 68*(1), 67-73.

Ekelund, R.B., Jackson, J. D., Ressler, R. W., & Tollison, R. D. (2006). Marginal Deterrence and Multiple Murder. *Southern Economic Journal*, *72*(3), 521-541.

Ghez, G., & Becker, G. S. (1975). *The Allocation of Time and Goods over the Life Cycle.* NBER Books.

Gill, M. A., & Michaels, J. R. (1992). Does Drug Use Lower Wages? *Industrial and Labor Relations Review, 45*(3), 419-434.

Goldin, C., & Katz, L. F. (2007). *The Race Between Education and Technology.* Cambridge, MA: Belknap Press.

Grodsky, E., & Pager, D. (2001). The Structure of Disadvantage: Individual and Occupational Determinants of the Black-White Wage Gap. *American Sociological Review, 66*(4), 542-567.

Grossman, M., Coate, D., & Arluck, G. (1987). Price Sensitivity of Alcoholic Beverages in the United States: Youth Alcohol Consumption. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities*.

Heinen, D., & Pompelli, G. (1989). The Demand for Alcoholic Beverages: Economic and Demographic Effects. *Southern Journal of Economics* (55).

Heywood, J. S. (1986). Labor Quality and the Concentration-Earnings Hypotheses. *Review of Economics and Statistics*, *68*(2), 342-346.

Hout, M. (2012). Social and Economic Returns to College Education in the United States. *Annual Review of Sociology*, 379-400.

Jackson, J. D., & Jones, E. B. (1990). College Grades and Labor Market Rewards. *Journal of Human Resources*, 253-266.

Johnston, L., O'Malley, P., & Bachman, J. (1981). Marijuana Decriminalization: The Impact on Youth 1975-1980.

Kandel, D. B. (1980). Drug and Drinking Behavior Among Youth. *Annual Review of Sociology*, 235-285.

Kendell, R. E., De Roumanie, M., & Ritson, E. B. (1983). Influence of an Increase in Excise Duty on Alcohol Consumption and its Adverse Effects. *BMJ, 287*(6395), 809-811.

Kenkel, D. S. (1993). Drinking, Driving, and Deterrence: The Effectiveness and Social Costs of Alternative Policies. *Journal of Law and Economics*, 877-913.

Lazear, E. (1977). Academic Achievement and Job Performance: Note. *American Economic Review*, *67*(2), 252-254.

Lee, L.-F. (1978). Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *International Economic Review, 19*(2), 415-433.

Leung, S. F., & Phelps, C. E. (1993). My Kingdom for a Drink...? A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages. *Economics and the Prevention of Alcohol-Related Problems*, 1-32.

Mincer, J. (1974). Schooling, Experience and Earnings. *NBER*.

Model, K. (1992). Crime Violence and Drug Policy. *unpublished paper*.

Model, K. (1993). The Effect of Marijuana Decriminalization on Hospital Emergency Room Drug Episodes: 1975-1978. *Journal of the American Statistical Association, 88*(423).

Mullahy, J., & Sindelar, J. L. (1994, July). Alcoholism, Work, and Income Over the Life Cycle. *Journal of Labor Economics, 11*(3), 494-520.

Murnane, R. J., Willett, J. B., & Levy, F. (1995, May). The Growing Important of Cognitive Skills in Wage Determination. *The Review of Economics and Statistics, 77*(2), 251-266.

Nakonezny, P. A., Shull, R. D., & Rodgers, J. L. (1995). The Effect of No-Fault Divorce Law on the Divorce Rate Across the 50 States and its Relation to Income, Education, and Religiosity. *Journal of Marriage and the Family*, 477-488.

Nisbet, C., & Vakil, F. (n.d.). Some Estimates of Price and Expenditure Elasticities of Demand for Marijuana Among UCLA Students. *Review of Economics and Statistics, 54*, 1972.

Pacula, R. L. (1998). Does Increasing the Beer Tax Reduce Marijuana Consumption? *Journal of Health Economics, 17*(5), 557-585.

Saffer, H., & Chaloupka, F. (1998). *Demographic Differentials in the Demand for Alcohol and Illicit Drugs*. (No. w6432). National Bureau of Economic Research.

SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health (NSDUH), 2013.

Sander, W. (1985). Women, Work, and Divorce. *The American Economic Review, 75*(3), 519-523.

Schiller JS, Ward BW, Freeman G. Early release of selected estimates based on data from the 2013 National Health Interview Survey. National Center for Health Statistics. June 2014. Available from: http://www.cdc.gov/nchs/nhis.htm.

Schmidt, C. M., & Zimmermann, K. F. (1991). Work Characteristics, Firm Size and Wages. *The Review of Economics and Statistics*, 705-710.

Schoen, R., Astone, N. M., Kim, Y. J., Rothert, K., & Standish, N. J. (2002). Women's Employment, Marital Happiness, and Divorce. *Social Forces, 81*(2), 643-662.

Sickles, R., & Taubman, P. (1991). Who Uses Illegal Drugs? *The American Economic Review*, 248-251.

Silverman, L. P., & Spruill, N. L. (1977). Urban Crime and the Price of Heroin. *Journal of Urban Economics, 4*(1), 80-103.

Skinner, S. J., Austin, W. A., & Watson, J. K. (n.d.). An Examination of Legal and Illegal Substance Use and Work Absenteeism. *Working Paper.*

Slichter, S. H. (1950). Notes on the Structure of Wages. *The Review of Economics and Statistics*, 80-91.

Smith, A. (1937). *The Wealth of Nations [1776]* (p. 421). na.

Stevenson, B. (2007). The Impact of Divorce Laws on Marriage-Specific Capital. *Journal of Labor Economics, 25*(1), 75-94.

Thies, C. F., & Register, C. A. (1993). Decriminalization of Marijuana and the Demand for Alcohol, Marijuana, and Cocaine. *The Social Science Journal*, *30*(4), 385-399.

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Office of Applied Studies. National Survey on Drug Use and Health, 2006. ICPSR21240-v6. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-06-21. http://doi.org/10.3886/ICPSR21240.v6.

Van Ours, J. C. (1995). The Price Elasticity of Hard Drugs: The Case of Opium in the Dutch East Indies, 1923-1938. *Journal of Political Economy*, 261-279.

Willis, R. J. (1986). Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Function. *Handbook of Labor Economics*, 505-602.

Wise, D. A. (1975). Academic Achievement and Job Performance. *American Economic Review*, *65*(3), 350-366.

Table 1. Determinants of the Probability of Divorce

Variable	Probit Coefficients	T-Statistics	Marginal Effects
Constant	-4.322	-51.673	-.467
$\boldsymbol{RegularDrinker}$	.075	4.444	.008
$\boldsymbol{EverUsedIllegalDrugs}$	.215	8.973	.023
$\boldsymbol{EverMisusedLegalDrugs}$	.112	3.298	.013
$\boldsymbol{CollegeGraduate}$	-.189	-6.826	-.019
$\boldsymbol{Age}$	.161	39.517	.017
$\boldsymbol{Age^2}$	-.002	-33.485	-.0002
$\boldsymbol{White}$	-.046	-1.969	-.005
$\boldsymbol{Fulltime}$	.041	1.477	.004
$\boldsymbol{NonMSA}$	-.034	-.957	-.004
$\boldsymbol{FamilySize}$	-.108	-13.065	-.012
$\boldsymbol{FamilyIncome}$	$-.112\text{x}10^{-4}$	-8.325	$-.121\text{x}10^{-5}$
$\boldsymbol{FamilyIncome^2}$	$.287\text{x}10^{-10}$	2.651	$.310\text{x}10^{-11}$
$\boldsymbol{WorkingWoman}$	.292	10.941	.036

Table 2. Determinants of the Probability of Being a Regular Drinker

Variable	Probit Coefficients	T-Statistics	Marginal Effects
Constant	.235	4.386	.091
$\boldsymbol{EverMisusedLegalDrugs}$	.185	10.312	.071
$\boldsymbol{EverUsedIllegalDrugs}$	.716	45.808	.272
$\boldsymbol{Divorce}$	.015	.598	.006
$\boldsymbol{CollegeGraduate}$	.355	18.510	.133
$\boldsymbol{Age}$	-.028	-9.602	-.011
$\boldsymbol{Age^2}$	$.002\text{x}10^{-1}$	6.192	$.807\text{x}10^{-4}$
$\boldsymbol{White}$	.242	15.560	.094
$\boldsymbol{Fulltime}$	.207	13.717	.080
$\boldsymbol{NonMSA}$	-.151	-6.248	-.059
$\boldsymbol{FamilySize}$	-.099	-17.774	-.038
$\boldsymbol{FamilyIncome}$	$.485\text{x}10^{-5}$	5.359	$.188\text{x}10^{-5}$
$\boldsymbol{FamilyIncome^2}$	$-761\text{x}10^{-11}$	-1.048	$-.295\text{x}10^{-11}$
$\boldsymbol{Smoker}$	.115	5.922	.044

Table 3. Determinants of the Probability of Misusing Legal Drugs

Variable	Probit Coefficients	T-Statistics	Marginal Effects
Constant	-1.476	-22.837	-.427
$\boldsymbol{RegularDrinker}$	.198	11.359	.056
$\boldsymbol{EverUsedIllegalDrugs}$	1.032	59.548	.293
$\boldsymbol{Divorce}$	.097	3.455	.029
$\boldsymbol{CollegeGraduate}$	-.119	-5.683	-.033
$\boldsymbol{Age}$	.003	.878	.0009
$\boldsymbol{Age^2}$	$-.16\text{x}10^{-3}$	-3.759	$-.477\text{x}10^{-4}$
$\boldsymbol{White}$	.286	16.285	.080
$\boldsymbol{Fulltime}$	.026	1.548	.007
$\boldsymbol{NonMSA}$	-.106	-3.798	-.030
$\boldsymbol{FamilySize}$	$-.16\text{x}10^{-3}$	-.026	-.463x10-4
$\boldsymbol{FamilyIncome}$	$-.185\text{x}10^{-5}$	-1.846	-.536x10-6
$\boldsymbol{FamilyIncome^2}$	$.479\text{x}10^{-11}$	.596	.138x10-11
$\boldsymbol{PsychologicalDistress}$	.376	18.551	.119

Table 4. Determinants of the Probability of Using Illegal Drugs

Variable	Probit Coefficients	T-Statistics	Marginal Effects
Constant	-1.850	-30.909	-.738
$\boldsymbol{RegularDrinker}$	.687	45	.269
$\boldsymbol{LegalDrugMisuse}$	1.06	58.346	.390
$\boldsymbol{Divorce}$	.243	8.939	..096
$\boldsymbol{CollegeGraduate}$	-.171	-8.937	-.068
$\boldsymbol{Age}$	.074	23.070	.030
$\boldsymbol{Age^2}$	-.001	-26.423	-.0004
$\boldsymbol{White}$	.168	10.448	.067
$\boldsymbol{Fulltime}$	.076	4.863	.030
$\boldsymbol{NonMSA}$	-.053	-2.077	-.021
$\boldsymbol{FamilySize}$	-.024	-4.119	-.009
$\boldsymbol{FamilyIncome}$	-.289x10^{-5}	-3.051	-.115x10^{-5}
$\boldsymbol{FamilyIncome^2}$	.271x10^{-10}	3.599	.108x10^{-10}
$\boldsymbol{Risky}$	.318	18.048	.126

Table 5. Determinants of Income

Variable	GDR Coefficients	T Statistics
Constant	-31,940.115	-17.598
$\boldsymbol{Divorce}$	-12,228.875	-16.176
$\boldsymbol{RegularDrinker}$	5,907.227	12.843
$\boldsymbol{EverMisusedLegalDrugs}$	-2,650.527	-5.247
$\boldsymbol{EverUsedIllegalDrugs}$	1,076.293	2.299
$\boldsymbol{CollegeGraduate}$	13,147.148	22.529
$\boldsymbol{Age}$	1,693.252	17.401
$\boldsymbol{Age^2}$	-13.881	-11.452
$\boldsymbol{White}$	14,019.79	30.383
$\boldsymbol{Fulltime}$	7,198.220	14.206
$\boldsymbol{NonMSA}$	-8,032.063	-10.814
$\boldsymbol{FamilySize}$	6,777.026	41.640
$\boldsymbol{Female}$	-4,451.124	-9.641
$\boldsymbol{WhiteCollar1}$	8,900.53	15.570
$\boldsymbol{BlueCollar}$	-3,664.693	-6.480
$\boldsymbol{OneEmployer}$	5,642.006	12.699
$\boldsymbol{SmallFirm}$	-2,659.443	-5.081
$\boldsymbol{LargeFirm}$	4,374.067	7.514
$\boldsymbol{SickRecently}$	-2,751.616	-5.594

Table 6. Descriptive Statistics

Variable	Mean	Minimum	Maximum	Standard Deviation	Number of Cases[22]
Divorce	.088	0	1	.284	2,351
Regular Drinker	.634	0	1	.482	16,867
Legal Drug Misuse	.270	0	1	.444	7,190
Illegal Drug Use	.538	0	1	.499	14,314
Income	48,825.20	-16,376.38	121,899.93	33,967.66	36,965
CollegeGrad	.230	0	1	.421	6,114
Age	32.01	21.5	72	12.89	36.965
White	.670	0	1	.470	17,825
Fulltime	.748	0	1	.434	19,904
NonMSA	.085	0	1	.279	2,269
FamilySize	3.172	1	6	1.354	36,965
WorkingWoman	.338	0	1	.473	8,995
Smoker	.188	0	1	.391	4.998
Risky	.266	0	1	.442	7,069
PsychDistress	.140	0	1	.347	3,720
Female	.496	0	1	.500	13,191
WhiteCollar1	.252	0	1	.434	6,707
BlueCollar	.241	0	1	.428	6,412
OneEmployer	.620	0	1	.486	16,488
SmallFirm	.486	0	1	.486	12,943
LargeFirm	.283	0	1	.283	7,543
SickRecently	.233	0	1	.233	6,193

[22] For binary variables, this represents the number of individuals who have a "1" for their answer, i.e. the number of divorced individuals in the sample.

Table 7. Reduced Form Estimates

Exogenous Variable	Divorce	Regular Drinker	Legal Drug Misuse	Illegal Drug Use	Income
Constant	-4.228 (-36.362)	.276 (3.915)	-1.355 (-16.823)	-1.175 (-16.057)	-27.922.761 (-15.156)
CollegeGrad	-.230 (-6.814)	.300 (12.948)	-.042 (-1.757)	-.021 (-.943)	13,236.328 (22.515)
Age	.147 (26.837)	.007 (1.752)	.026 (5.862)	.062 (15.619)	1,551,795 (16.048)
Age2	-.001 (-21.774)	-.0002 (-4.186)	-.0004 (-7.703)	-.0009 (-17.269)	-12.885 (-10.674)
White	-.152 (-5.613)	.326 (18.386)	.307 (15.662)	.267 (15.043)	15,295.343 (32.850)
Fulltime	.123 (2.123)	.158 (5.220)	.053 (1.672)	.130 (4.325)	9,400.058 (12.072)
NonMSA	-.011 (-.258)	-.211 (-7.317)	-.205 (-6.468)	-.189 (-6.504)	-8,197.340 (-10.995)
FamilySize	-.168 (-17.509)	-.085 (-13.726)	-.024 (-3.644)	-.043 (-6.940)	6,823.946 (42.068)
WorkingWoman	.134 (1.934)	-.125 (3.227)	-.025 (-.607)	-.064 (-1.671)	-3,170.705 (-3.201)
Smoker	.427 (14.884)	.312 (13.999)	.541 (25.281)	.801 (34.624)	-6,793.046 (-12.360)
Risky	-.010 (-.332)	.378 (18.837)	.403 (20.627)	.430 (21.956)	3,699.907 (7.419)
PsychDistress	.215 (6.273)	.006 (.270)	.350 (14.713)	.219 (9.073)	-4,176.187 (-6.840)
Female	.209 (3.216)	-.088 (-2.615)	-.009 (-.260)	-.020 (-.605)	-2,271.814 (-2.585)
WhiteCollar1	-.022 (-.663)	.113 (5.008)	.020 (.844)	.0178 (.808)	8,946.252 (15.560)
BlueCollar	.100 (2.960)	-.048 (-2.144)	.034 (1.444)	-.043 (-1.927)	-3,692.226 (-6.453)
OneEmployer	-.098 (-3.216)	-.070 (-3.966)	-.091 (-4.971)	-.113 (-6.528)	5,616.919 (-6.453)
SmallFirm	-.019 (-.597)	-.063 (-3.066)	.037 (1.703)	-.002 (-.122)	-2,570.286 (-4.884)
LargeFirm	.010 (.300)	-.037 (-1.626)	-.0007 (-.028)	.019 (.846)	4,315.605 (7.375)
SickRecently	.069 (2.396)	-.036 (-1.884)	.104 (5.177)	.110 (5.699)	-2,539.907 (-5.128)
Chi-Squared Statistic	2,701.205	2,330.501	2,582.986	3,485.085	7,309.84

Table 8. Structural Form Estimates

Variable	Divorce	Regular Drinker	Legal Drug Misuse	Illegal Drug Use	Income
Constant	-4.427 (-40.267)	.525 (7.935)	-1.099 (-14.016)	-1.254 (-17.577)	-30,923.551 (-16.491)
Divorce		-.267 (-1.698)	-.297 (1.760)	.427 (2.342)	-9,620.646 (-2.336)
Regular Drinker	.102 (2.422)		.184 (5.937)	.340 (12.816)	8,128.152 (11.255)
Legal Drug Misuse	.284 (6.538)	.242 (6.517)		.776 (20.603)	-5,793.330 (-7.277)
Illegal Drug Use	.250 (7.913)	.305 (14.779)	.452 (21.320)		-2,430.932 (-4.611)
Income	-.338x10-5 (-2.820)	.107x10-5 (1.377)	-.182x10-5 (-2.231)	-.316x10-5 (-4.108)	
$Income^2$	-.385x10-10 (-3.493)	.314x10-10 (4.334)	.752x10-11 (.988)	.365x10-10 (5.125)	
CollegeGrad	-.161 (-4.959)	.265 (12.291)	-.135 (-6.004)	-.179 (-8.628)	12,604.416 (21.142)
Age	.155 (27.737)	-.016 (-4.132)	.003 (.729)	.059 (15.104)	1,450.504 (14.744)
Age2	-.001 (-21.385)	.570x10-4 (1.208)	-.0001 (-2.283)	-.0008 (-16.307)	-11.237 (-9.039)
White	-.169 (-5.176)	.185 (9.493)	.190 (8.439)	.163 (7.989)	13,333.585 (24.290)
Fulltime	.045 (1.214)	.053 (2.722)	.029 (1.379)	.099 (5.177)	7,328.947 (14.282)
NonMSA	.015 (.339)	-.113 (-3.913)	-.092 (-2.897)	-.068 (-2.343)	-7,796.569 (-10.309)
FamilySize	-.109 (-10.628)	-.098 (-14.961)	-.009 (1.262)	-.027 (-4.177)	7,284.138 (42.784)
WorkingWoman	.327 (11.810)				
Smoker		.139 (5.087)			
Risky				.279 (13.563)	
PsychDistress			.277 (11.574)		
Female					-4,746.291 (-9.904)
WhiteCollar1					8,553.718 (14.819)
BlueCollar					-3,619.682 (-6.339)
OneEmployer					5,635.768 (12.435)

Table 8. Structural Form Estimates (Continued)

Variable	Divorce	Regular Drinker	Legal Drug Misuse	Illegal Drug Use	Income
SmallFirm					-2,188.227 (-4.145)
LargeFirm					4,613.899 (7.869)
SickRecently					-2,122.962 (-4.229)
Chi-Squared Statistic	2,847.333	2,208.445	1,947.592	2,622.167	7,252.96

Table 9. Marginal Effects and Beta Coefficients

Variable	Divorce	Regular Drinker	Legal Drug Misuse	Illegal Drug Use	Income
Constant	-.475	.196	-.353	-.497	
Divorce		-.103 -.013	-.086 -.015	.161 .021	-.016
Regular Drinker	.010 .041		.057 .073	.135 .135	.104
Legal Drug Misuse	.037 .078	.086 .067		.276 .214	-.052
Illegal Drug Use	.027 .125	.113 .152	.144 .226		.039
Income	-.362x10-6 -.115	.398x10-6 .036	-.586x10-6 -.062	-.125x10-5 -.107	
$Income^2$	-.413x10-11 -.139	.117x10-10 .114	.242x10-11 .027	.145x10-10 .132	
CollegeGrad	-.016 -.068	.096 .111	-.042 -.057	-.071 -.075	.172
Age	.017 1.995	-.006 -.200	.001 .003	.024 .765	.606
Age2	-.0002 -1.485	.212x10-4 .058	-.418x10-4 -.133	-.0003 -.837	-.372
White	-.019 -.079	.070 .087	.060 .089	.065 .077	.203
Fulltime	.005 .020	.020 .023	.009 .012	.039 .043	.103
NonMSA	.002 .004	-.043 -.032	-.029 -.026	-.027 -.019	-.071
FamilySize	-.012 -.147	-.036 -.132	.003 .012	-.011 -.037	.320
WorkingWoman	.038 .155				
Smoker		.051 .054			
Risky				.109 .123	
PsychDistress			.094 .096		
Female					-.077
WhiteCollar1					.139
BlueCollar					-.050
OneEmployer					.089

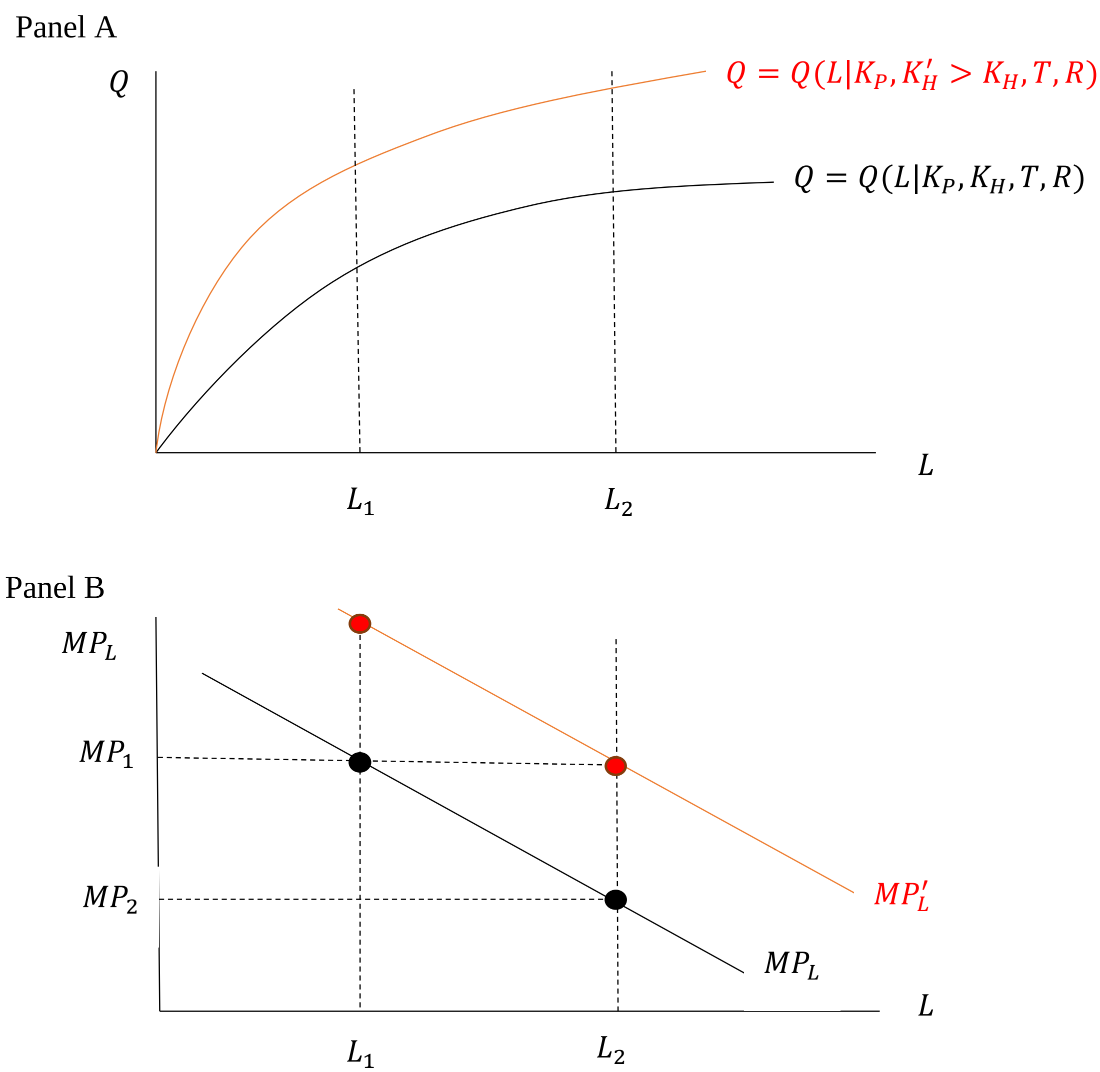
Table 9. Marginal Effects and Beta Coefficients (Continued)

Variable	Divorce	Regular Drinker	Legal Drug Misuse	Illegal Drug Use	Income
SmallFirm					-.035
LargeFirm					.067
SickRecently					-.029

Table 10. Comparison of Income Coefficients with and without Personal Problem Variables

Exogenous Variable	GDR Coefficients with Personal Problems Included	GDR Coefficients without Personal Problems
Constant	-31,940.115 (-17.598)	-24,963.117 (-13.953)
CollegeGrad	13,147.148 (22.529)	14,289.006 (24.355)
Age	1,693.252 (17.401)	1,456.491 (15.083)
Age2	-13.881 (-11.452)	-11.821 (-9.787)
White	14,019.79 (30.383)	14,774.788 (32.237)
Fulltime	7,198.220 (14.206)	7,216.667 (14.095)
NonMSA	-8,032.063 (-10.814)	-8,324.026 (-11.106)
FamilySize	6,777.026 (41.640)	6,935.219 (42.607)
Female	-4,451.124 (-9.641)	-5,556.825 (-11.966)
WhiteCollar1	8,990.53 (15.570)	9,198.175 (15.926)
BlueCollar	-3,664.693 (-6.480)	-4,011.951 (-7.009)
OneEmployer	5,642.006 (12.699)	5,738.453 (12.795)
SmallFirm	-2,659.443 (-5.081)	-2,672.747 (-5.051)
LargeFirm	4,374.067 (7.514)	4,283.704 (7,280)
SickRecently	-2,751.616 -5.594	-2,995.808 (-6.034)

Figure 1. Total Product, Marginal Product, and the Demand for Labor



Panel C

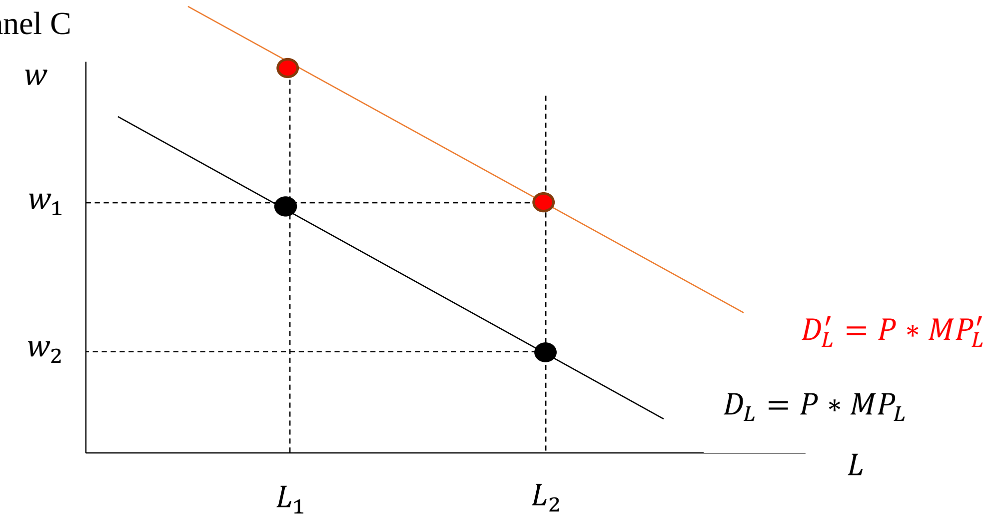
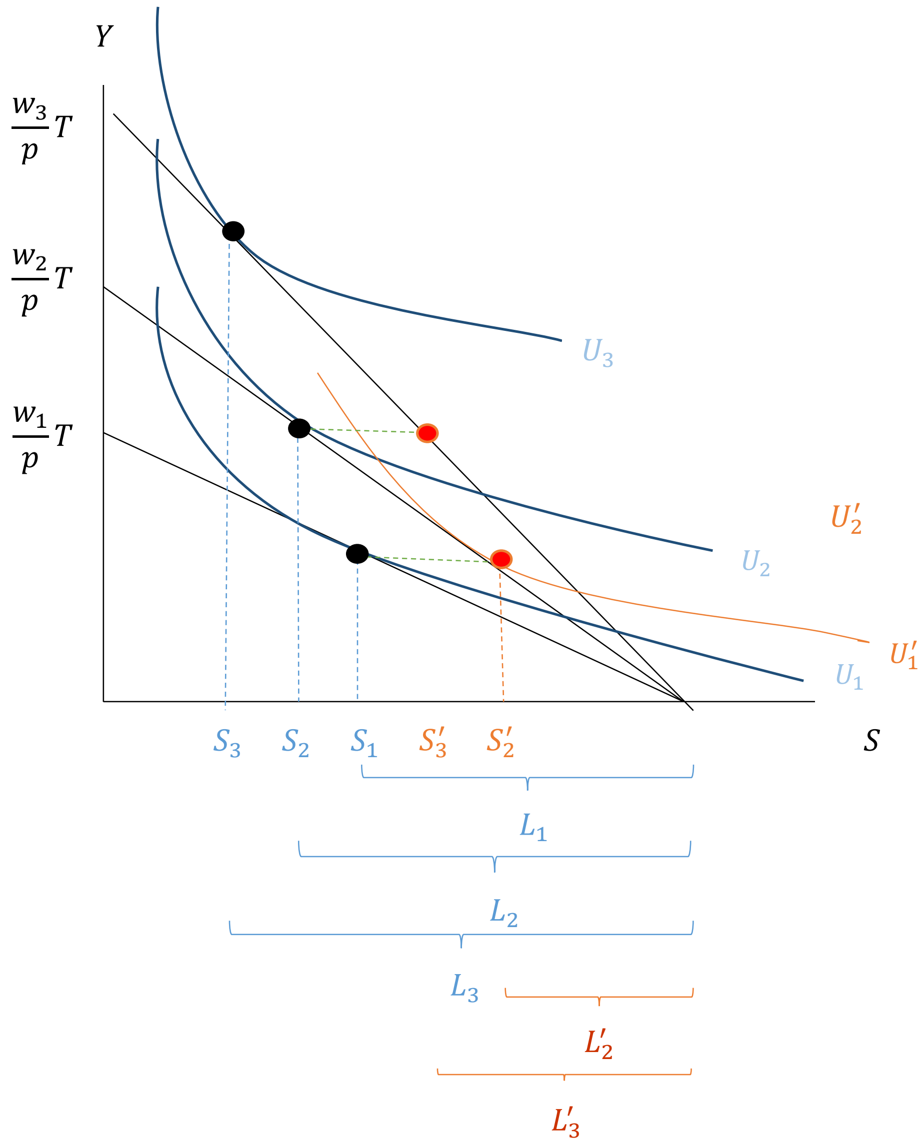


Figure 2. Labor/Leisure Choice and the Supply of Labor

Panel A



Panel B

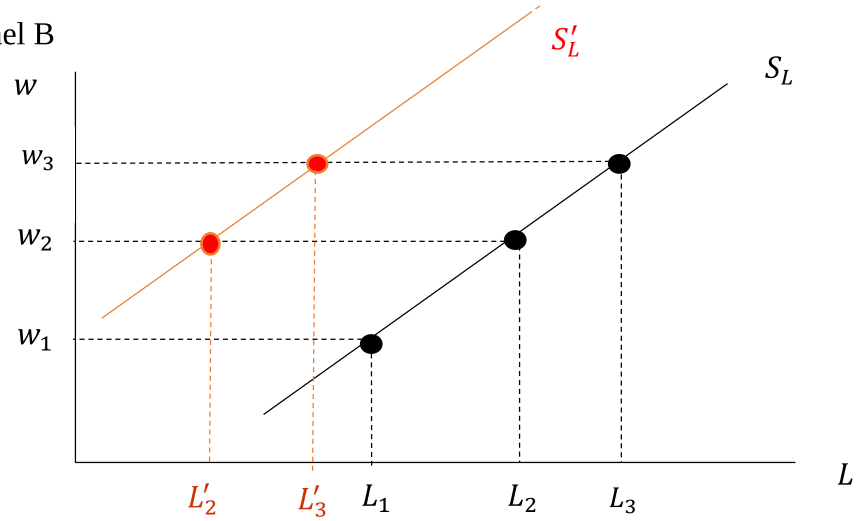


Figure 3. A Change in the Demand for Labor

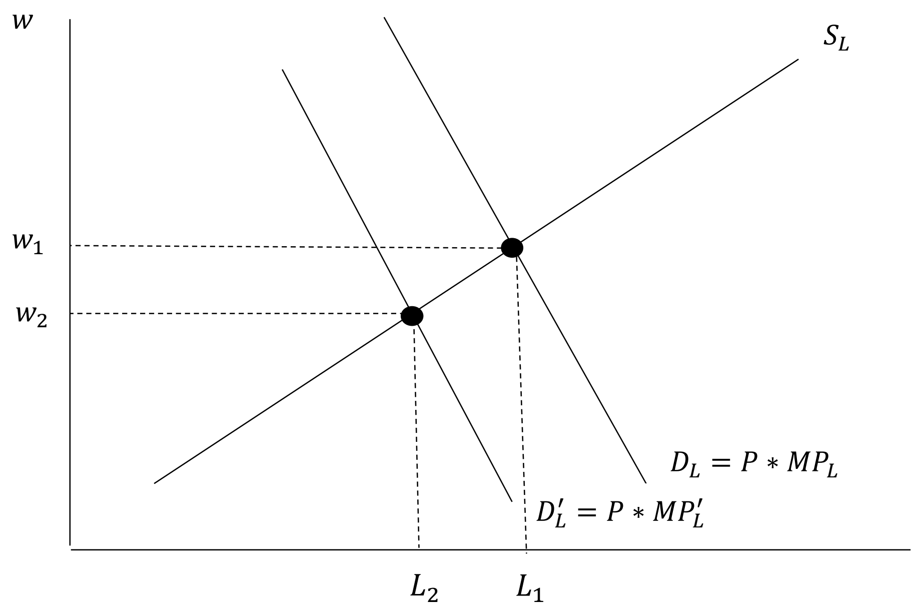


Figure 4. A Change in the Supply of Labor

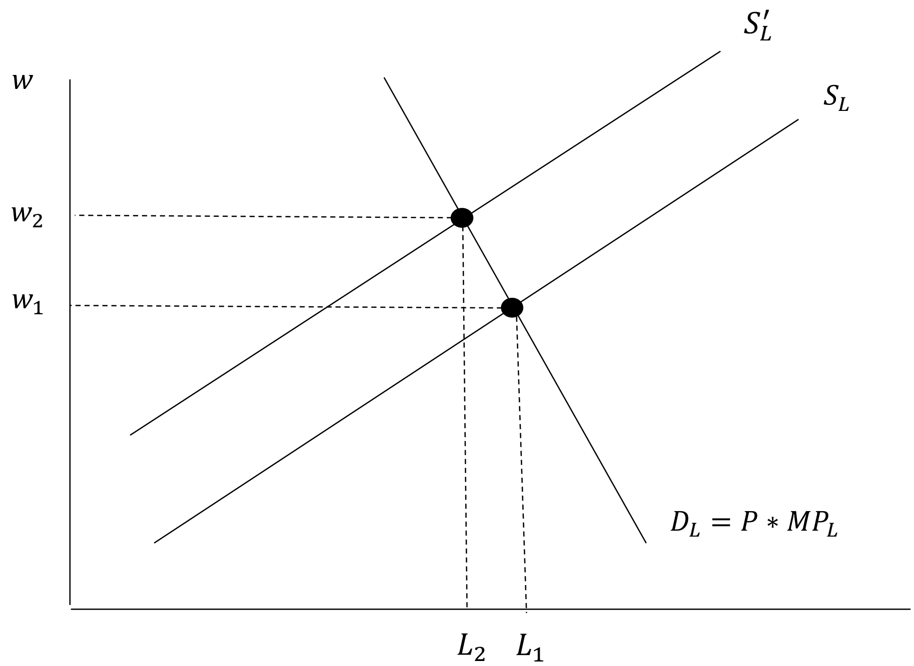
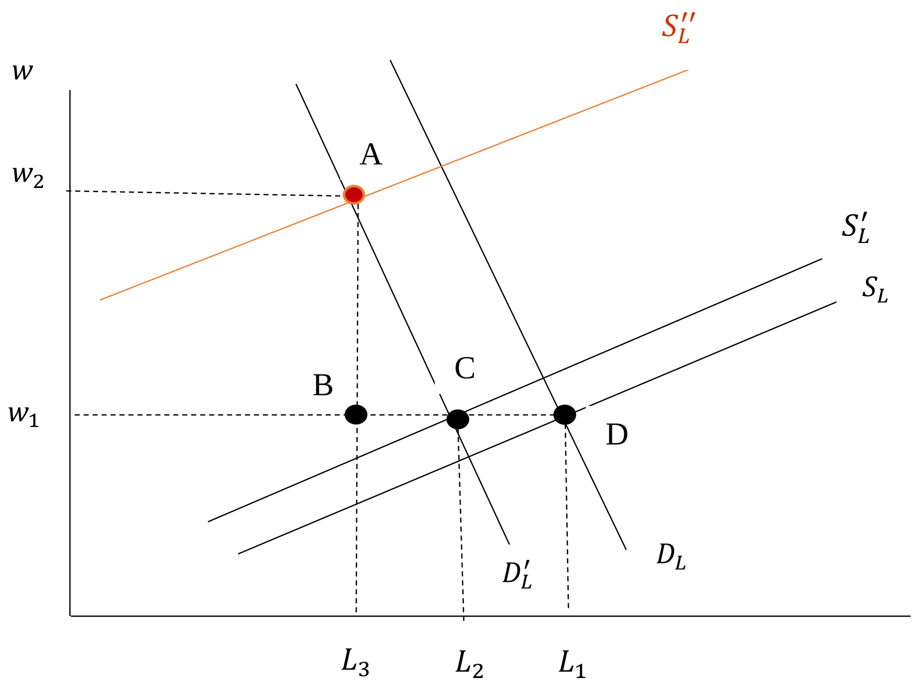


Figure 5. A Change in both the Demand and Supply of Labor



APPENDIX

Table A. Divorce Results including Standardized Coefficients and Marginal Effects

Variable	Probit Coefficients	T-Statistics	Marginal Effects	Standardized Probit Coefficients	Standardized Marginal Effects
Constant	-4.322	-51.673	-.467		
RegularDrinker	.075	4.444	.008	.037	.004
IllegalDrugUse	.215	8.973	.023	.107	.012
LegalDrugMisuse	.112	3.298	.013	.049	.005
CollegeGrad	-.189	-6.826	-.019	-.076	-.008
Age	.161	39.517	.017	2.437	.263
Age^2	-.002	-33.485	-.0002	-2.020	-.218
White	-.046	-1.969	-.005	-.022	-.002
Fulltime	.041	1.477	.004	.020	.002
NonMSA	-.034	-.957	-.004	-.0096	-.001
FamilySize	-.108	-13.065	-.012	-.148	-.016
FamilyIncome	$-.112\text{x}10^{-4}$	-8.325	$-.121\text{x}10^{-5}$	-.389	-.042
$FamilyIncome^2$	$.287\text{x}10^{-10}$	2.651	$.310\text{x}10^{-11}$	.122	.013
Working Woman	.292	10.941	.036	.126	.014

Table B. Regular Drinker Results including Standardized Coefficients and Marginal Effects

Variable	Probit Coefficients	T-Statistics	Marginal Effects	Standardized Probit Coefficients	Standardized Marginal Effects
Constant	.235	4.386	.091		
LegalDrugMisuse	.185	10.312	.071	.081	.031
IllegalDrugUse	.716	45.808	.272	.358	.139
Divorce	.015	.598	.006	.004	.002
CollegeGrad	.355	18.510	.133	.142	.055
Age	-.028	-9.602	-.011	-.420	-.163
Age^2	$.002x10^{-1}$	6.192	$.807x10^{-4}$	.270	.105
White	.242	15.560	.094	.115	.044
Fulltime	.207	13.717	.080	.103	.040
NonMSA	-.151	-6.248	-.059	-.043	-.017
FamilySize	-.099	-17.774	-.038	-.136	-.053
FamilyIncome	$.485x10^{-5}$	5.359	$.188x10^{-5}$	.169	.065
$FamilyIncome^2$	$-761x10^{-11}$	-1.048	$-.295x10^{-11}$	-.032	-.013
Smoker	.115	5.922	.044	.045	.017

Table C. Legal Drug Misuse Results including Standardized Coefficients and Marginal Effects

Variable	Probit Coefficients	T-Statistics	Marginal Effects	Standardized Probit Coefficients	Standardized Marginal Effects
Constant	-1.476	-22.837	-.427		
RegularDrinker	.198	11.359	.056	.097	.028
IllegalDrugUse	1.032	59.548	.293	.516	.149
Divorce	.097	3.455	.029	.027	.008
CollegeGrad	-.119	-5.683	-.033	-.048	-.014
Age	.003	.878	.0009	.047	.014
Age^2	$-.16x10^{-3}$	-3.759	$-.477x10^{-4}$	-.214	-.062
White	.286	16.285	.080	.136	.039
Fulltime	.026	1.548	.007	.013	.004
NonMSA	-.106	-3.798	-.030	-.030	-.009
FamilySize	$-.16x10^{-3}$	-.026	-.463x10-4	-.0002	-.00006
FamilyIncome	$-.185x10^{-5}$	-1.846	-.536x10-6	-.065	-.019
$FamilyIncome^2$	$.479x10^{-11}$	.596	.138x10-11	.020	.006
PsychDistress	.376	18.551	.119	.134	.039

Table D. Illegal Drug Use Results including Standardized Coefficients and Marginal Effects

Variable	Probit Coefficients	T-Statistics	Marginal Effects	Standardized Probit Coefficients	Standardized Marginal Effects
Constant	-1.850	-30.909	-.738		
RegularDrinker	.687	45	.269	.339	.131
LegalDrugMisuse	1.06	58.346	.390	.464	.179
Divorce	.243	8.939	..096	.069	.027
CollegeGrad	-.171	-8.937	-.068	-.069	-.027
Age	.074	23.070	.030	1.12	.436
Age^2	-.001	-26.423	-.0004	-1.13	-.516
White	.168	10.448	.067	.080	.031
Fulltime	.076	4.863	.030	.038	.015
NonMSA	-.053	-2.077	-.021	-.015	-.006
FamilySize	-.024	-4.119	-.009	-.033	-.013
FamilyIncome	$-.289x10^{-5}$	-3.051	$-.115x10^{-5}$	-.100	0.039
$FamilyIncome^2$	$.271x10^{-10}$	3.599	$.108x10^{-10}$	.116	.045
Risky	.318	18.048	.126	.138	.053