Three Essays on Econometric Modeling and Application: Health and Consumer Behaviors

Namhoon Kim

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics, Agricultural and Life Sciences

Travis P. Mountain, Chair

Klaus Moeltner

Kimberly L. Morgan

Hongxiao Zhu

March 14, 2018

Blacksburg, Virginia

Keywords: Non-Linear Regression, Vaccination, Paid Sick Leave, Financial Knowledge, Financial Behaviors, HPV, Papanicolaou (Pap) Smear

Three Essays on Econometric Modeling and Application: Health and Consumer Behaviors

Namhoon Kim

(Academic Abstract)

In the three chapters of my dissertation, I analyze the individual behaviors including health (vaccination and preventive care) and consumer (financial literacy) behaviors and the corresponding interventions by nonlinear econometric modeling.

In the first chapter, I suggest an appropriate econometric model that investigates the effect of paid sick leave on workers' decision to receive the seasonal flu vaccination. For this investigation, I apply a Bayesian non-linear structural regression model with one-outcome and two-endogenous equations. The results of my estimation indicate that having paid sick leave affects workers' vaccination decisions differently based on their income levels. Low-income workers are willing to be vaccinated because they perceive the high cost of claiming paid sick leave. However, high-income workers are willing to be vaccinated because paid sick leave reduces the cost of vaccination for seasonal flu.

In the second chapter, I suggest new econometric regression models that investigate the effect of "Don't Know" or "Refuse" (DK/RF) responses on parameter identification. I estimate the effect of group characteristics and financial education on the level of young respondents' objective financial knowledge and find the actual effects and biases by my suggested models. This study examines six questions about personal finance and selects covariates in the 2015 National Financial Capability Study (NFCS). Because these questions include DK/RF responses, a simple regression model that does not consider DK/RF responses could lead to misleading conclusions, such as gender/income difference and educational effectiveness in schools.

In the last chapter, I investigate the effect of three health-related interventions including a doctor's recommendation, information about human papillomavirus (HPV), and HPV vaccination, on the misuse of cervical cancer screening including too-early screening, unnecessary HPV test, annual Pap test, and no Pap smear that are not recommended for women younger than 30 years. I examine the National Health Interview Survey conducted in 2015 and applies binary and multinomial logistic regression models. From the estimation result, I observe that doctor's recommendation plays a significant role in increasing the probability of receiving cervical cancer screening while it induces the too-early screening, unnecessary HPV testing, and overuse of Pap smears.

Three Essays on Econometric Modeling and Application: Health and Consumer Behaviors

Namhoon Kim

(General Audience Abstract)

In the three chapters of my dissertation, I analyze the individual behaviors including health (vaccination and preventive care) and consumer (financial literacy) behaviors and the corresponding interventions.

In the first chapter, I investigate the effect of paid sick leave on workers' decision to receive the seasonal flu vaccination. From this investigation, I observe that low-income workers are willing to be vaccinated because they perceive the high cost of claiming paid sick leave while high-income workers are willing to be vaccinated because paid sick leave reduces the cost of vaccination for seasonal flu.

In the second chapter, I investigate the effect of "Don't Know" or "Refuse" (DK/RF) responses that are commonly ignored in many studies. From this investigation, I observe the disadvantage of rejecting financial education offers in obtaining proper financial knowledge and the ineffectiveness of formal and informal financial education on the level of objective financial knowledge. In addition, I observe few or no gender, income, and age differences in the level of objective financial knowledge among young adults.

In the last chapter, I investigate the effect of a doctor's recommendation, information about human papillomavirus (HPV), and HPV vaccination on the too-early screening, unnecessary HPV test, annual Pap test, and no Pap smear that are not recommended for women younger than 30 years. From this investigation, I observe that doctor's recommendation plays a significant role in increasing the probability of receiving cervical cancer screening while it induces the too-early screening, unnecessary HPV testing, and overuse of Pap smears.

Acknowledgments

I would never have been able to finish my dissertation without the support from my wife and children, Yeji and Yena, and guidance of my adviser and committee members.

I would like to thank my wife and best friend, Yeonsook, for the countless sacrifices she has made in helping me achieve this goal. She was always there and stood by me through the good and bad times.

I would like to express my gratitude to my advisor, Dr. Mountain, for his excellent guidance, tremendous academic support, and providing me many wonderful opportunities to enhance my research. I would also like to thank Dr. Moeltner, Dr. Morgan, and Dr. Zhu for all of their guidance and support.

Contents

List of Tables

List of Figures

Chapter 1

Paid Sick Leave, Vaccination, and Inequality: Bayesian Structural Equation Approach

1.1 Introduction

Although people tend to disregard flu infection as a mild infectious disease, its severity and mortality cannot be ignored. The World Health Organization (WHO) estimated that the flu epidemic results in about 3 to 5 million cases of severe illness, and about $250,000$ to $500,000$ annual deaths worldwide (WHO 2016). Annual vaccinations are the most effective way to prevent seasonal flu, and the sufficient vaccination coverage protects the high-risk population, such as children and older adults with chronic disease, from flu infection. However, global targets for vaccination coverage rate for this purpose have not been satisfied in most countries, particularly, for those aged 65 and older (OECD 2017; Palache et al. 2015). In the United States, the Centers for Disease Control and Prevention (CDC) estimated that there were approximately $12,000$ flu-associated deaths during the 2015-2016 flu season (Rolfes et al. 2016). However, at the same season, only 41.7% of adults were vaccinated (CDC 2015). This insufficiency may result in flu transmission to the high-risk population, such as children and older adults with chronic diseases (Shim et al. 2012).

Several models have been used to explain the insufficient vaccination coverage based on the gap between infectious disease and human behaviors (Funk, Salathé, and Jansen 2010; Verelst, Willem, and Beutels 2016). First, game theory blames selfish individuals who enjoy positive externalities without cost (Bauch, Galvani, and Earn 2003; Bauch and Earn 2004; Yamin and Gavious 2013). The game theory posits that free riders prevent society from achieving the optimal vaccination coverage rate. Second, the Health Belief Model (HBM) attributes insufficient coverage to poor perceptions of susceptibility and severity (Rosenstock 1974). This model suggests that weak perceptions of the risks related to flu result in failure to reach a sufficient vaccination rate. In the present study, I focus on economic factors, such as the cost and benefits of vaccination (Brito, Sheshinski, and Intriligator 1991). In this model, I regard the lack of financial benefit or accessibility to vaccination as a significant factor in the insufficient vaccination rate.

The lack of economic benefit leads to presenteeism, which is the cost associated with workers who are present in a workplace while suffering from diseases, which bears high costs to both employees and employers (Liao et al. 2012). If these costs are not covered by an employer, a worker with flu-like symptoms will be more willing to work to avoid losing one's pay. This behavior could lead to further flu endemics in the workplace (Kumar et al. 2013).

One of the economic interventions used to address this problem is paid sick leave. Paid sick leave is a paid absence from work due to sickness or disability. Also, for flu infections, simulation, and observational researchers have shown that, if paid sick leave is available and covers the potential loss of workers' income, it prevents them from spreading severe flu infections and the loss of workplace (Lovell 2004; Liao et al. 2012; DeRigne, Stoddard-Dare, Collins, et al. 2017). However, other researchers have argued that if a company offers paid sick leave, it might suffer financial hardship by paying for absent workers who are not truly ill. This ultimately could reduce workers' benefits and undermine their job because of the high cost of employing them (Drago and Lovell 2011; Colla et al. 2014; Nelsen 2014).

Further, it is also unclear whether offering paid sick leave affects workers' vaccination decisions. To the best of my knowledge, only one study has analyzed the relationship between paid sick leave and vaccination decisions (Wilson, Wang, and Stimpson 2014). In this study, the authors suggested first that workers with paid sick leave were likely to be vaccinated and second that paid sick leave had a positive effect on the economy and healthcare system. However, the authors did not focus on any differential benefit, such as who enjoys this benefit of paid sick leave for vaccination or who does not. Thus, I assumed that paid sick leave influenced the vaccination decision in two ways: reduced cost of receiving a vaccination now and the expected benefit of claiming paid sick leave after flu infection. I analyzed the differential benefit by identifying these two decisions.

In addition, the use of paid sick leave can be extended to include access to preventive healthcare. To my knowledge, two papers argued that whether workers have paid sick leave or not determines the frequency of preventive healthcare use (DeRigne, Stoddard-Dare, Collins, et al. 2017; Peipins et al. 2012). The authors empirically analyzed the National Health Interview survey dataset and concluded that the lack of paid sick leave could be a potential barrier to access to preventive healthcares, in particular, including mammography and Pap testing that are closely related to women's health. Thus, I can extend my analysis to a relationship between paid sick leave and the use of preventive heath care.

Thus, the purpose of this study is to investigate the differential benefit of paid sick leave depending on income levels, as well as the effect of paid sick leave on the decision to obtain a vaccination. I construct an endogenous structural regression model with an instrumental variable. I then estimate the marginal effects of the reduced cost and expected benefit in subgroups classified by income levels. Based on the estimation results, I investigate and discuss which subgroup benefit most from paid sick leave and which does not.

In the following sections, I introduce a theoretical model of the vaccination decision by applying the expected utility framework. Next, I describe the analysis of the survey data

and present the econometric model used in this study. In the results and discussion sections, I present and discuss my estimation results, and follow with several key conclusions and recommendation for healthcare policymakers.

1.2 Vaccination Decision Model

I assume that an individual will decide to obtain a vaccination if the utility from vaccination is greater than that from remaining unvaccinated. Based on this assumption, I propose a vaccination model with an expected utility framework (Brito, Sheshinski, and Intriligator 1991). Let $u_i(y_i)$ be the utility of income y_i of individual i with the properties given by:

$$u_i'(y_i) > 0 \text{ and } u_i''(y_i) < 0 \tag{1.1}$$

Then, the utility if an individual i accepts vaccination ($T_i = 1$), is given by:

$$v_i(y_i^H, \theta_i | T_i = 1) = u_i(y_i^H) - \theta_i \tag{1.2}$$

where y_i^H is the income of a healthy individual and θ_i are additive parameters that represent the individual cost of vaccination. On the other hand, let $p(z)$ be the perceived probability of flu infection without vaccination[1] (PPFI), where z is a parameter that affect this probability. Then, the expected utility if individual i rejects vaccination ($T_i = 0$) is given by:

$$v_i(y_i, z_i, q_i | T_i = 0) = u_i(y_i^H) \cdot [1 - p_i(z_i)] + u_i(y_i^I) \cdot p_i(z_i) \tag{1.3}$$

[1]The use of $p(z)$ is justifiable for my individual vaccination model. In epidemiology, $p(\phi)$ that is derived from various epidemic models was used rather than $p(z)$ for infectious disease model. $p(\phi)$ denotes "the probability that an unvaccinated individual will eventually be infected if the vaccine coverage level in the population is ϕ" (Bauch and Earn 2004). However, individuals do not know the actual or epidemiological coverage level. Thus, I assume that they are vaccinated based on the perception of the probability rather than the epidemically calculated rate.

where y_i^l is the income of an infected individual with the condition that $y_i^I \leq y_i^H$.

Let the excess utility of being vaccinated over unvaccinated be the gap between the utility of being vaccinated in (1.2) and unvaccinated in (1.3) and other relevant factors that influence the utilities (ϵ_i), given by:

$$\xi_i(y_i, z_i, \rho_i, \theta_i) = -\theta_i + [u_i(y_i^H) - u_i(y_i^I)] \cdot p_i(z_i) + \epsilon_i \tag{1.4}$$

[2]Thus, I could conclude that if the excess utility is greater than zero, this individual will be willing to be vaccinated. On the other hand, s/he will reject vaccination if the excess utility is less than zero. Furthermore, consider paid sick leave as a policy, represented by s. I assumed that paid sick leave will affect individuals' vaccination decisions through an additive cost parameter, θ_i, and the income of an infected worker, $y_i^I(x)$[3]. Thus, the excess

[2]This excess of the utility assumes that vaccine efficacy is 100%. Vaccine efficacy (VE) is estimated in a laboratory, as follows:

$$VE = (1 - RR) \times 100 \tag{1.5}$$

where RR is the relative risk of disease among the vaccinated group calculated by:

$$RR = \frac{ARV}{ARU} \tag{1.6}$$

where ARV and ARU are the unvaccinated and vaccinated study cohorts. Vaccine effectiveness is similar to vaccine efficacy. This is estimated by evaluating a set of infected cases, not in a laboratory (Weinberg and Szilagyi 2010). If the perceived effectiveness of vaccine is a revealed variable for vaccine efficacy rate, then, the excess of utility in (1.4) is rewritten by:

$$\xi_i(y_i, z_i, \rho_i, \theta_i) = -\theta_i + (1 - \rho_i) \cdot [u_i(y_i^H) + u_i(y_i^I)] \cdot p_i(z_i) \cdot p_i^e \tag{1.7}$$

where p_i^e is the perceived effectiveness of the vaccine. This study ignores the effect of the perceived effectiveness of the vaccine. Both $p_i(z_i)$ and p_i^e are not continuous and it is difficult to construct the model. One justification for ignorance of p_i^e is that most individuals in the sample used in this study show high or very high p_i^e and perceive vaccination to be very effective. Thus, I assume that they perceive the vaccination to be 100% effective, that is, $p_i^e = 1$. Instead, this variable is added to the model as one of the covariates for outcome equation.

[3]The assumption that y_i^I is a function of the policy is reasonable. For example, I can construct the income of an infected worker as follows:

$$y_i^I = y_i^H - L_i(s) \tag{1.8}$$

This construction shows that the income of an infected worker consists of the income of a healthy individual, y_i^H, and an individual loss function of L_i. This loss function represents a comprehensive loss induced by disease infection. It includes direct and indirect losses, such as the financial burden of reducing work time

utility is rewritten as:

$$\xi_i(y_i, z_i, \rho_i, \theta_i|s) = -\theta_i(s) + [u_i(y_i^H) - u_i(y_i^I(s))] \cdot p_i(z_i) + \epsilon_i \tag{1.10}$$

After differentiating (1.10) with respect s, the corresponding marginal effect of the paid sick leave policy is expressed mathematically as:

$$\frac{d\xi}{ds} = -\frac{\partial \theta_i}{\partial s} - \frac{\partial u_i}{\partial y_i^I} \cdot \frac{dy_i^I}{ds} \cdot p_i(z_i) \tag{1.11}$$

The marginal effect of the excess utility in (1.11) includes two parts: cost and income effects. First, for the cost effect, I assume that if a worker receives paid sick leave, s/he perceives the low average cost of the vaccine as a form of preventive care and is more willing to receive a flu vaccination. Thus, paid sick leave increases the vaccination probability, as follows:

$$-\frac{\partial \theta_i}{\partial s} \geq 0 \tag{1.12}$$

The income effect is realized through the PPFI, the marginal utility with respect to a worker's income, and the worker's marginal income with respect to paid sick leave. The first two factors are always positive because of the characteristics of probability and the assumption of a strictly increasing utility function. Thus, the direction of the income effect depends on the last factor, the worker's marginal income against paid sick leave. If workers perceive a high cost of using paid sick leave when they are sick, the marginal income is

and job insecurity when workers are absent. The derivative of y_i^I with respect to s is given by:

$$\frac{dy_i^I}{ds} = -\frac{dL_i}{ds} \tag{1.9}$$

Thus, the marginal income for an infected worker depends on the marginal potential loss when s/he is eligible for paid sick leave. If this worker perceives that the loss increases because of sick leave, s/he has a positive marginal loss. On the other hand, if this worker expects that the loss decreases because of sick leave, s/he has a negative marginal loss.

negative, and the income effect is positive. In this case, they would be willing to be vaccinated and their probability of being vaccinated increases. However, if workers perceive a high benefit of claiming paid sick leave when they are sick, the marginal income is positive and the income effect is negative. In this case, they would be unwilling to be vaccinated and the probability of vaccination decreases. Thus, the income effect can be specified by:

$$-\frac{\partial u_i}{\partial y_i^I} \cdot \frac{dy_i^I}{ds} \cdot p_i(z_i) = \begin{cases} > 0 & \text{if } dy_i^I/ds < 0 \\ = 0 & \text{if } dy_i^I/ds = 0 \\ < 0 & \text{if } dy_i^I/ds > 0 \end{cases} \tag{1.13}$$

By combining the cost effect and income effect, I conclude that the probability of vaccination always increases if the income effect is positive. I also concluded that the probability of vaccination is unclear if the income effect is negative. In this case, the vaccination decision depends on the scale difference between the cost and income effects.

I could explain the meaning of the cost and income effect that were decomposed by my vaccination model. The cost effect depends on whether or not workers have paid sick leave. If paid sick leave is given to workers, it could reduce the direct or indirect cost of vaccination. Thus, this effect is the actual effect of paid sick leave that increases the probability of their receiving a vaccination. However, the income effect depends on the change in the PPFIs of workers who have paid sick leave. If the PPFIs of workers who have paid sick leave increases, it could increase the cost of claiming paid sick leave. The income effect is the response to the increase in this cost. Thus, this effect is not the actual effect of paid sick leave but the expected losses from claiming paid sick leave that increases the probability of their vaccination. Thus, I argue that the income effect is a cognitive effect from perceiving potential cost that is not realized now when workers would claim paid sick leave.

1.3 Econometric Model

1.3.1 Model Specification

Section 1.2 concludes that workers' final decision of seasonal flu vaccination depends on the interaction of the cost and income effects. In this section, I construct the econometric model that realizes the vaccination decision and empirically evaluates the marginal effect of paid sick leave.

I consider two characteristics in my econometric model. First, it is difficult to identify the cost and income effects. This difficulty is attributable to the fact that the estimated marginal effect of paid sick leave is a combination of these two effects. The strategy to address this issue is to use the conditional perceived probability of flu infection without vaccination (cPPFI), which is an interaction term between the PPFI and a paid sick leave indicator. The income effect consists of marginal utility, marginal income, and the PPFI. Then, I determine that the marginal effect of the cPPFI would be the income effect. On the other hand, the marginal effect of the paid sick leave indicator would be the cost effect. If I estimate both marginal effects, the cPPFI term can divide the policy effect into the cost and income effects.

Another issue is endogeneity. I doubt that the PPFI and cPPFI could be endogenous variables as represented in (1.10). One possible unobserved confounder would be the vaccination coverage rate in neighbors. Many studies have argued that strategic behavior is used in vaccination, including free-riding (Hershey et al. 1994; Bauch, Galvani, and Earn 2003; Bauch and Earn 2004), altruism (Hershey et al. 1994; Shim et al. 2012), bandwagon (Hershey et al. 1994), and the peer-effect (Brunson 2013). These strategic decisions could influence both the PPFI and vaccination decision. Thus, the PPFI and the corresponding interaction term, cPPFI, are endogenous, and should be controlled by appropriate econometric model. I would like to employ a structural equation with the instrumental

variables using a Bayesian inference.

Let y_i and r_i be a vaccination and paid sick leave indicator, respectively. Further, let y_{1i} and w_i be the probability of infection without vaccination and the corresponding set of dummy variables for it, respectively. Similarly, let y_{2i} and $w_{i|r_i=1}$ be the cPPFI and the corresponding set of dummy variables for it, respectively. Then, my econometric regression model for the vaccination decision is written by the following structural equations:

$$
\begin{aligned}
y_i =& \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{w}'_i\boldsymbol{\beta}_2 + \boldsymbol{w}'_{i|r_i=1}\boldsymbol{\beta}_3 + \beta_4 r_i + \varepsilon_i \\
y_{1i} =& \boldsymbol{z}'_{1i}\boldsymbol{\gamma}_1 + \varepsilon_{1i} \\
y_{2i} =& \boldsymbol{z}'_{2i}\boldsymbol{\gamma}_2 + \varepsilon_{2i}
\end{aligned} \tag{1.14}
$$

where x_i denotes the covariates included in this model, and z_{1i} and z_{2i} denote the set of covariates including the instrumental variable and selected covariates from x_i for each endogenous equation. In this system, the first equation is the outcome equation that explained the binary vaccination decision with several covariates and perception variables. I refer to the last two equations as the endogenous equation that randomize the endogenous variable included in the outcome equation. I estimate this system of equations using latent variables with a Bayesian inference.

The estimation strategy for quantifying these two effects is as follow. I focus on two parameters, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ in (1.14). The former represents the parameter for the income effect estimation while the latter describes the parameter for the cost effect estimation. For the cost effect evaluation, I defined the cost effect as an increase in the probability that workers get vaccinated if they obtain paid sick leave given their PPFIs remain the same. For the income effect evaluation, I define the income effect as an increase in the probability that workers who have paid sick leave get vaccinated if their PPFIs become higher. To estimate these effect, I construct the marginal effects from the posterior predictive densities with the

corresponding coefficient estimates, which is explained in Section 1.3.2.

1.3.2 Estimation Method

1.3.2.1 Structural Equations and Reparamterization

I specify my suggested econometric model that addresses the system of equations in (1.14) as follows:

$$\begin{aligned} y_i^* =& \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ y_{1i}^* =& \boldsymbol{z}_{1i}'\boldsymbol{\gamma}_1 + \varepsilon_{1i} \\ y_{2i}^* =& \boldsymbol{z}_{2i}'\boldsymbol{\gamma}_2 + \varepsilon_{2i} \end{aligned} \tag{1.15}$$

where

$$\boldsymbol{x}_i'\boldsymbol{\beta} = \boldsymbol{x}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{w}_i'\boldsymbol{\beta}_2 + \boldsymbol{w}_{i|r_i=1}'\boldsymbol{\beta}_3 + \beta_4 r_i \tag{1.16}$$

The binary outcome variable y_i is relevant to the latent variable y_i^* as follows:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{1.17}$$

Similarly, two ordered treatment variable y_{1i} and y_{2i} are also related to the latent variable y_{1i}^* and y_{2i}^* as follows:

$$\begin{aligned} y_{1i} = k \text{ iff } \alpha_{1k} < y_{1i}^* \leq \alpha_{1k+1},\ k \in \{1, \cdots, K\} \\ y_{2i} = l \text{ iff } \alpha_{2l} < y_{2i}^* \leq \alpha_{2l+1},\ l \in \{1, \cdots, L\} \end{aligned} \tag{1.18}$$

where $\{\alpha_{1k}\}_{k=1}^K$ and $\{\alpha_{2l}\}_{l=1}^L$ are called cutpoints (or thresholds) and K and L are the numbers of such cutpoints. In addition, I set selected cutpoints to fixed numbers without

loss of generality, as follows:

$$\alpha_{m1} = -\infty,\ \alpha_{m2} = 0,\ \alpha_{mM} = \infty,\ m \in \{1, 2\}\ M \in \{K, L\} \tag{1.19}$$

Thus, the number of unknown cutpoints in these structural equations is $K-3$ for the second equation and $L-3$ for the third equation. Of the error term in (1.15), I assume joint normality as follows:

$$\begin{bmatrix} \varepsilon_i \\ \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} |\boldsymbol{x}_i, \boldsymbol{z}_{1i}, \boldsymbol{z}_{2i} \overset{ind}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} \right), \quad \text{where} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & 1 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & 1 \end{bmatrix} \tag{1.20}$$

The identification is one of the restrictions when I estimate the parameters of interest. If there exists a significant correlation among error terms, the parameters in this system could not be identifiable. For this purpose, I use the concept of the instrumental variable. I add some variables that act like potential instruments to the second and third equations in this system.

Given the independence assumption in the error terms, I can specify the likelihood function to address my model as follows:

$$\mathcal{L}(\Theta|\boldsymbol{y}, \boldsymbol{y}_1, \boldsymbol{y}_2) \propto \prod_{i=1}^{n} \prod_{j=0}^{1} \prod_{k=1}^{K} \prod_{l=1}^{L} \Pr(y_{1i} = k, y_{2i} = l, y_i = j|\Theta) \tag{1.21}$$

where $\Pr(\cdot)$ denotes a joint probability of several events occurred. I could calculate the joint

probability within this likelihood function as follows: if $j = 0$, then,

$$
\begin{aligned}
&\Pr(y_{1i} = k, y_{2i} = l, y_i = 0|\Theta) \\
=&\Pr(\alpha_{1k} < y_{1i}^* \leq \alpha_{1k+1}, \alpha_{2l} < y_{2i}^* \leq \alpha_{2l+1}, y_i^* \leq 0|\Theta, X_i) \\
=&\Pr(\alpha_{1k} - \boldsymbol{z}_{1i}'\boldsymbol{\gamma}_1 < \varepsilon_{1i} \leq \alpha_{1k+1} - \boldsymbol{z}_{1i}'\boldsymbol{\gamma}_1 \\
&\quad , \alpha_{2l} - \boldsymbol{z}_{2i}'\boldsymbol{\gamma}_2 < \varepsilon_{2i} \leq \alpha_{2l+1} - \boldsymbol{z}_{2i}'\boldsymbol{\gamma}_2, \varepsilon_i \leq -\boldsymbol{x}_i'\boldsymbol{\beta}|\Theta, \boldsymbol{X}_i) \\
=&\int_{\alpha_{1k+1}-z_{1i}'\gamma_1}^{\alpha_{1k}-z_{1i}'\gamma_1} \int_{\alpha_{2l+1}-z_{2i}'\gamma_2}^{\alpha_{2l}-z_{2i}'\gamma_2} \int_{-\infty}^{-x_i'\beta} f(\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_i) d\varepsilon_{1i} d\varepsilon_{2i} d\varepsilon_i
\end{aligned}
\tag{1.22}
$$

To employ classical estimation methods including maximum likelihood estimation, I should solve these integrals and use proper approximation and convergence solutions. In this study, I apply a Bayesian approach, specifically, Monte Carlo integration by employing truncated normal sampling with data augmentation rather than the classical method.

Before I start my estimation procedures, I apply the following reparameterization to reduce the lagged autocorrelations and a slow algorithm problem that induces high cost of inefficiency to reach the convergence (Nandram and Chen 1996). I set the third cutpoint for the second and third equations to one by reparameterizing it as follows:

$$
\sigma_1 = (1/\alpha_{13}^2) \quad \text{and} \quad \sigma_2 = (1/\alpha_{23}^2) \tag{1.23}
$$

By using these reparameterized parameters, I reparameterize all parameters relevant to the second and third endogenous equations. For the coefficient parameters, I reparameterize them as follows:

$$
\tilde{\boldsymbol{\gamma}}_j = \sqrt{\sigma_j}\boldsymbol{\gamma}_j \quad \text{and} \quad \tilde{y}_{ji}^* = \sqrt{\sigma_j} y_{ji}^*, \ j \in \{1, 2\} \tag{1.24}
$$

For the cutpoints, I also reparameterize them as follows:

$$\tilde{\alpha}_{1k} = \sqrt{\sigma_1}\alpha_{1k} \quad \text{and} \quad \tilde{\alpha}_{2l} = \sqrt{\sigma_2}\alpha_{2l},\ k \in \{3, \cdots, K\},\ l \in \{3, \cdots, L\} \tag{1.25}$$

Further, the correlation parameters appearing in the error terms can be parameterized by:

$$\tilde{\sigma}_{01} = \sqrt{\sigma_1}\sigma_{01},\ \tilde{\sigma}_{02} = \sqrt{\sigma_2}\sigma_{02},\ \tilde{\sigma}_{12} = \sqrt{\sigma_1\sigma_2}\sigma_{12} \tag{1.26}$$

Considering the reparameterization from (1.23) to (1.26), the model specified in (1.15) can be rewritten by:

$$\begin{aligned} y_i^* =& \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ \tilde{y}_{1i}^* =& \boldsymbol{z}_{1i}'\tilde{\boldsymbol{\gamma}}_1 + \tilde{\varepsilon}_{1i} \\ \tilde{y}_{2i}^* =& \boldsymbol{z}_{2i}'\tilde{\boldsymbol{\gamma}}_2 + \tilde{\varepsilon}_{2i} \end{aligned} \tag{1.27}$$

where

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{1.28}$$

and

$$\begin{aligned} y_{1i} = k \text{ iff } \tilde{\alpha}_{1k} < \tilde{y}_{1i}^* \leq \tilde{\alpha}_{1k+1},\ k \in \{1, \cdots, K\} \\ y_{2i} = l \text{ iff } \tilde{\alpha}_{2l} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2l+1},\ l \in \{1, \cdots, L\} \end{aligned} \tag{1.29}$$

with cutpoints

$$\tilde{\alpha}_{m1} = -\infty,\ \tilde{\alpha}_{m2} = 0,\ \tilde{\alpha}_{m3} = 1, \tilde{\alpha}_{mM} = \infty,\ m \in \{1, 2\}\ M \in \{K, L\} \tag{1.30}$$

The error terms with the reparameterized covariance parameters take the following form:

$$\begin{bmatrix} \varepsilon_i \\ \tilde{\varepsilon}_{1i} \\ \tilde{\varepsilon}_{2i} \end{bmatrix} |\boldsymbol{x}_i, \boldsymbol{z}_{1i}, \boldsymbol{z}_{2i} \overset{ind}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tilde{\boldsymbol{\Sigma}} \right), \quad \text{where} \quad \tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} 1 & \tilde{\sigma}_{01} & \tilde{\sigma}_{02} \\ \tilde{\sigma}_{01} & \sigma_1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{02} & \tilde{\sigma}_{12} & \sigma_2 \end{bmatrix} \tag{1.31}$$

where $\sigma_1$and σ_2 are the reparameterized parameters and to be estimated in the sampling process.

1.3.2.2 Posterior Sampling

My suggested model is similar to the regime-switching (or type 5 Tobit) model (Amemiya 1984) because of three equations, two endogenous variables, and correlation parameters. The regime-switching model with ordered outcome was developed by a Bayesian estimation method and produced highly reliable results (Li and Tobias 2008). I use the sampling method used in Li and Tobias (2008), in particular, when I draw conditional multivariate normal density for the latent variables.

First, I introduce several priors for the parameters to be estimated in my posterior samplers. The prior for all regression coefficient $\tilde{\boldsymbol{B}} = (\boldsymbol{\beta}, \tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2)'$ is given by:

$$p(\boldsymbol{\beta}) = (2\pi)^{-\frac{k}{2}} |\boldsymbol{B}_0|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left[(\tilde{\boldsymbol{B}} - \boldsymbol{B}_0)' \boldsymbol{V}_0^{-1} (\tilde{\boldsymbol{B}} - \boldsymbol{B}_0) \right] \right\} \tag{1.32}$$

that follows the multivariate normal distribution with mean $\boldsymbol{B}_0$ and variance $\boldsymbol{V}_0$. The prior for the covariance $\tilde{\boldsymbol{\Sigma}}$ is given by:

$$p(\tilde{\boldsymbol{\Sigma}}) = \left[2^{\nu_0} \pi^{\frac{1}{2}} \prod_{i=1}^{2} \Gamma\left(\frac{\nu_0 + 1 - i}{2} \right) \right]^{-1} |\boldsymbol{S}_0|^{\frac{\nu_0}{2}} |\tilde{\boldsymbol{\Sigma}}|^{-\frac{\nu_0+4}{2}} \exp\left\{ -\frac{1}{2} tr\left(S_0 \cdot \tilde{\boldsymbol{\Sigma}}^{-1} \right) \right\} I(\tilde{\boldsymbol{\Sigma}}_{11} = 1) \tag{1.33}$$

that follows the Inverse Wishart (IW) distribution with the degree of freedom ν_0 and scale

parameter $\boldsymbol{S}_0$. This density no longer follow the IW distribution because there exist one diagonal restriction in the covariance matrix ($\tilde{\boldsymbol{\Sigma}}_{11} = 1$). In the posterior sampler, the IW random sampler automatically adjusts this restriction. However, to obtain fast and accurate estimation, I use the special algorithm proposed by Nobile (2000). The priors for cutpoints $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2)$ are given by:

$$p(\tilde{\boldsymbol{\alpha}}_{kl}) = \prod_{k=3}^{K-1} \left[\frac{\exp\left\{ -\frac{1}{2\sigma_c^2} \tilde{\alpha}_{1k}^2 \right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)} \right] \cdot \prod_{l=3}^{L-1} \left[\frac{\exp\left\{ -\frac{1}{2\sigma_c^2} \tilde{\alpha}_{2l}^2 \right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)} \right] \tag{1.34}$$

that follows the truncated normal distribution bounded $[0, 1]$ with mean 0 and standard deviation 1.

The full observation over all individuals can be written by stacking all observations, as follows:

$$\tilde{\boldsymbol{Y}}^* = \begin{bmatrix} \boldsymbol{y}^* \\ \tilde{\boldsymbol{y}}_1^* \\ \tilde{\boldsymbol{y}}_2^* \end{bmatrix} \quad \text{and} \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} \tag{1.35}$$

Then, by combining the priors in (1.32) through (1.34) and likelihood function in (1.21), the

augmented joint posterior kernel takes the following form:

$$
\begin{aligned}
&p(\tilde{\boldsymbol{B}}, \tilde{\Sigma}, \tilde{\boldsymbol{Y}}^*, \tilde{\boldsymbol{\alpha}} | \boldsymbol{Y}, \boldsymbol{X}) \\
&\propto |\tilde{\boldsymbol{\Sigma}}|^{-\frac{\nu_0+4}{2}} \exp\left\{-\frac{1}{2} tr\left(S_0 \cdot \tilde{\boldsymbol{\Sigma}}^{-1}\right)\right\} I(\tilde{\boldsymbol{\Sigma}}_{11} = 1) \\
&\times \exp\left\{-\frac{1}{2}\left[(\tilde{\boldsymbol{B}} - \boldsymbol{B}_0)' \boldsymbol{V}_0^{-1} (\tilde{\boldsymbol{B}} - \boldsymbol{B}_0)\right]\right\} \\
&\times |\tilde{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})' \Sigma^{-1} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})\right\} \\
&\times \prod_{k=3}^{K-1} \left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2} \tilde{\alpha}_{1k}^2\right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)}\right] \cdot \prod_{l=3}^{L-1} \left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2} \tilde{\alpha}_{2l}^2\right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)}\right] \\
&\times \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{l=1}^{L} \begin{pmatrix} I(y_{1i} = k) I(\tilde{\alpha}_{1k} < \tilde{y}_{1i}^* \leq \tilde{\alpha}_{1k+1}) \\ \cdot I(y_{2i} = l) I(\tilde{\alpha}_{2l} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2l+1}) \\ \cdot I(y_i = 0)(\tilde{y}_i^* \leq 0) \\ + I(y_{1i} = k) I(\tilde{\alpha}_{1k} < \tilde{y}_{1i}^* \leq \tilde{\alpha}_{1k+1}) \\ \cdot I(y_{2i} = l) I(\tilde{\alpha}_{2l} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2l+1}) \\ \cdot I(y_i = 1)(\tilde{y}_i^* > 0) \end{pmatrix}
\end{aligned}
\tag{1.36}
$$

Based on the joint posteriors, I suggest four different sampling processes to estimate all parameters and augmented latent variables.

Step 1: The conditional posterior kernel for $\tilde{\boldsymbol{B}}$ takes a form as follows:

$$
\begin{aligned}
&p(\tilde{\boldsymbol{B}} | \tilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{Y}}^*, \boldsymbol{X}) \\
&\propto \exp\left\{-\frac{1}{2}\left[(\tilde{\boldsymbol{B}} - \boldsymbol{B}_0)' \boldsymbol{V}_0^{-1} (\tilde{\boldsymbol{B}} - \boldsymbol{B}_0) + \sum_{i=1}^{n} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})' \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})\right]\right\}
\end{aligned}
\tag{1.37}
$$

Thus, based on this kernel density, I can derive the conditional posterior moments of multivariate normal density, given by:

$$\tilde{\boldsymbol{B}}|\tilde{\boldsymbol{\Sigma}}, \tilde{\boldsymbol{Y}}^*, \boldsymbol{X} \sim N(\boldsymbol{\mu}_1, \boldsymbol{V}_1) \tag{1.38}$$

with the parameters

$$\boldsymbol{V}_1 = \Big(\boldsymbol{V}_0^{-1} + \sum_{i=1}^{n} \boldsymbol{X}_i' \tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X}_i\Big) \quad \text{and} \quad \boldsymbol{\mu}_1 = \boldsymbol{V}_1 \Big(\boldsymbol{V}_0^{-1} \boldsymbol{\mu}_0 + \sum_{i=1}^{n} \boldsymbol{X}_i' \tilde{\Sigma}^{-1} \boldsymbol{X}_i\Big) \tag{1.39}$$

Step 2: The conditional posterior for $\tilde{\boldsymbol{\Sigma}}$ also takes a form as follows:

$$\begin{aligned} &p(\tilde{\boldsymbol{\Sigma}}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{Y}}^*, \boldsymbol{X}) \\ &\propto |\tilde{\boldsymbol{\Sigma}}|^{-\frac{\nu_0+4+n}{2}} \exp\Big\{-\frac{1}{2}\Big[tr(S_0 \cdot \tilde{\boldsymbol{\Sigma}}^{-1}) + \sum_{i=1}^{n} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})' \Sigma^{-1} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})\Big]\Big\} I(\tilde{\boldsymbol{\Sigma}}_{11} = 1) \end{aligned} \tag{1.40}$$

Thus, the corresponding conditional posterior moments takes a form of the IW density as:

$$\tilde{\boldsymbol{\Sigma}}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{Y}}^*, \boldsymbol{X} \sim IW(\nu_1, \boldsymbol{S}_1) I(\tilde{\boldsymbol{\Sigma}}_{11} = 1) \tag{1.41}$$

with

$$\nu_1 = \nu_0 + n \quad \text{and} \quad S_1 = S_0 + \sum_{i=1}^{n} (\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})(\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i \tilde{\boldsymbol{B}})' \tag{1.42}$$

The following steps in the posterior simulator include joint sampling process for the latent

variables. The conditional posteriors for the augmented latent variables are given by:

$$
\begin{aligned}
&p(\tilde{\boldsymbol{Y}}^*|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha}^*, \boldsymbol{X}, \boldsymbol{Y}) \\
&\propto |\tilde{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i\tilde{\boldsymbol{B}})'\Sigma^{-1}(\tilde{\boldsymbol{Y}}_i^* - \boldsymbol{X}_i\tilde{\boldsymbol{B}})\right\} \\
&\times \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{l=1}^{L}\left(\begin{array}{l}
I(y_{1i}=k)I(\tilde{\alpha}_{1k} < \tilde{y}_{1i}^* \leq \tilde{\alpha}_{1k+1}) \\
\cdot I(y_{2i}=l)I(\tilde{\alpha}_{2l} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2l+1}) \\
\cdot I(y_i=0)(\tilde{y}_i^* \leq 0) \\
+I(y_{1i}=k)I(\tilde{\alpha}_{1k} < \tilde{y}_{1i}^* \leq \tilde{\alpha}_{1k+1}) \\
\cdot I(y_{2i}=l)I(\tilde{\alpha}_{2l} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2l+1}) \\
\cdot I(y_i=1)(\tilde{y}_i^* > 0)
\end{array}\right)
\end{aligned}
\tag{1.43}
$$

To derive the posterior moments for the latent variables, I use the conditional trivariate normal density (Li and Tobias 2008; Greene 2008). I obtain a convenient algorithm for Monte Carlo integration from the truncated normal distribution to obtain the posterior samples for latent variables.

Step 3-1: The conditional posterior moment for y_i^* takes a form as follows:

$$
y_i^*|y_{1i}^*, y_{2i}^*, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha}^*, \boldsymbol{X}, \boldsymbol{Y} \sim \begin{cases} TN_{(-\infty,0)}(\mu_{0|12}, \sigma_{0|12}) & \text{if } y_i = 0 \\ TN_{(0,\infty)}(\mu_{0|12}, \sigma_{0|12}) & \text{if } y_i = 1 \end{cases}
\tag{1.44}
$$

where

$$
\mu_{0|12} = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{w}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{w}_{2i}'\boldsymbol{\beta}_2 + \begin{bmatrix}\tilde{\sigma}_{01} & \tilde{\sigma}_{02}\end{bmatrix}\begin{bmatrix}\sigma_{11} & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \sigma_2\end{bmatrix}^{-1}\begin{bmatrix}\tilde{y}_{1i}^* - \boldsymbol{z}_{1i}'\tilde{\boldsymbol{\gamma}}_1 \\ \tilde{y}_{2i}^* - \boldsymbol{z}_{2i}'\tilde{\boldsymbol{\gamma}}_2\end{bmatrix}
\tag{1.45}
$$

and

$$\sigma_{0|12} = 1 - \begin{bmatrix} \tilde{\sigma}_{01} & \tilde{\sigma}_{02} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{01} \\ \tilde{\sigma}_{02} \end{bmatrix} \tag{1.46}$$

Step 3-2: Similarly, the conditional posterior moment for y_{1i}^* is given by:

$$y_{1i}^* | y_{2i}^*, y_i^*, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha}^*, \boldsymbol{X}, \boldsymbol{Y} \sim TN_{(\alpha_{1k}, \alpha_{1k+1})}(\mu_{1|02}, \sigma_{1|02}) \tag{1.47}$$

with

$$\mu_{1|02} = \boldsymbol{z}_{1i}' \tilde{\boldsymbol{\gamma}}_1 + \begin{bmatrix} \tilde{\sigma}_{01} & \tilde{\sigma}_{12} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\sigma}_{02} \\ \tilde{\sigma}_{02} & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} y_i^* - \boldsymbol{x}_i' \boldsymbol{\beta} \\ \tilde{y}_{2i}^* - \boldsymbol{z}_{2i}' \tilde{\boldsymbol{\gamma}}_2 \end{bmatrix} \tag{1.48}$$

and

$$\sigma_{1|02} = \sigma_{11} - \begin{bmatrix} \tilde{\sigma}_{01} & \tilde{\sigma}_{12} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\sigma}_{02} \\ \tilde{\sigma}_{02} & \sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{01} \\ \tilde{\sigma}_{12} \end{bmatrix} \tag{1.49}$$

Step 3-3: Further, the conditional posterior moment for y_{2i}^* is also given by:

$$y_{2i}^* | y_{1i}^*, y_i^*, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\alpha}^*, \boldsymbol{X}, \boldsymbol{Y} \sim TN_{(\alpha_{2k}, \alpha_{2k+1})}(\mu_{2|01}, \sigma_{2|01}) \tag{1.50}$$

where

$$\mu_{2|01} = \boldsymbol{z}_{2i}' \tilde{\boldsymbol{\gamma}}_2 + \begin{bmatrix} \tilde{\sigma}_{02} & \tilde{\sigma}_{12} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\sigma}_{01} \\ \tilde{\sigma}_{01} & \sigma_1 \end{bmatrix}^{-1} \begin{bmatrix} y_i^* - \boldsymbol{x}_i' \boldsymbol{\beta} \\ \tilde{y}_{1i}^* - \boldsymbol{z}_{1i}' \tilde{\boldsymbol{\gamma}}_1 \end{bmatrix} \tag{1.51}$$

and

$$\sigma_{2|01} = \sigma_{22} - \begin{bmatrix} \tilde{\sigma}_{02} & \tilde{\sigma}_{12} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\sigma}_{01} \\ \tilde{\sigma}_{01} & \sigma_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{02} \\ \tilde{\sigma}_{12} \end{bmatrix} \tag{1.52}$$

The last step is to derive the conditional posteriors for cutpoints $\tilde{\boldsymbol{\alpha}}$ that takes forms as

follows:

$$
\begin{aligned}
&p(\tilde{\boldsymbol{\alpha}}|\tilde{\boldsymbol{B}},\tilde{\boldsymbol{\Sigma}},\tilde{\boldsymbol{Y}}^*) \\
&\propto \prod_{k=3}^{K-1}\left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{1k}^2\right\}}{\Phi(0,\sigma_c,1)-\Phi(0,\sigma_c,0)}\right]\cdot\prod_{l=3}^{L-1}\left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{2l}^2\right\}}{\Phi(0,\sigma_c,1)-\Phi(0,\sigma_c,0)}\right] \\
&\times\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{l=1}^{L}\left(\begin{array}{l}
I(y_{1i}=k)I(\tilde{\alpha}_{1k}<\tilde{y}_{1i}^*\leq\tilde{\alpha}_{1k+1}) \\
\cdot I(y_{2i}=l)I(\tilde{\alpha}_{2l}<\tilde{y}_{2i}^*\leq\tilde{\alpha}_{2l+1}) \\
\cdot I(y_i=0)(\tilde{y}_i^*\leq 0) \\
+I(y_{1i}=k)I(\tilde{\alpha}_{1k}<\tilde{y}_{1i}^*\leq\tilde{\alpha}_{1k+1}) \\
\cdot I(y_{2i}=l)I(\tilde{\alpha}_{2l}<\tilde{y}_{2i}^*\leq\tilde{\alpha}_{2l+1}) \\
\cdot I(y_i=1)(\tilde{y}_i^*>0)
\end{array}\right)
\end{aligned}
\tag{1.53}
$$

The remaining steps include the sampling processes for the cutpoints in three equations. I use truncated normal densities and apply the Metropolis-Hastings (NH) method for these steps.

Step 4-1: The conditional posterior for α_{1k} is given by:

$$
\begin{aligned}
&p(\alpha_{1k}|\tilde{\boldsymbol{B}},\tilde{\boldsymbol{\Sigma}};\alpha_{1k}<y_{1i}^*\leq\alpha_{1k+1}) \\
&\propto\left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{1k}^2\right\}}{\Phi(0,\sigma_c,1)-\Phi(0,\sigma_c,0)}\right]\prod_{i=1}^{n}I(\tilde{\alpha}_{1k}<\tilde{y}_{1i}^*\leq\tilde{\alpha}_{1k+1}) \\
&\propto\left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{1k}^2\right\}}{\Phi(0,\sigma_c,1)-\Phi(0,\sigma_c,0)}\right]\prod_{i=1}^{n}\left[\Phi(\frac{\tilde{\alpha}_{1k+1}-\mu_{1|02}}{\sqrt{\sigma_{1|02}}})-\Phi(\frac{\tilde{\alpha}_{1k+1}-\mu_{1|02}}{\sqrt{\sigma_{1|02}}})\right]
\end{aligned}
\tag{1.54}
$$

In this case, there is no known density kernel for this posterior. Thus, I use the MH method to draw a candidate and determine the acceptance for the cutpoints as follows:

$$p_a(\alpha_{1k}^{(o)} \to \alpha_{1k}^{(c)}) = \min\left\{\frac{p(\alpha_{1k}^{(c)}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{Z}_1)q(\alpha_{1k}^{(o)}|\alpha_{1k}^{(c)})}{p(\alpha_{1k}^{(o)}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{Z}_1)q(\alpha_{1k}^{(c)}|\alpha_{1k}^{(o)})}, 1\right\} \tag{1.55}$$

where $q(\cdot)$ denotes a candidate generation function.

Step 4-2: I apply the identical method for the posteriors for α_{2k}. The conditional posterior for α_{2k} is given by:

$$\begin{aligned}
&p(\alpha_{2k}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}; \alpha_{2k} < y_{2i}^* \leq \alpha_{2k+1}) \\
&\propto \left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{2k}^2\right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)}\right] \prod_{i=1}^{n} I(\tilde{\alpha}_{2k} < \tilde{y}_{2i}^* \leq \tilde{\alpha}_{2k+1}) \\
&\propto \left[\frac{\exp\left\{-\frac{1}{2\sigma_c^2}\tilde{\alpha}_{2k}^2\right\}}{\Phi(0, \sigma_c, 1) - \Phi(0, \sigma_c, 0)}\right] \prod_{i=1}^{n} \left[\Phi(\frac{\tilde{\alpha}_{2k+1} - \mu_{2|01}}{\sqrt{\sigma_{2|01}}}) - \Phi(\frac{\tilde{\alpha}_{1k+1} - \mu_{2|01}}{\sqrt{\sigma_{2|01}}})\right]
\end{aligned} \tag{1.56}$$

and the acceptance probability for a candidate is written by:

$$p_a(\alpha_{2k}^{(o)} \to \alpha_{2k}^{(c)}) = \min\left\{\frac{p(\alpha_{2k}^{(c)}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{Z}_1)q(\alpha_{2k}^{(o)}|\alpha_{2k}^{(c)})}{p(\alpha_{2k}^{(o)}|\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{Z}_1)q(\alpha_{2k}^{(c)}|\alpha_{2k}^{(o)})}, 1\right\} \tag{1.57}$$

My posterior simulator through steps 1 to 4 generates proper sampling draws to estimate the parameters in which I am interested. To recover the original parameters from the reparameterized parameters, I invert the identifies described in (1.23) through (1.26).

1.3.2.3 Marginal Effect Estimation

I describe the estimation strategy of quantifying the cost and income effects in Section 1.3.2. The posterior predictive densities of the marginal effects that are corresponding the cost and income effects are follows. Let J be the number of ordered variable for the probability of

infection without vaccination. Then, the marginal effect when paid sick leave is given to a worker was estimated by: for each $\bar{\boldsymbol{w}}_j \in \{1, \cdots, J\}$

$$
\begin{aligned}
&p\Big(p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, r_p = 1\big) - p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, r_p = 0\big)\Big) \\
=&\int_{\beta_2,\beta_3} \Big[\Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1 + \boldsymbol{w}_{p|r=1}'\boldsymbol{\beta}_2 + \beta_3\big) - \Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1\big)\Big] d\boldsymbol{\beta}_2 d\beta_3
\end{aligned} \tag{1.58}
$$

where $\boldsymbol{x}_p, \boldsymbol{w}_p, \boldsymbol{w}_{p|r=1}$ and r_p denote target covariates I am interested in for $\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{w}_{|r=1}$ and r. However, this marginal effect is not informative in identifying the cost and income effects. Thus, I suggest a strategy to decompose this effect. First, the cost effect is evaluated by fixing the interaction term to zero vector ($\boldsymbol{w}_{p|r=1} = 0$), as follows: for each $\bar{\boldsymbol{w}}_j \in \{1, \cdots, J\}$

$$
\begin{aligned}
&p\Big(p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, \boldsymbol{w}_{p|r_i=1} = \boldsymbol{0}, r_p = 1\big) - p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, r_p = 0\big)\Big) \\
=&\int_{\beta_2,\beta_3} \Big[\Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1 + \beta_3\big) - \Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1\big)\Big] d\boldsymbol{\beta}_2 d\beta_3
\end{aligned} \tag{1.59}
$$

In this case, the result describe only the effect of the change in the cost of vaccine and remove the effect related to the change in the perceived probability of infection without vaccination. One other hand, the income effect is calculated by fixing the interaction term and paid sick leave indicator to a target vector ($\boldsymbol{w}_{p|r=1} = \bar{\boldsymbol{w}}_{p,j}$) and zero ($r_p = 0$), respectively, as follows: for each $\bar{\boldsymbol{w}}_j \in \{1, \cdots, J\}$

$$
\begin{aligned}
&p\Big(p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, \boldsymbol{w}_{p|r=1} = \bar{\boldsymbol{w}}_{p,j}, r_p = 1\big) - p\big(y_i = 1|\boldsymbol{x}_p, \boldsymbol{w}_p = \bar{\boldsymbol{w}}_j, \boldsymbol{w}_{p|r_i=1} = \boldsymbol{0}, r_p = 1\big)\Big) \\
=&\int_{\beta_2,\beta_3} \Big[\Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1 + \boldsymbol{w}_{p|r=1}'\boldsymbol{\beta}_2 + \beta_3\big) - \Phi\big(\boldsymbol{x}_p'\boldsymbol{\beta} + \boldsymbol{w}_p'\boldsymbol{\beta}_1 + \beta_3\big)\Big] d\boldsymbol{\beta}_2 d\beta_3
\end{aligned} \tag{1.60}
$$

In this case, the result describe only the effect of a change in expected income and remove the effect related to the change in the cost of vaccination.

I also schematize the estimation strategy in Figure 1.1. In this figure, I set up three

different groups based on whether they have paid sick leave and their PPFIs. The distance between points on the line represents the measure of marginal effect.

[Figure 1.1 about here]

1.4 Data and Measures

I examined the National 2009 H1N1 Flu Survey (NHFS) to estimate the marginal individual benefit function (DHHS and NCHS 2012). This survey was conducted jointly by the National Center for Immunization and Respiratory Diseases (NCIRD), the National Center for Health Statistics (NCHS), and the CDC, and was administered from October 2009 through June 2010. In addition to information about H1N1 vaccination, it contains seasonal flu vaccination information, such as a vaccination indicator, respondents' socio-demographic characteristics, and perceptions about vaccination effectiveness, side-effect, and infection probability. This study focuses only on vaccination for the seasonal flu and thus selected only seasonal flu-related variables. The descriptive statistics are shown in Table 1.1.

[Table 1.1 about here]

The purpose of this study is to investigate the causal relationship between the decision to obtain a vaccination for the seasonal flu and paid sick leave. I use and focus on the subsample of employed and young ($18-64$ years) respondents only because I am interested in a vaccination policy that focuses on a younger population whose vaccination coverage is much less than the older population. Moreover, I consider the observations answered in 2010 but not those in 2009 for two reasons. First, I am concerned that most individuals decide whether to be vaccinated before the end of the year. Thus, I might include a respondent who had not received the flu vaccine before the interview date but was vaccinated thereafter. Another reason is that an important factor, an indicator of health insurance, was added in

only 2010, but not in 2009. Also, I drop observations if they included missing variables. Ultimately, I examine $11,702$ of $70,944$ observations in this survey.

This study also investigates the heterogeneous cost and income effects according to workers' household income levels. Thus, I separate the selected samples into three subsamples: low-income ($\leq \$25,000$), middle income ($\$25,001 - \$75,000$), and high-income ($> \$75,000$). The 2009 poverty guidelines for the 48 Contiguous States and the District of Columbia showed that the threshold for the poverty line was $\$22,050$ per family of four in 2009 (ASPE 2009). Thus, the cutpoint for low-income workers is $\$25,000$ that is the closest categories in this survey.

The vaccination rate for low-income workers is much lower than was that of middle- and high-income workers. Only 29.15% of low-income workers were vaccinated, while 42.16% and 52.03% of middle- and high-income workers, respectively, received vaccinations. The proportion of low-income workers who had health insurance is also lower than those of other income groups. Approximately 56% of low-income workers were insured, while approximately 97% of high-income workers were insured. Another characteristic of income difference is paid sick leave. Only 32% of low-income workers could claim paid sick leave, while approximately 74% of high-income workers had paid sick leave.

1.4.1 Dependent variables

The dependent variable for the outcome equation is the vaccination for seasonal flu. I dichotomize this variable as "1" if a respondent is vaccinated and "0" if not. The dependent variable for the first endogenous equation is the PPFI. This variable is the respondents' answer to the survey question, "If you [had not gotten/do not get] a seasonal flu vaccination this fall or winter, what [would have been/are] your chances of getting sick with the seasonal flu?" The possible responses to this question are "very high," "somewhat high," "somewhat

low," and "very low." The dependent variable for the second endogenous variable is the cPPFI that described the conditional PPFI given that a worker has paid sick leave. This variable consists of "No paid sick leave," "loss with very low probability," "loss with somewhat low probability" "loss with high probability," and "loss with very high probability."

1.4.2 Independent variables

A key independent variable in this study is an indicator of having paid sick leave. This variable is generated from the question, "Workers sometimes receive benefits in addition to their wages. Whether you receive them or not, please tell me whether you are ELIGIBLE to receive sick leave with full pay." If they could use sick leave with full pay, the variable was coded as "1," as "0" if not.

The covariates in the first equation include a constant term, socio-demographic factors, and individual perceptions about the flu vaccine and infection risk. The socio-demographic factors include gender, age group ($18 - 44$ years or $45 - 64$), marital status (married or unmarried), number of children in the family, number of adults in the family, educational degree (high school or below, or college or higher), chronic disease status (chronic disease or not), and health insurance (insured or not). One of the perception variables is the perceived side-effects of the seasonal flu vaccine. This variable is the respondents' answer to the survey question, "How worried [were/are] you about getting sick from the seasonal flu vaccine?" The possible answers are "very high," "somewhat high," "somewhat low," and "very low." Another perception variable was the perceived effectiveness of the seasonal flu vaccine. This variable is the respondents' answer to the survey question, "How effective do you think the seasonal flu vaccination [was/is] in preventing the seasonal flu?" The possible answers were "very high," "somewhat high," "somewhat low," and "very low." These independent variables except two perception variables about the vaccine are also commonly used in two endogenous

equations.

The instrumental variable I choose is the PPFI for the H1N1 flu. This variable is the respondents' answer to the survey question, "If you [had not gotten/do not get] a seasonal flu vaccination this fall or winter, what [would have been/are] your chances of getting sick with the H1N1?" The possible responses are "very high," "somewhat high," "somewhat low," and "very low." A suitable instrument in this study should affect the vaccination decision only through the endogenous variable, not directly.

1.5 Estimation Results

All posteriors are obtained from $10,000$ replications; $5,000$ burn-in replications are discarded and $5,000$ are retained. Because this study employs a Bayesian estimation method, I introduce alternative measures for coefficient estimates and marginal effects. First, a classical significance measure, such as the standard error or p-value, cannot be used for a Bayesian estimation. Thus, I use an alternative measure that represents the probability that the mass of the posteriors drawn from the samplers is placed over the positive value (Koop, Poirier, and Tobias 2007). If this measure approaches zero, I suggest that the corresponding posterior mean is negatively significant. On the other hand, if the measure is close to one, I suggest that the posterior mean is positively significant. Based on this concept, the measure is "highly significant" if the measure is greater than 0.975 or less than 0.025. Also, The measure is "marginally significant" if it is greater than 0.950 or less than 0.050. Second, to calculate the predictive probability and the corresponding marginal effect, I use Bayesian measures of Marginal Effects at the Means (MEMs) and Average Marginal Effects (AMEs) with the posterior draws (Williams et al. 2012). Table 1.2 provides the posterior means and standard deviations for the outcome equation.

[Table 1.2 about here]

First, I observe that the covariance/correlation terms between the outcome and endogenous equations are highly significant for high-income, but not for low-income workers. The posterior means of the covariance terms for high- and middle-income workers are highly significant with -0.131 and -0.280, respectively. I conclude that the regression model I used to address the endogeneity and estimate the parameters of interest would be appropriate.

Second, I observe that paid sick leave is a critical factor that increases the probability of flu vaccination among high-income workers, although it is not significant among low-income workers. The coefficient estimate of paid sick leave for high-income workers is 0.643, statistically significant based on the significance measure. In contrast, the estimate of paid sick leave for low-income workers is 0.013 and statistically insignificant. This result indicates that the cost effect of paid sick leave matters only to high-income workers.

I also observe that the effect of cPPFI differs for different income groups. As Table 1.2 showed, only the coefficients of "very high" cPPFI for low- and middle- income workers are significant. This fact shows that although low-income workers do not consider paid sick leave as a cost saving, they consider their sick leave contracts significantly via the PPFI. In other words, although paid sick leave is a significant factor for high-income workers, they do not consider the future loss from paid sick leave when they determine the vaccination. Thus, I conclude that paid sick leave increases the probability of vaccination among low-income workers because they would perceive a high cost of claiming paid sick leave if they became infected by flu.

Table 1.3 and 1.4 shows the MEMs and AMEs of paid sick leave on the vaccination decision according to worker's income levels.

[Table 1.3 and 1.4 about here]

First, giving paid sick leave to workers increases their probability of obtaining a

vaccination. The total effect in Table 1.3 and 1.4 shows positive values for all income levels and perceptions. Further, most values in the high and very high PPFI are significant, with low standard deviations of posterior draws. Thus, I conclude that the policy of giving paid sick leave encourages workers to obtain vaccinations and prevents them from unexpected infection through presenteeism.

However, the aspects of cost and income effects differ significantly depending on workers' income levels. The relevant result shows that low-income workers have a significant income effect of paid sick leave on increasing the probability of vaccination. The marginal income for low-income workers is positively significant, ranging from 0.145 to 0.222 in MEMs, and 0.123 to 0.196 in AMEs when their PPFIs are high or very high. On the other hand, the cost effects were not significant for both MEMs and AMEs at all levels of the PPFIs. This result indicates that low-income workers decide to get vaccinated because they perceive a high expected loss of claiming paid sick leave and ignore any current benefit of the low cost of vaccination.

I also observe a different aspect of cost and income effect among high-income workers. The estimation result shows that the income effect is much higher than the cost effect for high-income workers. It also shows that the total effect is greater than zero despite negative income effect because the cost effect dominates the negative income effect. High-income workers' marginal income is close to zero and even negative for the low level of the PPFI, ranging from -0.064 to 0.035 based on both MEMs and AMEs. However, their cost effect ranges from 0.085 to 0.237 depending on the level of PPFI. This result indicates that high-income workers make their vaccination decision by considering the current benefit of vaccination, but disregarding the expected loss from their future use of paid sick leave.

1.6 Discussion

I showed in the decision model that the marginal effect of paid sick leave depends on the perception of current benefit (cost effect) and cognitive loss (income effect) of having paid sick leave. To identify the cost and income effects from the estimation results, I evaluated and decomposed the marginal effect of paid sick leave. Based on these results, I discuss several issues regarding the vaccination inequality by workers' income level.

First, the estimation results of marginal effect explain that high-income workers have higher vaccination coverage than do low-income workers. About 74% of high-income workers have a paid sick leave contract, and it increased their vaccination coverage because the total effect for them is positive. Conversely, Only about 32% of low-income workers have a paid sick leave contract, and the positive total effect was less effective to increase their vaccination coverage. Thus, I conclude that the inequality in possessing paid sick leave by workers' income level is one of the critical reasons for the inequality in vaccination decision and the corresponding vaccination coverage disparity.

Second, I argue whether or not paid sick leave is an actual factor to increase the vaccination coverage for seasonal flu, in particular, for low-income workers. This study identified and decomposed the source to affect the probability of seasonal flu vaccination. These identification and decomposition results show that high-income workers could obtain benefits of a reduction in direct and indirect costs of vaccination for seasonal flu from a paid sick leave contract. Thus, paid sick leave could increase high-income workers' vaccination through the cost effect, and such effect is actual. However, these results also show that low-income workers perceived the high cognitive cost of claiming paid sick leave when they would be sick or infected. They would be an underprivileged group to obtain an actual effect of paid sick leave on the vaccination decision. Thus, the total effect of paid sick leave for low-income workers accounts for the income effect, and the effect of paid sick

leave would not be actual but cognitive illusion of workers who have paid sick leave.

Some studies have investigated on the positive effect of paid sick leave by asserting that paid sick leave increases timely access to healthcare and less absenteeism due to illness in the workplace (Cook 2011; DeRigne, Stoddard-Dare, and Quinn 2016). However, I find a different conclusion from these studies arguing that there is an income inequality that affects workers' vaccination in vaccination or preventive care decision. I contend that this inequality would be attributable to job security that represents whether workers could use paid sick leave without restrictions or not and yield a positive income effect for low-income workers. Several studies support this argument that low-income workers' sick leave increases job insecurity, such as unemployment or fewer chances of promotion in the workplace (Hesselius 2007; Balchin and Wooden 1995; Brown and Sessions 1996; DeLeire and Manning 2004). If a worker has flu-like symptoms and is absent from the workplace, his/her employer may hire other available workers rather than waiting for this worker to return to the workplace. In this case, if employers are able to supply labor easily because the labor market is flexible, they may hire substitutes and the current worker might be laid off temporarily or permanently. However, because the job security of high-income workers would be stable than that of low-income workers, they can use sick leave without any loss of salary or risk of losing their positions when they are sick. Thus, I finally conclude that low-income workers' job security was less stable than that of high-income workers, and therefore, the income effect for low-income workers was greater than that for high-income workers.

It remains unclear which factors produce the cost effect. This survey does not include cost-related variables that are affected by paid sick leave. A possible explanation is a preventive care included in the paid sick leave contract. Low-income workers would be restricted to access vaccination locations during working hours, and few can utilize the benefit of claiming paid sick leave. Conversely, high-income workers could obtain more chances to utilize preventive care than low-income workers could, which encourages

high-income workers to be vaccinated than low-income workers.

1.7 Data Limitations

This study has several limitations. First, the indicator of paid sick leave does not describe any details of the contract. The survey includes a simple question, whether paid sick leave is offered or not. Thus, I cannot identify, if any, what differences existed between low- and high-income workers. Another limitation is income level. The income level used in this survey is household income, not actual salaries or wages. Thus, the classification of income level does not reflect the worker's wage precisely. Moreover, this survey does not include any variables about the cost of vaccination except whether respondents are insured or not. Thus, it is not possible to estimate or analyze the effect of the direct vaccination cost, such as the price of a flu shot.

1.8 Conclusions

Through this analysis, I conclude that paid sick leave increases the probability of workers' vaccination because of the positive total effect. However, the estimation results indicate that low-income workers are willing to be vaccinated because their cognitive loss of claiming paid sick leave is significantly high. High-income workers are willing to be vaccinated because they consider a reduction of the direct and indirect costs of vaccination by having paid sick leave, but not expected the loss when they are sick. I also explained the potential cause of this inequality by workers' job security.

This study provides an implication. Low-income workers pay attention to the income effect more than the cost effect. This tendency shows that low-income workers have less job security, so they may be worried about their employment status. Such job insecurity

encourages these workers to be vaccinated before flu season rather than waiting for the expected benefit of paid sick leave. Thus, eliminating restricted access to clinics and the cost of the burden is an efficient policy to support low-income workers by stimulating the cost effect. One solution to this limitation is a financial incentive for visiting community pharmacies or supermarket medical facilities. These operate during off-clinic hours and offer visits with nurse practitioners where patients can be vaccinated at more convenient times (Goad et al. 2013; Wilson, Wang, and Stimpson 2014). Furthermore, providing vaccination opportunities in non-traditional locations, including the workplace, would be an effective policy to reduce low-income workers' limited access to vaccination locations (Kim and Mountain 2017).

Finally, these suggestions are also effective for other forms of preventive healthcare use, in particular, cancer screening for female worker's health because vaccination is a kind of preventive care. Breast cancer and cervical cancer can be prevented by preventive medical care including regular Mamograms and Pap testing, respectively. However, low-income female workers have a potential barrier to access to the preventive medical services. Thus, the eliminating restricted access would be a possible solution to utilize the cost effect and improve their health.

Acknowledgments

A part of this paper was published with Mountain in *Social Science and Medicine* (Kim and Mountain 2018).

References

Amemiya, Takeshi (1984). "Tobit models: A survey". *Journal of econometrics* 24.1-2, pp. 3–61.

ASPE (2009). "The 2009 HHS poverty guidelines". *Federal Register* 74.14, pp. 4199–4201.

Balchin, Jeffrey and Wooden, Mark (1995). "Absence penalties and work attendance". *Australian Economic Review* 28.4, pp. 43–58.

Bauch, Chris T and Earn, David JD (2004). "Vaccination and the theory of games". *Proceedings of the National Academy of Sciences of the United States of America* 101.36, pp. 13391–13394.

Bauch, Chris T, Galvani, Alison P, and Earn, David JD (2003). "Group interest versus self-interest in smallpox vaccination policy". *Proceedings of the National Academy of Sciences* 100.18, pp. 10564–10567.

Brito, Dagobert L, Sheshinski, Eytan, and Intriligator, Michael D (1991). "Externalities and compulsary vaccinations". *Journal of Public Economics* 45.1, pp. 69–90.

Brown, Sarah and Sessions, John G (1996). "The economics of absence: theory and evidence". *Journal of Economic Surveys* 10.1, pp. 23–53.

Brunson, Emily K (2013). "The impact of social networks on parentsv́accination decisions". *Pediatrics*.

CDC (2015). *Flu Vaccination Coverage, United States, 2015-16 Influenza Season*. https://www.cdc.gov/flu/fluvaxview/coverage-1516estimates.htm (last visited: 6/14/2017).

Colla, Carrie H, Dow, William H, Dube, Arindrajit, and Lovell, Vicky (2014). "Early effects of the San Francisco paid sick leave policy". *American journal of public health* 104.12, pp. 2453–2460.

Cook, Won Kim (2011). "Paid sick days and health care use: An analysis of the 2007 national health interview survey data". *American journal of industrial medicine* 54.10, pp. 771–779.

DeLeire, Thomas and Manning, Willard (2004). "Labor market costs of illness: prevalence matters". *Health economics* 13.3, pp. 239–250.

DeRigne, LeaAnne, Stoddard-Dare, Patricia, Collins, Cyleste, and Quinn, Linda (2017). "Paid sick leave and preventive health care service use among US working adults". *Preventive Medicine* 99, pp. 58–62.

DeRigne, LeaAnne, Stoddard-Dare, Patricia, and Quinn, Linda (2016). "Workers without paid sick leave less likely to take time off for illness or injury compared to those with paid sick leave". *Health Affairs* 35.3, pp. 520–527.

DHHS and NCHS (2012). *National 2009 H1N1 Flu Survey.* https://www.cdc.gov/nchs/nis/data_files_h1n1.htm (last visited: 6/14/2017).

Drago, Robert and Lovell, Vicky (2011). "San Franciscoś Paid Sick Leave Ordinance: Outcomes for Employers and Em ployees".

Funk, Sebastian, Salathé, Marcel, and Jansen, Vincent AA (2010). "Modelling the influence of human behaviour on the spread of infectious diseases: a review". *Journal of the Royal Society Interface*, rsif20100142.

Goad, Jeffery A, Taitel, Michael S, Fensterheim, Leonard E, and Cannon, Adam E (2013). "Vaccinations administered during off-clinic hours at a national community pharmacy: implications for increasing patient access and convenience". *The Annals of Family Medicine* 11.5, pp. 429–436.

Greene, William H (2008). "Econometric Analysis, Sixth Edition". Upper Saddle River, NJ: Pearson Education. Chap. 23. Models for Discrete Choice, pp. 780–781.

Hershey, John C, Asch, David A, Thumasathit, Thi, Meszaros, Jacqueline, and Waters, Victor V (1994). "The roles of altruism, free riding, and bandwagoning in vaccination decisions". *Organizational behavior and human decision processes* 59.2, pp. 177–187.

Hesselius, Patrik (2007). "Does sickness absence increase the risk of unemployment?" *The Journal of Socio-Economics* 36.2, pp. 288–310.

Kim, Namhoon and Mountain, Travis P (2017). "Role of non-traditional locations for seasonal flu vaccination: Empirical evidence and evaluation". *Vaccine* 35.22, pp. 2943–2948.

— (2018). "Do we consider paid sick leave when deciding to get vaccinated?" *Social Science & Medicine* 198, pp. 1–6.

Koop, Gary, Poirier, Dale J, and Tobias, Justin L (2007). *Bayesian econometric methods*. Cambridge University Press.

Kumar, Supriya, Grefenstette, John J, Galloway, David, Albert, Steven M, and Burke, Donald S (2013). "Policies to reduce influenza in the workplace: impact assessments using an agent-based model". *American journal of public health* 103.8, pp. 1406–1411.

Li, Mingliang and Tobias, Justin L (2008). "Bayesian analysis of treatment effects in an ordered potential outcomes model". *Modelling and Evaluating Treatment Effects in Econometrics*. Emerald Group Publishing Limited, pp. 57–91.

Liao, Shaojuan, Ma, Yifei, Chen, Jiangzhuo, and Marathe, Achla (2012). "Paid Sick-Leave: Is It a Good Way to Control Epidemics?" *Complex*. Vol. 12. Springer, pp. 213–227.

Lovell, Vicky (2004). *No time to be sick: Why everyone suffers when workers don't have paid sick leave*. Institute for Women's Policy Research Washington, DC.

Nandram, Balgobin and Chen, Ming-Hui (1996). "Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence". *Journal of Statistical Computation and Simulation* 54.1-3, pp. 129–144.

Nelsen, Maxford (2014). *The Effect of Mandatory Paid Sick Leave Policies - Reviewing the Evidence.*

Nobile, Agostino (2000). "Comment: Bayesian multinomial probit models with a normalization constraint". *Journal of Econometrics* 99.2, pp. 335–345.

OECD (2017). *Influenza vaccination rates (indicator).* http://www.who.int/mediacentre/factsheets/fs211/en/ (last visited: 9/20/2017).

Palache, Abraham, Oriol-Mathieu, Valerie, Fino, Mireli, Xydia-Charmanta, Margarita, et al. (2015). "Seasonal influenza vaccine dose distribution in 195 countries (2004–2013): little progress in estimated global vaccination coverage". *Vaccine* 33.42, pp. 5598–5605.

Peipins, Lucy A., Soman, Ashwini, Berkowitz, Zahava, and White, Mary C. (July 2012). "The lack of paid sick leave as a barrier to cancer screening and medical care-seeking: results from the National Health Interview Survey". *BMC Public Health* 12.1, p. 520. DOI: 10.1186/1471-2458-12-520.

Rolfes, MA, Foppa, IM, Garg, S, Flannery, B, Brammer, L, Singleton, JA, et al. (2016). *Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths Averted by Vaccination in the United States.*

Rosenstock, Irwin M (1974). "The health belief model and preventive health behavior". *Health education monographs* 2.4, pp. 354–386.

Shim, Eunha, Chapman, Gretchen B, Townsend, Jeffrey P, and Galvani, Alison P (2012). "The influence of altruism on influenza vaccination decisions". *Journal of The Royal Society Interface*, rsif20120115.

Verelst, Frederik, Willem, Lander, and Beutels, Philippe (2016). "Behavioural change models for infectious disease transmission: a systematic review (2010–2015)". *Journal of The Royal Society Interface* 13.125, p. 20160820.

Weinberg, Geoffrey A and Szilagyi, Peter G (2010). "Vaccine epidemiology: efficacy, effectiveness, and the translational research roadmap". *The Journal of infectious diseases* 201.11, pp. 1607–1610.

WHO (2016). *Influenza (Seasonal) Fact sheet.* http://www.who.int/mediacentre/factsheets/fs211/en/ (last visited: 9/20/2017).

Williams, Richard et al. (2012). "Using the margins command to estimate and interpret adjusted predictions and marginal effects". *Stata Journal* 12.2, p. 308.

Wilson, Fernando A, Wang, Yang, and Stimpson, Jim P (2014). "Universal paid leave increases influenza vaccinations among employees in the US". *Vaccine* 32.21, pp. 2441–2445.

Yamin, Dan and Gavious, Arieh (2013). "Incentiveséffect in influenza vaccination policy". *Management Science* 59.12, pp. 2667–2686.

Tables

Table 1.1: Descriptive statistics for dependent variable and covariates by income level

Variable Name	Description	All	Low	Middle	High
Vaccination	coverage	44.77%	29.23%	42.16%	51.92%
Gender	0: male, 1: female	55.13%	57.03%	56.84%	52.89%
Age group	0: 18-44y 1: 45-64y	52.26%	36.89%	51.54%	57.51%
Marital status	0: single 1: married	59.73%	25.18%	50.27%	79.21%
Having children	0: no, 1: yes	40.44%	37.41%	36.25%	45.42%
Education	0: below college 1: college	76.86%	46.96%	72.67%	89.81%
Chronic disease	0: no, 1: yes	20.89%	21.52%	21.93%	19.68%
Insurance	0: no, 1: yes	88.10%	56.51%	88.55%	97.05%
Perceived	very low	34.49%	36.23%	32.36%	36.04%
side-effect	low	50.59%	46.04%	52.24%	50.33%
of vaccine	high	10.55%	11.51%	11.03%	9.78%
	very high	4.38%	6.21%	4.36%	3.84%
Perceived	very low	5.97%	12.16%	5.88%	4.23%
effectiveness	low	18.79%	22.30%	19.96%	16.61%
of vaccine	high	30.88%	29.23%	30.59%	31.64%
	very high	44.36%	36.30%	43.58%	47.52%
Perceived	very low	22.06%	26.23%	23.50%	19.43%
probability	low	36.78%	31.72%	37.32%	37.75%
of infection	high	29.91%	27.01%	27.62%	33.00%
without vaccination	very high	11.25%	15.04%	11.55%	9.82%
Paid sick leave	0: no, 1: yes	63.86%	31.85%	63.02%	74.18%

Table 1.2: Posterior means and standard deviation for the outcome equation by income level

Variable	**All**		**Low**		**Middle**		**High**	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Female	0.070**	0.027	0.150*	0.081	0.096**	0.043	0.044	0.040
Old (45-64y)	0.273**	0.029	0.477**	0.081	0.226**	0.045	0.237**	0.044
Married	0.083**	0.029	-0.136†	0.091	0.110**	0.043	0.016	0.052
Having child	-0.002	0.031	0.073	0.090	-0.043	0.049	-0.014	0.046
College	0.181**	0.034	0.183**	0.080	0.046	0.049	0.294**	0.069
Unhealthy	0.133**	0.034	0.179**	0.094	0.132**	0.051	0.117**	0.052
Health insured	0.476**	0.054	0.513**	0.092	0.509**	0.082	0.222*	0.130
Low PSV	-0.301**	0.033	-0.136	0.105	-0.266**	0.052	-0.351**	0.047
High PSV	-0.376**	0.039	-0.135	0.115	-0.413**	0.059	-0.353**	0.060
Very High PSV	-0.603**	0.064	-0.279*	0.143	-0.454**	0.100	-0.829**	0.117
Low PEV	-0.412**	0.088	-0.517**	0.238	-0.341**	0.134	-0.455**	0.140
High PEV	0.104	0.078	-0.013	0.193	0.122	0.118	0.145	0.127
Very High PEV	0.678**	0.084	0.668**	0.206	0.701**	0.126	0.698**	0.139
Low PPFI	0.796**	0.069	0.487**	0.161	0.687**	0.108	1.115**	0.119
High PPFI	1.627**	0.074	0.921**	0.193	1.608**	0.119	2.040**	0.129
Very High PPFI	1.895**	0.101	1.046**	0.235	1.944**	0.159	2.483**	0.197
Low cPPFI	-0.016	0.079	0.225	0.244	-0.096	0.125	-0.149	0.134
High cPPFI	0.031	0.082	0.372	0.243	-0.129	0.132	-0.175	0.143
Very High cPPFI	0.467**	0.120	0.596**	0.297	0.320*	0.193	0.160	0.235
Paid sick leave	0.469**	0.074	0.013	0.211	0.531**	0.110	0.643**	0.128
Covariance: σ_{y,y_1}	-0.195**	0.032	-0.043	0.100	-0.131**	0.049	-0.280**	0.050
Covariance: σ_{y,y_2}	-0.196**	0.033	-0.063	0.107	-0.130**	0.050	-0.280**	0.051

** and * denote high and marginal significance based on the Bayesian significance measure, and numbers in parenthesis () denotes standard deviations. I selected and display three core covariates that were related to the analysis. PSV denotes the perceived side-effect of vaccine. PEV denotes the perceived effectiveness of vaccine. PPFI denotes the perceived probability of flu infection without vaccination. cPPFI denotes the conditional PPFI that represents the PPFI conditional on the fact that paid sick leave is given.

Table 1.3: Marginal effect at means of paid sick leave on vaccination decision by income level

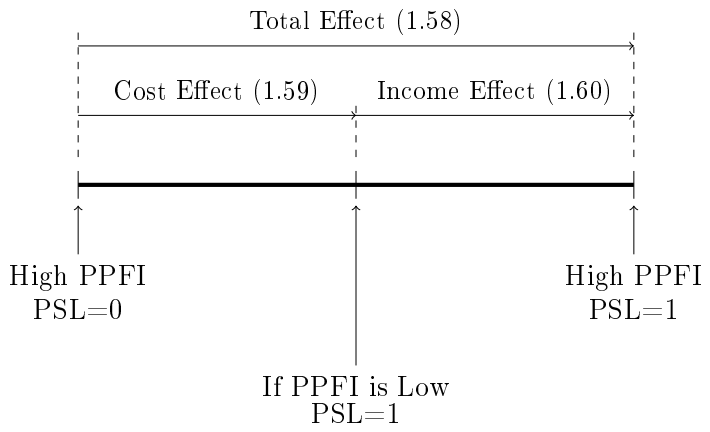
PPFI	**Effect**	**All**		**Low**		**Middle**		**High**	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
Very Low	Total	0.072	0.010	0.002	0.029	0.090	0.017	0.085	0.014
	Income	-	-	-	-	-	-	-	-
	Cost	0.072	0.010	0.002	0.029	0.090	0.017	0.085	0.014
Low	Total	0.152	0.020	0.068	0.061	0.136	0.028	0.179	0.034
	Income	-0.004	0.023	0.067	0.072	-0.023	0.031	-0.044	0.039
	Cost	0.159	0.027	0.008	0.054	0.172	0.040	0.237	0.049
High	Total	0.174	0.023	0.150	0.079	0.145	0.037	0.142	0.036
	Income	0.012	0.031	0.145	0.094	-0.051	0.052	-0.064	0.053
	Cost	0.164	0.024	0.007	0.078	0.184	0.034	0.180	0.032
Very High	Total	0.210	0.029	0.227	0.099	0.178	0.043	0.120	0.043
	Income	0.130	0.033	0.222	0.104	0.086	0.052	0.035	0.051
	Cost	0.130	0.020	0.005	0.082	0.129	0.027	0.104	0.029

Table 1.4: Average margin effect of paid sick leave on vaccination decision by income level

PPFI	**Effect**	**All**		**Low**		**Middle**		**High**	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std
Very Low	Total	0.078	0.011	0.002	0.032	0.095	0.018	0.093	0.015
	Income	-	-	-	-	-	-	-	-
	Cost	0.078	0.011	0.002	0.032	0.095	0.018	0.093	0.015
Low	Total	0.142	0.019	0.064	0.058	0.129	0.027	0.165	0.032
	Income	-0.004	0.022	0.062	0.067	-0.023	0.030	-0.042	0.038
	Cost	0.147	0.025	0.006	0.052	0.162	0.037	0.218	0.045
High	Total	0.157	0.022	0.128	0.068	0.132	0.034	0.134	0.034
	Income	0.010	0.028	0.123	0.080	-0.045	0.046	-0.059	0.048
	Cost	0.148	0.022	0.006	0.067	0.168	0.032	0.172	0.031
Very High	Total	0.206	0.027	0.200	0.089	0.177	0.041	0.127	0.043
	Income	0.122	0.031	0.196	0.093	0.081	0.049	0.035	0.051
	Cost	0.123	0.019	0.005	0.072	0.125	0.025	0.108	0.027

Figures

Figure 1.1: Estimation strategy for effect decomposition



Chapter 2

Financial Knowledge, "Don't Know" Response, and Misunderstood Bias: Binomial Latent Regression Approach

2.1 Introduction

Many studies have argued that sufficient financial knowledge improves appropriate financial decisions and behaviors, in particular, in the case of young adults (Lusardi and Mitchelli 2007; Hilgert, Hogarth, and Beverly 2003; Robb and Sharpe 2009; Robb and Woodyard 2011). However, people in the United States appear to lack satisfactory knowledge about personal finance and are less prepared for proper financial behaviors, such as proper behaviors of using credit cards, debt management, and long-term financial planning, compared with those in other countries (Klapper, Lusardi, and Van Oudheusden 2015; OECD 2016a; OECD 2017).

One of the popular surveys regarding financial knowledge is the state-by-state National Financial Capability Study (NFCS) survey conducted by the Financial Industry Regulatory Authorities (FINRA) Investor Education Foundation (Foundation 2016). Although it has been used for many studies, a critical concern exists in this survey. The factual questions that assess respondents' financial knowledge include "Don't Know" or "Refuse" (DK/RF)

responses along with substantive responses. Most existing studies that have examined this survey data have regarded DK/RF responses as incorrect answers and use binary or binomial regression models with a probit or logit link (Lusardi and Mitchell 2011; Bassa Scheresberg 2013; Xiao, Chen, and Sun 2015; Johnson and Lamdin 2015). This procedure is likely to neglect the effect of DK/RF responses and lead to misleading conclusions if there exist non-random choices of DK/RF responses.

The effect of DK/RF responses in factual questions used to determine respondents' knowledge level has been analyzed primarily in political science (Delli Carpini and Keeter 1993; Mondak 1999; Mondak 2001; Mondak and Davis 2001). The authors argue that when a respondent is not fully knowledgeable, s/he has two possible options to choose: a DK/RF response or substantive answer by guessing randomly. This decision decreases the validity of using the number of correct answers as a dependent variable because it ignores the unobserved difference between a substantive answer and DK/RF response in political knowledge questions.

Thus, the purpose of this study is to suggest a new regression model that prevents a distortion of DK/RF responses. in proposing my econometric models, I focus on the unobserved true number of correct answers rather than those observed. This identifies each individual's actual financial knowledge and estimates the true parameters that are hidden in DK/RF responses. The second purpose of this study is to investigate the actual effect of educational interventions and group differences on the level of objective financial knowledge. I estimate the coefficients of interest with a typical binomial regression model and my suggested models. Then, I compare these estimates to discover what the typical binomial regression model ignores which results in unwanted conclusions about the level of objective financial knowledge.

In the following section, I review the literature related to financial knowledge definition, the appropriate measure of knowledge level, and methods used to control DK/RF responses

and propose several assumptions. In the subsequent two sections, I describe the survey data and the econometric models used in this study. This section includes a simulation study to test my suggested econometric models. In the results and discussion sections, I present the results of the estimations and discuss the advantages of my suggested models. The financial section provides conclusions and policy implications resulting from this study.

2.2 Background Knowledge and Literature Review

Financial knowledge and literacy have been defined in terms of human capital. Financial knowledge is the stock of knowledge or human capital about personal finance, obtained by education and experience (Huston 2010; OECD 2016b). Financial literacy, considers another dimension, that of the individual's confidence and ability to apply financial knowledge to appropriate financial decisions and activities, as well as the proper understanding of financial information.

Financial education is one of the influential and popular interventions used to instill appropriate financial behaviors and improve financial literacy. However, the effect of these education programs on financial knowledge is unclear. Some researchers emphasize the effectiveness of financial educations in schools (Danes, Huddleston-Casas, and Boyce 1999; Varcoe et al. 2005; Peng et al. 2007; Walstad, Rebeck, and MacDonald 2010). They argue that financial education accumulates financial knowledge and it would result in better financial literacy and financial decision making over the long-term. In contrast, other researchers disagree with the effectiveness of the financial education in schools (Mandell 2006; Mandell and Klein 2007; Fernandes, Lynch Jr, and Netemeyer 2014; Willis 2008; Willis 2011). Thus, financial education would be less effective than researchers expected, and the corresponding cost would be too high given limited improvements in financial literacy.

Political science studies have investigated the effect of DK/RF responses to factual questions to measure the actual individual political knowledge in survey data. A conventional procedure used by political science researchers recommends that survey questionnaires should encourage DK/RF responses to factual questions. This procedure asserts that encouraging DK/RF responses increases the reliability of knowledge scales by reducing noise and avoiding random guessing in surveys (Carpini and Keeter 1993).

However, several studies have asserted that including DK/RF answers in survey questions is problematic when a count of correct answers is used as a measure of political knowledge. First, researchers could suffer from complex underlying knowledge states in political studies if DK/RF responses are encouraged (Mondak and Davis 2001). As shown in Table 2.1, Mondak and Davis (2001) assume four behavioral knowledge states: fully informed (knowing the correct answer), partially informed (having incomplete knowledge), misinformed (misunderstanding the correct answer), and uninformed (having no knowledge). The difficulty in measuring the proper knowledge level is from the fact that three possible choices - correct, incorrect, and DK/RF - cannot identify these knowledge states perfectly. Thus, adding DK/RF responses to factual questions could worsen the identification of the appropriate level of respondents' knowledge.

[Table 2.1 about here]

Second, respondents have different propensities to choose DK/RF responses rather than selecting a substantive answer despite the same level of knowledge (Mondak 2001). For example, two respondents with identical knowledge levels would obtain different scores if one has a higher propensity to choose DK/RF answer than does the other. Thus, Mondak (2001) suggests that DK/RF responses should be discouraged when political knowledge is measured in surveys.

However, not all researchers agree with this argument. A small experimental study

conducted in the UK provided evidence that partially knowledgeable individuals do not intend to conceal their knowledge level with DK/RF responses (Sturgis, Allum, and Smith 2007). From the interview of 1,006 respondents via telephone in 2014, Sturgis, Allum, and Smith (2007) find that a respondent who chooses DK/RF responses has no greater chance to choose correct answers even after they are forced to choose a substantive answer.

2.3 Econometric Model

2.3.1 Model Assumptions

Suppose that I evaluate the effect of educational interventions or predictors on the level of objective financial knowledge. A typical dependent variable would be the observed number of correct answers. Typically, researchers then analyze the change in the probability of answering correctly and conclude that the interventions or predictors affected objective financial knowledge based the probability estimated. However, if the factual questions in the survey include DK/RF responses, respondents can obtain another chance of choosing these responses. In this case, researchers could under- or overestimate the effect if the distribution of the DK/RF responses is not random. Thus, I propose several assumptions about the statistical distribution and dependent variable to evaluate actual effectiveness properly. First, I need distributional assumptions about the answers to financial questions, as follows:

A1: The difficulty of all financial questions in a survey is equal.

This assumption is very strong because there is no standard method to evaluate the difficulty of questions perfectly. The purpose of considering the number of correct answers is to evaluate the level of objective financial literacy. Thus, instead of A1, I assume that, for evaluators, each question has the same weight to represent the level of objective

financial knowledge. I suggest a general probability of answering correctly that represents the participants' overall level of objective financial knowledge rather than the equal level of difficulty. Thus, I suggest an alternative assumption for A1, as follows:

A1a: Each question has the same weight for the level of objective financial knowledge.

A2: The number of correct answers for each individual is binomially distributed and random.

Based on these two assumptions, I can apply a binomial regression model to estimate the effect of the educational interventions or covariates of interest. Second, I also need latent variable assumptions to evaluate the actual effect from the observed outcomes, as follows:

A3: The true number of correct answers for each individual includes the observed number of correct answers and the number of correct answers hidden in the DK/RF responses.

A4: These numbers of correct answers are independent.

2.3.2 Model Specification

Let Q_i and Q_i^{DK} be the total number of questions and DK/RF responses, respectively. Further, let p_i^{true} and p_i^{DK} be the probabilities of answering correctly for the true number of correct answers and the unobserved number of correct answers hidden in DK/RF responses, respectively. Then, A1a and A2 indicate that the random variable y_i^{true} and y_i^{DK} are given, respectively, by:

$$\begin{aligned} y_i^{true} &\sim Bin(Q_i, p_i^{true}) \\ y_i^{DK} &\sim Bin(Q_i, p_i^{DK}) \end{aligned} \tag{2.1}$$

Moreover, A3 also indicates that the number of correct answers observed is equal to the true number of correct answers, with the number of correct answers unobserved in DK/RF responses subtracted. Let y_i^{obs} and y_i^{true} be the observed and true number of correct answers

for each individual i, respectively. Also, let y_i^{DK} be the number of correct answers hidden in DK/RF responses. Then, I can specify A3 by:

$$y_i^{obs} = y_i^{true} - y_i^{DK} \tag{2.2}$$

While the actual support of y_i^{obs} is $(-Q_i^{DK}, Q_i)$, this study focuses on a positive support $(0, Q_i)$ only. The probability density function of the binomial distribution for a number of successes y with a number of trials Q , and a success probability p is given by:

$$f(y;\theta) = \binom{Q}{y} p^y (1-p)^{Q-y} \tag{2.3}$$

where θ denotes all parameters in this distribution. Based on this probability density function, the probability density function for the number of answering correctly observed, y_i^{obs}, is calculated by:

$$\begin{aligned} f(y_i) &= \sum_{k=0}^{Q_i^{DK}} f(y_i^{true} = y_i^{obs} + k,\ y_i^{DK} = k) \\ &= \sum_{k=0}^{Q_i^{DK}} f(y_i^{true} = y_i^{obs} + k) f(y_i^{DK} = k) \\ &= \sum_{k=0}^{Q_i^{DK}} \left[\begin{array}{l} \binom{Q_i}{y_i^{obs}+k} (p_i^{true})^{y_i^{obs}+k} (1-p_i^{true})^{Q_i - y_i^{obs} - k} \\ \cdot \binom{Q_i^{DK}}{k} (p_i^{DK})^k (1-p_i^{DK})^{Q_i^{DK}-k} \end{array} \right] \end{aligned} \tag{2.4}$$

This density function indicates the expected number of correct answers for each individual over the number of correct answers observed.

To estimate the regression parameters of the individual effect of the treatments and covariates, I attach a vector parameter, $\boldsymbol{\beta}$, that represents the effect of educational

interventions and group differences, x_i, with the true number of correct answers, as follows:

$$p_i^{true} = \Phi(x_i'\boldsymbol{\beta}) \tag{2.5}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative density function as a link function. Finally, to estimate the parameter $\boldsymbol{\beta}$, I construct a likelihood function with the number of correct answers observed with the probability density function as follows:

$$\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{y}^{obs}, \boldsymbol{X}) \propto \prod_{i=1}^{n} \sum_{k=0}^{Q_i^{DK}} \left[\begin{array}{l} \binom{Q_i}{y_i^{obs}+k} \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})^{y_i^{obs}+k} \Phi(-\boldsymbol{x}_i'\boldsymbol{\beta})^{Q_i - y_i^{obs} - k} \\ \cdot \binom{Q_i^{DK}}{k} (p_i^{DK})^k (1 - p_i^{DK})^{Q_i^{DK}-k} \end{array} \right] \tag{2.6}$$

A critical issue for the parameter estimation is to determine the probability of answering correctly in DK/RF responses, p_i^{DK}. I suggested two methods to define this probability. For the first method, I assume that a respondent who choose a DK/RF response does not know the correct answer and s/he would choose one of the substantive responses randomly if s/he is forced to pick one rather than the DK/RF responses. In this case, the probability of answering correctly in the absence of the DK/RF response is fixed as the probability of randomly guessing, as follows:

$$p_i^{DK} = \bar{p}_i \tag{2.7}$$

I refer to this case as Model I.

On the other hand, the second model assumes that a respondent who selects DK/RF responses would have the equivalent probability of answering correctly as for the true responses. Thus, I rewrite the probability of answering correctly in DK/RF responses, as follows:

$$p_i^{DK} = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta}) (= p_i^{true}) \tag{2.8}$$

I refer to this case as Model II.

I use maximum likelihood estimation (MLE) to obtain the parameter estimates, $\hat{\boldsymbol{\beta}}$, and the corresponding standard errors by using the inverse of the Hessian matrix achieved in the estimation. Moreover, I also estimate the marginal effects and the corresponding standard errors. All marginal effects are from dummy variables. Thus, the standard error for the marginal effects can be evaluated from the estimation of the asymptotic variances as follow:

$$aVar(\Delta\hat{F}) = \left(\frac{\partial\Delta\hat{F}}{\partial\hat{\beta}}\right)' \boldsymbol{V} \left(\frac{\partial\Delta\hat{F}}{\partial\hat{\beta}}\right) \tag{2.9}$$

where $\Delta\hat{F}$ denotes the estimated marginal effect for a dummy variable, and $\boldsymbol{V}$ denotes a asymptotic variance of $\hat{\beta}$ in which I am interested in (Greene 2008).

2.3.3 Simulated Data Experiment

I carry out a simulation to examine the performance of the suggested model with the MLE. For the true number of correct answers, I generate $1,000$ observations with the total number of trials, $Q = 10$, the regression parameter, $\beta = (1.2, 2.1, 0.8, -1.5)'$, and the probability of success, $p_i = \Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$.

For the number of correct answers hidden in the DK/RF responses, I add two restrictions. First, for Model I, I set the probability of success to arbitrary random chance, say, $p_i^{DK} = 0.25$, and generate the number of correct answers hidden in the DK/RF responses, y^{DK}, with this probability. Then, I calculate the number of correct answers observed by subtracting y^{DK} from y^{true}. I estimate the true regression parameters, $\boldsymbol{\beta}$, by using only the number of correct answers observed that is generated by the simulation procedure. As Table 2.2 indicates, the suggested model identifies the true parameters. I obtain very precise estimates and posterior means for the true parameters.

Second, for Model II, the simulation data are generated with the same parameters and procedures as in Model I, except for the probability of success in the DK/RF responses,

$p_i^{DK} = \Phi(\boldsymbol{x}_i\boldsymbol{\beta})$. I estimate the true regression parameters with the generated data and obtain the estimates and posterior means. As Table 2.2 shows, the suggested model with MLE identifies the true parameters successfully.

[Table 2.2 about here]

2.4 Data and Measures

I examine the 2015 NFCS survey data that included a total sample size of $27,564$. I restrict the analysis to young adults whose ages range from 18 to 34. The number of states that offer financial courses or require students to take them has increased since approximately 2000 (CFEE 2016). Thus, I use only the young adult sample whose ages range from 18 to 34 for this study. Moreover, I also focus on financial education in high school and college, excluding the workplace and military facilities. The military sample is too small to obtain reliable estimates. I am also unsure whether education in the workplace was actually related to personal finance. Therefore, the total number of observations used in this study is $5,951$.

2.4.1 Dependent variable

The dependent variable is the level of objective financial knowledge. In this study, I measure financial literacy by evaluating the score based on a personal finance quiz with six multiple-choice questions, as follows:

1. **Interest rate question**: Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?

2. **Inflation question**: Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to

buy with the money in this account?

3. **Bond price question**: If interest rates rise, what will typically happen to bond prices?

4. **Interest compounding question**: Suppose you owe $1,000 on a loan and the interest rate you are charged is 20% per year compounded annually. If you didn't pay anything off, at this interest rate, how many years would it take for the amount you owe to double?

5. **Mortgage question**: A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less.

6. **Risk question**: Buying a single company's stock usually provides a safer return than a stock mutual fund.

There are three possible answers for the interest rate and inflation question, in addition to the DK/RF options. There are four possible answers to the bond price and compounding interest questions, in addition to the DK/RF responses. Also, there are two possible answers to the mortgage and risk questions, in addition to the DK/RF choices. The compounding interest question is a new one added to the survey conducted in 2015.

As Table 2.3 shows, I observe many DK/RF answers. In this table, the numbers in diagonal cells represent the case where respondents choose substantive answers only. In contrast, the numbers in off-diagonal cells represent the case where respondents choose at least one DK/RF response instead of choosing substantive answers. Thus, if I use the number of correct answers by considering substantive responses, I could ignore the true information of respondents' financial knowledge level and reach the misleading results. To avoid this unwanted situation, I count both the number of correct answers and the number of DK/RF

answers observed in the subsample and use both as dependent variables with my suggested model.

[Table 2.3 about here]

2.4.2 Independent Variables

I consider several independent variables, including socio-demographic factors, self-assessed skills, and formal and informal financial education. I drop any observations in which a DK/RF response appeared for any independent variable outlined below.

First, the socio-demographic factors include age group (18-24 years or 25-34 years), gender (male or female), race (white or non-white), education level (attended high school only or some college including associate degree, or college or higher degree), and household income (high$> \$25,000$ and low $\leq \$25,000$).

Second, I consider two self-assessed abilities as independent variables: mathematical skill and financial knowledge level. The question, "How strongly do you agree or disagree with the following statements? I am pretty good at math" evaluates the self-assessed math skill. The possible responses are measured on seven scales that ranged from "1- Strongly Disagree" to "7- Strongly Agree." If respondents answer "7-Strongly Agree," which is above the median ($= 6$), I classify them as having "High math skill." For the self-assessed financial knowledge level, respondents are asked, "On a scale from 1 to 7, where 1 means very low and 7 means very high, how would you assess your overall financial knowledge?" Then, they select answers on scales that range from "1-Very Low" to "7-Very High." If they have a higher level of self-assessed financial knowledge than the median ($= 5$), they are regarded as having "High subjective knowledge."

For the educational interventions, I consider several variables, including whether a respondent has rejected any financial education offers, two types of formal financial

education (in high school or in college), and informal education provided by parents or guardians. Respondents are asked, "Was financial education offered by a school or college you attended, or a workplace where you were employed?" There are three possible answers, "Yes, but I did not participate in the financial education of," "Yes, and I did participate in the financial education," and "No." If they choose the first answer, I scored them as "rejecting financial education." If they choose the second answer, they are allowed to move to other questions that asked, "When did you receive that financial education?" The possible responses to this question are, "In high school," "In college," "From an employer," and "From the military." If they choose "In high school," I label them as "Having formal education in high school." If they select "In college," I label them as "Having formal education in college." As mentioned previously, I disregard the last two answers and drop those observations. Another education treatment is informal financial education. In this survey, respondents are asked, "Did your parents or guardians teach you how to manage your finances?" The responses are "Yes" or "No." If they choose "Yes," I regard them as "Having informal education."

2.5 Results

2.5.1 Sample Descriptive Statistics

The descriptive statistics indicates that the concern that DK/RF responses should be considered is justified. Table 2.4 and 2.5 show the mean of the independent variables by the number of correct and incorrect answers, respectively.

[Table 2.4 and 2.5 about here]

Although the means of covariates increase (decrease) as the numbers of correct answered increase (decrease), I observe that the means of covariates increase (decrease) as the numbers

of incorrect answers increase (decrease). For example, the mean of the binary variable for informal education increases monotonically from 0.444 to 0.566 as the number of correct answers increases. However, I found that the mean of this variable increases from 0.500 to 0.549 as the number of incorrect answers increases. I also found a similar pattern in gender. In the female sample, both the number of correct and incorrect answers decrease. I observe this pattern in several other covariates and conclude that researchers would obtain misleading results if they use the number of correct answers as a dependent variable and ignore DK/RF responses.

2.5.2 Estimation Results

I estimate the parameters using three econometric regression models for each estimation method: binomial regression with the probit link (Reference model), binomial-latent regression with a random probability (Model I), and binomial-latent regression with true probability (Model II). Table 2.6 and 2.7 show the estimates of the coefficient and the corresponding marginal effect on the probability of success estimated, respectively, by the MLE method. These tables also include the standard errors of the estimates and marginal effects.

[Table 2.6 to 2.7 about here]

From these coefficient estimates and the corresponding marginal effects, I observe that the results from the suggested model are inconsistent with those from the reference model. First, I investigate the existence of demographic group differences in the level of objective financial knowledge. I observed no gender difference in the level of objective financial knowledge in Model II, while there is a significant gender difference when I use the reference model. I find that female respondents have a 1% lower probability of answering correctly than what male respondents do. On the other hand, if I use the reference model, I observe that female

respondents have a 5.5% lower probability of answering correctly than what male respondents do. Model I still generates a significant decrease in the probability that female respondents would answer correctly. However, this is about 2.3% less than in the reference model.

In addition, I find no income difference in my suggested models. Low income is a significant factor that undermined the level of objective financial knowledge in the reference model. However, when I apply the suggested models, the effect on the level of objective financial knowledge decreases or disappears. If a respondent has a low income, less than $\$25,000$, then s/he has a 4.2% lower probability of answering correctly according to the reference model. However, this probability is only 2.3% and 1.1% when I use the suggested regression models. Thus, it is necessary to consider DK/RF responses because the decision between substantive and DK/RF answers has a significant effect in this sample clearly.

Second, I analyze the effect of self-assessed skills on financial knowledge. High self-assessed financial knowledge would be a case opposite to that of the demographic factors mentioned above. This variable does not significantly increase the level of objective financial knowledge relative to the reference model. There is only a 1.17% decrease in the knowledge level if a respondent has high self-assessed financial knowledge. However, if I consider DK/RF responses, I find that this variable significantly reduces the level of objective financial knowledge by 5.3% for Model I and 11.7% for Model II.

Furthermore, I evaluate the effectiveness of informal and formal financial education as educational interventions. The most important educational variable in this sample is whether respondents reject educational offers or not. If a respondent rejects any educational offers, his/her objective financial knowledge insignificantly decreases by 1.4% in the reference model compared to no financial education. However, in my suggested models, this variable is highly significant to decrease the level of objective financial knowledge compared to no financial education. A rejection of financial education decreases the level of objective financial knowledge by 4.4% and 8.7% in Model I and Model II when DK/RF responses are considered.

Both of these effects are highly significant.

Financial education provided in college is not significant in the reference model or in Model I, while it is marginally significant in Model II and shows a negative effect on the level of objective financial knowledge compared to no financial education. If a respondent receives financial education in college, this education reduces his/her level of objective financial knowledge by 2.2%. This decrease indicates that college financial curricula have an adverse effect on the level of objective financial knowledge.

Informal financial education is highly significant compared to no financial education when the reference model is used for this sample. A respondent who has received informal education shows a 2.5% higher probability of answering correctly than does one who has not. However, the suggested models show only marginal effects with 1.5% increases in the probability of answering correctly compared with no financial education.

2.6 Discussion

First, the suggested regression models show the advantage of identifying the hidden preference between guessing randomly and choosing a DK/RF response. The levels of significance of regression parameters and the corresponding marginal effects in the reference model differ from those in the suggested models. This suggests that researchers cannot neglect DK/RF responses when they evaluate group differences and the effect of educational interventions. I review several papers that discuss the importance of DK/RF responses to factual questions in survey data. If the preference to answer DK/RF is not random, researchers should consider these responses to avoid obtaining unreliable results and making misleading interpretations. Thus, they need to identify the number or proportion of DK/RF responses and appropriately address this issue.

Second, I discuss the findings in the suggested models that the reference model fails to

find, including group differences and the effectiveness of the educational interventions on the level of objective financial knowledge.

The suggested models, especially Model II, indicates that there is no gender difference in the level of objective financial knowledge among young adults. There is a highly significant difference in objective financial knowledge between males and females according to the reference model. However, because DK/RF responses are significant in explaining female characteristics, researchers should have considered the DK/RF responses seriously. The underestimation of female financial knowledge is attributable to their preference to select DK/RF responses rather than guessing randomly and selecting substantive answers, even though their ability does not differ from that of males. Thus, if they choose a substantive answer rather than a DK/RF response, their true understanding of personal finance would be equal to that of males, which would eliminate the gender bias in the level of financial knowledge. the suggested models identify this tendency and draw more accurate conclusions because researchers consider the correct information hidden in DK/RF responses. Several studies have investigated the gender difference in objective financial knowledge (Chen and Volpe 2002; Danes and Haberman 2007; Fonseca et al. 2012), scholastic Assessment Test (SAT) (Baldiga 2013), and the undergraduate-level mathematics test (Anderson 1989). However, many studies have argued that females prefer to choose DK/RF responses rather than substantive answers. A study evaluates the gender gap in an educational science test and finds that the difference is attributable to the higher tendency to answer DK/RF than to any actual knowledge difference (Linn et al. 1987). Furthermore, the conventional belief that females have a lower level of political knowledge than males do is illusory because researchers have not considered female respondents' high propensity to choose DK/RF answers (Mondak and Anderson 2004; Lizotte and Sidman 2009). In addition, another study finds that when there is a small penalty for wrong answers, women answer significantly fewer questions than men do (Baldiga 2013).

To my knowledge, a direct effect of low income on the level of objective financial knowledge has not been investigated. In this study, the reference model concludes that low household income reduces the level of objective financial knowledge. This appears to be reasonable because low household income would offer fewer educational opportunities and, in turn, decreases the knowledge level, including financial knowledge. However, the suggested models indicate that income level has no effect on the level of objective financial knowledge.

The estimation result from the suggested models also indicates that whether a respondent rejects or accepts financial education is a very significant factor that affects young adults' level of objective financial knowledge. The reference model fails to address this argument because it neglects the information in the DK/RF responses. Respondents who reject any formal financial education had the low probability of answering correctly. Furthermore, I find an adverse effect in which their financial knowledge is less than the knowledge of those who have not received any educational offers. Thus, I conclude that it may produce a selection problem, in which less-motivated or previously less-informed students reject educational offers and are highly likely to accumulate insufficient knowledge in personal finance. There is an empirical analysis that mandating financial education helps individuals accumulate financial knowledge and gradually affects financial behaviors in the long term (Bernheim, Garrett, and Maki 2001). Moreover, another study finds that states with a specific financial education mandate generates higher scores than do those with either a general or no mandate (Tennyson and Nguyen 2001). However, if researchers ignore the effect of DK/RF responses and use the reference model, they could reach a misleading conclusion about educational policy opposite to these findings in which rejecting educational offers does not lower the level of objective financial knowledge significantly. This result would mislead policymakers. In the United States, twenty-two states offer a high school course in personal finance, while seventeen of these states require students to take a course in personal finance to graduate

high school in 2015 (CFEE 2016).

In addition, I find that if the DK/RF responses are considered in the suggested models, formal and informal financial education do not affect the level of objective financial knowledge as much the reference model does. Moreover, I observe that college financial education have more adverse effects, in which it decreases the level of objective financial knowledge by 2.2% compared to no financial education. Thus, researchers could place too much confidence in financial education and curricula if the DK/RF responses are ignored in surveys. The suggested models can catch this kind of error and provide more reliable results to policymakers than the current model. Some studies have argued that financial curricula in high school and college affect objective financial knowledge positively, and improve financial literacy. However, other studies also have argued that financial education is effective only in the short term (Mandell 2006; Mandell and Klein 2007; Mandell 2008; Fernandes, Lynch Jr, and Netemeyer 2014), and ineffective methods of delivery undermine its effectiveness. The result that formal education is less effective than expected addresses these arguments by considering respondents' DK/RF preference (Willis 2008; Willis 2011).

In addition, high self-assessed financial knowledge is a significant factor that influences the level of objective financial knowledge. However, I cannot conclude that there is a causal relationship between high self-assessed financial knowledge and the level of objective financial knowledge. Instead, this estimation argues that self-assessed financial knowledge could be a critical factor in choosing DK/RF responses or randomly guessing. If respondents perceive that they are highly knowledgeable regardless of their actual knowledge, they may be more likely to choose substantive responses randomly. Thus, if this preference is considered, I would observe a high negative correlation between self-assessed and actual levels of objective financial knowledge.

2.7 Data Limitations

I face several limitations from the survey data used in this study. First, respondents are not asked to indicate which type of education they have rejected, high school or college. Second, they were not instructed to indicate what education is provided in high school and college financial education courses. Each state and college has various financial and economic curricula and standards for their student that the survey data does not capture. Moreover, it does not indicate whether accepting financial education is forced or voluntary. This restriction prevents us from analyzing the effect of mandatory financial education on the level of objective financial knowledge. In addition, the survey does not include information about the states in which respondents received their financial education. Thus, I am unable to evaluate the effect of regional characteristics, including required versus voluntary education on the level of objective financial knowledge.

2.8 Conclusions

In my analysis, I find several significant results hidden in DK/RF responses in the survey when I use an appropriate econometric model. I propose appropriate regression models and compare them to the reference model researchers typically have used to analyze this survey data. My suggested models reveal several interesting results that are different from those included in the reference model. For the educational interventions, rejecting financial education offers is a disadvantage in acquiring knowledge. I find that less motivated or less informed students reject education offers and are highly likely to be less knowledgeable about personal finance. Furthermore, I find from the suggested model that formal and informal education would less effective to improve the level of financial knowledge, which is different results from the reference model. For group differences, I observe that my models show few

or no gender, income, and age differences compared to the significant findings the reference model obtains.

Some policy implications are related to this study. DK/RF responses generate a critical problem for researchers and policymakers. Adding DK/RF responses to factual questions in the survey could yield unexpected and misleading conclusions if these responses are not random. Thus, researchers need to argue whether or not factual questions in surveys include DK/RF responses to address the selection problem. Also, we need a policy that requires students to take a course in financial education before graduation rather than simply offering the education. A voluntary financial education course decreases financial knowledge level for those who decline to take a personal finance course. Thus, mandated school curricula in personal finance would likely improve financial knowledge levels and decision-making among participants.

References

Anderson, Johnston (1989). "Sex-related differences on objective tests among undergraduates". *Educational Studies in Mathematics* 20.2, pp. 165–177.

Baldiga, Katherine (2013). "Gender differences in willingness to guess". *Management Science* 60.2, pp. 434–448.

Bassa Scheresberg, Carlo de (2013). "Financial literacy and financial behavior among young adults: Evidence and implications". *Numeracy* 6.2, p. 5.

Bernheim, B Douglas, Garrett, Daniel M, and Maki, Dean M (2001). "Education and saving: The long-term effects of high school financial curriculum mandates". *Journal of public Economics* 80.3, pp. 435–465.

Carpini, Michael X Delli and Keeter, Scott (1993). "Measuring political knowledge: Putting first things first". *American Journal of Political Science*, pp. 1179–1206.

CFEE (2016). *Survey of the States: The state of K-12 economic and financial education in the United States*. http://councilforeconed.org/wp/wp-content/uploads/2016/02/sos-16-final.pdf (last visited: 8/26/2017).

Chen, Haiyang and Volpe, Ronald P (2002). "Gender differences in personal financial literacy among college students". *Financial services review* 11.3, p. 289.

Danes, Sharon M and Haberman, Heather (2007). "Teen financial knowledge, self-efficacy, and behavior: A gendered view".

Danes, Sharon M, Huddleston-Casas, Catherine, and Boyce, Laurie (1999). "Financial planning curriculum for teens: Impact evaluation". *Journal of Financial Counseling and Planning* 10.1, p. 26.

Delli Carpini, Michael X and Keeter, Scott (1993). "Measuring political knowledge: Putting first things first". *American Journal of Political Science*, pp. 1179–1206.

Fernandes, Daniel, Lynch Jr, John G, and Netemeyer, Richard G (2014). "Financial literacy, financial education, and downstream financial behaviors". *Management Science* 60.8, pp. 1861–1883.

Fonseca, Raquel, Mullen, Kathleen J, Zamarro, Gema, and Zissimopoulos, Julie (2012). "What explains the gender gap in financial literacy? The role of household decision making". *Journal of Consumer Affairs* 46.1, pp. 90–106.

Foundation, FINRA Investor Education (2016). *2015 State-by-State Survey National Financial Capability Study.*

Greene, William H (2008). "Econometric Analysis, Sixth Edition". Upper Saddle River, NJ: Pearson Education. Chap. 23. Models for Discrete Choice, pp. 780–781.

Hilgert, Marianne A, Hogarth, Jeanne M, and Beverly, Sondra G (2003). "Household financial management: The connection between knowledge and behavior". *Fed. Res. Bull.* 89, p. 309.

Huston, Sandra J (2010). "Measuring financial literacy". *Journal of Consumer Affairs* 44.2, pp. 296–316.

Johnson, Mark A and Lamdin, Douglas J (2015). "Gauging the financial capability of Americans". *Business Economics* 50.3, pp. 135–146.

Klapper, Leora, Lusardi, Annamaria, and Van Oudheusden, Peter (2015). "Financial Literacy Around the World". *Standard & Poorś Ratings Services Global Financial Literacy Survey.* http://gflec.org/wp-content/uploads/2015/11/Finlit_paper_16_F2_singles.pdf (last visited: 8/26/2017).

Koop, Gary, Poirier, Dale J, and Tobias, Justin L (2007). *Bayesian econometric methods.* Cambridge University Press.

Linn, Marcia C, De Benedictis, Tina, DeLucchi, Kevin, Harris, Abigail, and Stage, Elizabeth (1987). "Gender differences in national assessment of educational progress science items:

What does ‘i don’t know’ really mean?” *Journal of Research in Science Teaching* 24.3, pp. 267–278.

Lizotte, Mary-Kate and Sidman, Andrew H (2009). “Explaining the gender gap in political knowledge”. *Politics & Gender* 5.2, pp. 127–151.

Lusardi, Annamaria and Mitchell, Olivia S (2011). “Financial literacy and retirement planning in the United States”. *Journal of Pension Economics & Finance* 10.4, pp. 509–525.

Lusardi, Annamaria and Mitchelli, Olivia S (2007). “Financial literacy and retirement preparedness: Evidence and implications for financial education”. *Business economics* 42.1, pp. 35–44.

Mandell, Lewis (2006). “Financial Literacy: If It’s So Important, Why Isn’t It Improving?”

— (2008). “The financial literacy of young American adults”. *The Jumpstart Coalition for Personal Financial Literacy.*

Mandell, Lewis and Klein, Linda Schmid (2007). “Motivation and financial literacy”. *Financial services review* 16.2, p. 105.

Mondak, Jeffery J (1999). “Reconsidering the measurement of political knowledge”. *Political Analysis* 8.1, pp. 57–82.

Mondak, Jeffery J and Anderson, Mary R (2004). “The knowledge gap: A reexamination of gender-based differences in political knowledge”. *The Journal of Politics* 66.2, pp. 492–512.

Mondak, Jeffery J and Davis, Belinda Creel (2001). “Asked and answered: Knowledge levels when we will not take ‘dońs know’ for an answer”. *Political Behavior* 23.3, pp. 199–224.

Mondak, Jeffrey J (2001). “Developing valid knowledge scales”. *American Journal of Political Science*, pp. 224–238.

OECD (2016a). “OECD/INFE International Survey of Adult Financial Literacy Competencies”.

OECD (2016b). “PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy”.

— (2017). “PISA 2015 Results (Volume IV)”.

Peng, Tzu-Chin Martina, Bartholomae, Suzanne, Fox, Jonathan J, and Cravener, Garrett (2007). “The impact of personal finance education delivered in high school and college courses”. *Journal of family and Economic Issues* 28.2, pp. 265–284.

Robb, Cliff A and Sharpe, Deanna L (2009). “Effect of personal financial knowledge on college studentsćredit card behavior”.

Robb, Cliff A and Woodyard, Ann (2011). “Financial knowledge and best practice behavior”.

Sturgis, Patrick, Allum, Nick, and Smith, Patten (2007). “An experiment on the measurement of political knowledge in surveys”. *Public Opinion Quarterly* 72.1, pp. 90–102.

Tennyson, Sharon and Nguyen, Chau (2001). “State curriculum mandates and student knowledge of personal finance”. *Journal of Consumer Affairs* 35.2, pp. 241–262.

Varcoe, Karen P, Martin, Allen, Devitto, Zana, and Go, Charles (2005). “Using a financial education curriculum for teens”.

Walstad, William B, Rebeck, Ken, and MacDonald, Richard A (2010). “The effects of financial education on the financial knowledge of high school students”. *Journal of Consumer Affairs* 44.2, pp. 336–357.

Willis, Lauren E (2008). “Against financial-literacy education”. *Iowa L. Rev.* 94, p. 197.

— (2011). “The financial education fallacy”. *The American Economic Review* 101.3, pp. 429–434.

Xiao, Jing Jian, Chen, Cheng, and Sun, Lei (2015). “Age differences in consumer financial capability”. *International journal of consumer studies* 39.4, pp. 387–395.

Tables

Table 2.1: Relation between knowledge states and responses to multiple-choice questions

<table>
<tr><th rowspan="2">Knowledge States</th><th colspan="3">Possible Answers</th></tr>
<tr><th>Answer Correctly</th><th>Answer Incorrectly</th><th>Answer DK/RF</th></tr>
<tr><td>Fully-Informed</td><td>Knows correct answers and reveals their knowledge</td><td>Not possible</td><td>Knows answers and conceals his/her knowledge</td></tr>
<tr><td>Partially-informed</td><td colspan="2">Does not know answer, but highly likely to choose the correct answer</td><td rowspan="3">Does not know answer and conceals knowledge</td></tr>
<tr><td>Randomizing</td><td colspan="2">Does not know answer and guesses answer randomly</td></tr>
<tr><td>Partially-misinformed</td><td colspan="2">Does not know answer, but highly possible to choose the incorrect answer</td></tr>
<tr><td>Misinformed</td><td>Not possible</td><td>Has wrong belief about correct answer</td><td>Has wrong belief about correct answer and conceals his/her knowledge</td></tr>
</table>

I modified the table described in the original paper. I changed some terms and reclassified the knowledge state based on the likelihood of choosing the correct answers

Table 2.2: True parameters and parameter estimation by Maximum Likelihood Estimation

True Parameters		**Estimation**			
		Model I		**Model II**	
Parameter	**Value**	Est.	S.E.	Est.	S.E.
β_0	1.200	1.086	0.065	1.115	0.068
β_1	2.100	2.120	0.057	2.148	0.049
β_2	0.800	0.836	0.023	0.831	0.021
β_3	-1.500	-1.548	0.059	-1.504	0.060

Table 2.3: Distribution of respondents' answers

Type	**Number of Correct Answers**								
Number of Incorrect Answers	#	0	1	2	3	4	5	6	Total
	0	232	175	182	184	281	270	242	1566
	1	84	159	354	417	369	408		1791
	2	46	168	339	357	450			1360
	3	33	132	229	348				742
	4	22	86	268					376
	5	17	89						106
	6	10							10
	Total	444	809	1372	1306	1100	678	242	5951

Table 2.4: Descriptive statistics by the number of correct answers

Variable	All	Lowest	Lower	Median	Upper	Highest
		(0-1)	(2)	(3)	(4)	(5-6)
Socio-demographics	0.619	0.551	0.589	0.624	0.644	0.721
Age: 25-34 years	0.590	0.624	0.634	0.612	0.574	0.464
Female	0.407	0.504	0.430	0.419	0.343	0.303
Non-white	0.439	0.482	0.474	0.457	0.409	0.341
Live with partners	0.503	0.460	0.515	0.534	0.515	0.484
Income: below $25K						
Self-assessed skill						
High math skill	0.346	0.225	0.317	0.342	0.400	0.495
High knowledge	0.360	0.326	0.361	0.349	0.372	0.408
General education						
Associate degree	0.352	0.215	0.278	0.327	0.433	0.589
College degree	0.509	0.421	0.510	0.529	0.542	0.563
Financial education						
(Informal education)						
From parents/guardians	0.513	0.444	0.494	0.529	0.554	0.566
(Formal education)						
Rejected offers	0.191	0.169	0.222	0.205	0.183	0.165
High school	0.131	0.090	0.115	0.129	0.147	0.191
College	0.131	0.104	0.106	0.123	0.130	0.215

Table 2.5: Descriptive statistics by the number of incorrect answers

Variable	All	Lowest	Lower	Median	Upper	Highest
		(0-1)	(2)	(3)	(4)	(5-6)
Socio-demographics						
Age: 25-34 years	0.619	0.643	0.612	0.599	0.617	0.626
Female	0.590	0.615	0.611	0.588	0.569	0.467
Non-white	0.407	0.356	0.390	0.427	0.476	0.476
Live with partners	0.509	0.492	0.507	0.510	0.535	0.533
Income: below $25K	0.503	0.496	0.500	0.518	0.513	0.478
Self-assessed skill						
High math skill	0.346	0.325	0.361	0.337	0.357	0.366
High knowledge	0.360	0.264	0.301	0.365	0.497	0.663
General education						
Associate degree	0.439	0.411	0.448	0.469	0.442	0.413
College degree	0.352	0.411	0.352	0.321	0.303	0.325
Financial Education						
(Informal education)						
From parents/guardians	0.513	0.508	0.500	0.509	0.543	0.549
(Formal education)						
Rejected offers	0.191	0.151	0.156	0.190	0.245	0.370
High school	0.131	0.133	0.140	0.131	0.117	0.112
College	0.131	0.131	0.126	0.135	0.147	0.108

Table 2.6: Estimates and standard errors by Maximum Likelihood Estimation

Variable	Ref. Model		Model I		Model II	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Socio-demographics						
Age: 25-34 years	0.043**	0.015	0.025	0.017	0.017	0.023
Female	-0.140**	0.014	-0.084**	0.015	-0.028	0.020
Non-white	-0.162**	0.014	-0.136**	0.015	-0.169**	0.020
Income: below $25K	-0.105**	0.016	-0.058**	0.018	-0.032	0.025
Self-assessed skill						
High math skill	0.223**	0.015	0.164**	0.016	0.163**	0.021
High knowledge	-0.044**	0.015	-0.136**	0.016	-0.320**	0.021
General education						
Associate or lower	0.163**	0.018	0.144**	0.021	0.203**	0.029
College or higher	0.409**	0.020	0.349**	0.023	0.419**	0.031
Financial Education						
(Informal education)						
From parents/guardians	0.064**	0.014	0.038*	0.016	0.028	0.021
(Formal education)						
Rejected offers	-0.036	0.018	-0.110**	0.020	-0.235**	0.025
High school	0.186**	0.021	0.131**	0.024	0.112**	0.031
College	0.030	0.022	0.000	0.024	-0.062*	0.030

*p<0.05, **p<0.01

Table 2.7: Marginal effects on the success probability by Maximum Likelihood Estimation

Variable	Ref. Model		Model I		Model II	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Socio-demographics						
Age: 25-34 years	0.017**	0.006	0.010	0.007	0.006	0.008
Female	-0.056**	0.006	-0.033**	0.006	-0.010	0.007
Non-white	-0.064**	0.005	-0.053**	0.006	-0.061**	0.007
Income: below $25K	-0.042**	0.006	-0.023**	0.007	-0.011	0.009
Self-assessed skill						
High math skill	0.089**	0.006	0.064**	0.006	0.058**	0.007
High knowledge	-0.017**	0.006	-0.053**	0.006	-0.117**	0.008
General education						
Associate or lower	0.063**	0.007	0.057**	0.008	0.077**	0.011
College or higher	0.162**	0.008	0.136**	0.009	0.152**	0.011
Financial Education						
(Informal education)						
From parents/guardians	0.025**	0.005	0.015*	0.006	0.010	0.007
(Formal education)						
Rejected offers	-0.014	0.007	-0.044**	0.008	-0.087**	0.010
High school	0.074**	0.009	0.050**	0.009	0.038**	0.010
College	0.012	0.009	0.000	0.009	-0.022*	0.011

*p<0.05, **p<0.01

Appendix A: Bayesian Estimation

I use a Bayesian estimation method as well as the MLE. I choose the usual multivariate normal distribution as a prior for $\boldsymbol{\beta}$ as follows:

$$p(\boldsymbol{\beta}) = (2\pi)^{-\frac{k}{2}}|\boldsymbol{B}_0|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{B}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\} \tag{2.10}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{B}_0$ are the parameters of a multivariate normal density. Also, k denotes the dimension of $\boldsymbol{\beta}$. Then, I combine this prior and likelihood shown in (2.6), and obtain the posterior kernel for $\boldsymbol{\beta}$ as follows:

$$\begin{aligned} p(\boldsymbol{\beta}|\boldsymbol{y}^{obs},\boldsymbol{X}) \propto & \mathcal{L}(\boldsymbol{\beta}|\boldsymbol{y}^{obs})p(\boldsymbol{\beta}) \\ \propto & (2\pi)^{-\frac{k}{2}}|\boldsymbol{B}_0|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{B}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\} \\ & \times\prod_{i=1}^{n}\sum_{k=0}^{Q_i^{DK}}\left[\begin{array}{l}\binom{Q_i}{y_i^{obs}+k}\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})^{y_i^{obs}+k}\Phi(-\boldsymbol{x}_i'\boldsymbol{\beta})^{Q_i-y_i^{obs}-k} \\ \cdot\binom{Q_i^{DK}}{k}(p_i^{DK})^k(1-p_i^{DK})^{Q_i^{DK}-k}\end{array}\right] \end{aligned} \tag{2.11}$$

Because this kernel is not a well-known density, I cannot obtain a convenient method to draw posteriors from it, such as a Gibbs sampler. I apply the Metropolis-Hastings (MH) algorithm in which I obtain a candidate vector $\boldsymbol{\beta}_c$ from a candidate generating function $q(\boldsymbol{\beta}_c)$. I retain this candidate with a high acceptance probability $p_a(\boldsymbol{\beta}_o \to \boldsymbol{\beta}_c)$ where $\boldsymbol{\beta}_o$ is a current value of the parameter. The acceptance probability takes the form of:

$$p_a(\boldsymbol{\beta}_o \to \boldsymbol{\beta}_c) = \min\left\{\frac{p(\boldsymbol{\beta}_c|\boldsymbol{y}^{obs},\boldsymbol{X})q(\boldsymbol{\beta}_o|\boldsymbol{\beta}_c)}{p(\boldsymbol{\beta}_o|\boldsymbol{y}^{obs},\boldsymbol{X})q(\boldsymbol{\beta}_c|\boldsymbol{\beta}_o)},1\right\} \tag{2.12}$$

In this study, I choose a multivariate normal candidate generating distribution with mean equal to the current value $\boldsymbol{\beta}_0$ and the same variance matrix. Thus, I have $q(\boldsymbol{\beta}_o|\boldsymbol{\beta}_c) = q(\boldsymbol{\beta}_c|\boldsymbol{\beta}_o)$ from a symmetric characteristics of the normal density. The final acceptance probability to

obtain the posterior draws for $\boldsymbol{\beta}$ has the form of:

$$p_a(\boldsymbol{\beta}_o \to \boldsymbol{\beta}_c) = \min\left\{\frac{p(\boldsymbol{\beta}_c|\boldsymbol{y}^{obs}, \boldsymbol{X})}{p(\boldsymbol{\beta}_o|\boldsymbol{y}^{obs}, \boldsymbol{X})}, 1\right\} \tag{2.13}$$

I also estimate the posterior densities and marginal effects of $\boldsymbol{\beta}$ by using the draws from the MH procedure.

I carry out a simulation to examine the performance of the suggested model with MH as well as MLE. As Table 2.8 shows, the suggested model with MH also identifies the true parameters successfully.

[Table 2.8 about here]

Table 2.9 shows the posterior means and standard deviations of the coefficients in which I am interested in estimated by the MH method. Table 2.10 also shows the corresponding marginal effects on the probability of success from the MH algorithm.

[Table 2.9 to 2.10 about here]

For the estimation of MH method, all posteriors are obtained from $200,000$ replications; $100,000$ burn-in replications were discarded and $100,000$ were retained. In this case, I do not use a typical significance measure, such as the standard error or p-value. Thus, I use an alternative measure that represents the probability that the mass of the posteriors drawn from the samplers is placed over the positive value (Koop, Poirier, and Tobias 2007). If this measure approaches zero, I suggest that the corresponding posterior mean was negatively significant. On the other hand, if the measure is close to one, I suggest that the posterior mean is positively significant. Based on this concept, I determine that it was "highly significant" if the measure is greater than 0.975 or less than 0.025. Also, I determine that it is "marginally significant" if it is greater than 0.950 or less than 0.050. I observe that the

estimation results, marginal effects and the corresponding significance of parameters from the MH are almost consistent with those from the MLE.

Appendix B: Tables

Table 2.8: True parameters and estimations by Metropolis-Hastings method

True Parameters		**Estimation**			
		Model I		**Model II**	
Parameter	**Value**	Mean	Std.	Mean	Std.
β_0	1.200	1.089	0.067	1.109	0.068
β_1	2.100	2.123	0.058	2.154	0.051
β_2	0.800	0.836	0.024	0.837	0.021
β_3	-1.500	-1.549	0.059	-1.544	0.059

Table 2.9: Posterior means and standard deviations by Metropolis-Hastings

Variable	Ref. Model		Model I		Model II	
	Mean	Std	Mean	Std	Mean	Std
Socio-demographics						
Age: 25-34 years	0.043	0.016	0.023	0.018	0.015	0.023
Female	-0.139**	0.013	-0.084**	0.016	-0.028	0.019
Non-white	-0.161**	0.013	-0.135**	0.016	-0.171**	0.021
Income: below $25K	-0.105**	0.017	-0.058**	0.018	-0.031	0.025
Self-assessed skill						
High math skill	0.223**	0.015	0.163**	0.016	0.162**	0.021
High knowledge	-0.044**	0.016	-0.136**	0.017	-0.319**	0.022
General education						
Associate or lower	0.160**	0.019	0.142**	0.021	0.199**	0.03
College or higher	0.408**	0.020	0.349**	0.024	0.416**	0.031
Financial Education						
(Informal education)						
From parents/guardians	0.063**	0.014	0.038**	0.016	0.029	0.02
(Formal education)						
Rejected offers	-0.037*	0.019	-0.113**	0.02	-0.236**	0.025
High school	0.186**	0.021	0.132**	0.023	0.113**	0.03
College	0.030	0.022	-0.001	0.023	-0.062**	0.029

** and * denote high and marginal significance representing that the 97.5% and 95.0% of the mass of the posterior were placed over the positive or negative values, respectively.

Table 2.10: Marginal effects on the probability of success by Metropolis-Hastings

Variable	Ref. Model		Model I		Model II	
	Mean	Std	Mean	Std	Mean	Std
Socio-demographics						
Age: 25-34 years	0.017**	0.006	0.009	0.007	0.006	0.008
Female	-0.055**	0.005	-0.033**	0.006	-0.010	0.007
Non-white	-0.064**	0.005	-0.053**	0.006	-0.062**	0.008
Income: below $25K	-0.042**	0.007	-0.023**	0.007	-0.011	0.009
Self-assessed skill						
High math skill	0.089**	0.006	0.063**	0.006	0.058**	0.007
High knowledge	-0.018**	0.006	-0.053**	0.006	-0.117**	0.008
General education						
Associate or lower	0.062**	0.007	0.056**	0.008	0.076**	0.011
College or higher	0.161**	0.008	0.136**	0.009	0.151**	0.012
Financial Education						
(Informal education)						
From parents/guardians	0.025**	0.006	0.015**	0.006	0.011	0.007
(Formal education)						
Rejected offers	-0.015*	0.007	-0.044**	0.008	-0.088**	0.009
High school	0.074**	0.008	0.051**	0.009	0.039**	0.010
College	0.012	0.009	-0.001	0.009	-0.022**	0.010

** and * denote high and marginal significance representing that the 97.5% and 95.0% of the mass of the posterior were placed over the positive or negative values, respectively.

Chapter 3

Health Interventions and Misuse of Cervical Cancer Screening among Young Women: Multinomial Regression Approach

3.1 Introduction

Cervical cancer is the second leading cause of cancer death in women aged 20 to 39 years in the United States (American Cancer Society 2017a; Siegel, Miller, and Jemal 2017). In 2017, the American Cancer Society published an annual report that provided the estimated numbers of new cancer cases and deaths (Siegel, Miller, and Jemal 2017). In this report, they estimated that 12,820 cases of cervical cancer would be newly diagnosed and 4,210 women would die from cervical cancer in the United States in 2018.

Cervical cancer is an abnormal growth of the cells on the cervix. There are several risk factors for cervical cancer, such as human papillomaviruses (HPV) infection, smoking, age, family history of cervical cancer, and so forth. Among them, infection by human papillomaviruses (HPV) is the most common risk factor for cervical cancer (American Cancer Society 2017b). Despite no cure for HPV infection, we can treat the abnormal cell growth by cervical cancer preventions including Papanicolaou (Pap) screening and the corresponding

early detection of cervical cancer (Saslow et al. 2002). When detected at an early stage, the 5-year survival rate for women with invasive cervical cancer is 91% compared with 68% of average survival rate (Siegel, Miller, and Jemal 2017). As a result, over the last 40 years, the death rate from this disease has plunged by more than 50%.

For this screening, the U.S. Preventive Services Task Force (USPSTF) suggests several recommendations on cervical cancer screening for average-risk women in 2012[1] (Moyer 2012; Centers for Disease Control and Prevention n.d.). First, the USPSTF recommends that women aged 21–65 receive Pap smear testing every three years, or women aged 30–65 should receive an HPV co-test every five years, with either an HPV co-test or a Pap test every 3 years, regardless of the age of onset of sexual activity or other risk factors. The USPSTF also recommends against screening for cervical cancer with HPV testing in women younger than age 30 years. In addition, the USPSTF recommends against screening for cervical cancer in women younger than age 21 years. Table 3.1 shows recommendations summary suggested by the USPSTF.

[Table 3.1 about here]

Despite this guideline, many women underuse or misuse Pap smear tests (Watson et al. 2017; White 2017). Several studies argued that the cause of this is insufficient Pap smear knowledge among high-risk populations and some socioeconomic/demographic characteristics (Chen et al. 2012; Daley et al. 2013). The 2015 National Health Interview Survey (NHIS) also reported that the main reason women skip Pap smear screening is a lack of recommendations from a doctor or a lack of relevant information about HPV (CDC 2015).

[1]The USPSTF's guideline is for average-risk women but does not explain who is average-risk. Sawaya et al. (2015) suggested that these populations include women with no history of a precancerous lesion (cervical intraepithelial neoplasia [CIN] grade 2 or a more severe lesion) or cervical cancer, those who are not immunocompromised (including being HIV-infected), and those without in utero exposure to diethylstilbestrol.

Another concern is too-frequent Pap testing for average-risk women. Experts argued that annual Pap test can harm the recipients' health. A false positive Pap test result may lead to more-frequent colposcopies, a follow-up procedure which may damage the cervix. In addition, annual screening does not significantly increase the detection of cervical cancer comparing to screening every three years because the cancer is relatively slow to develop (Moyer 2012). In addition, women who completed HPV vaccination prior to first intercourse require less frequent screenings (Massad et al. 2009). A study conducted in 2004 found the majority of American women were screened for cervical cancer by the Pap smear test more frequently than recommended; 55% of women with no history of abnormal smears nevertheless underwent annual Pap smears (Sirovich and Welch 2004). Similarly, the NHIS conducted in 2015 showed that 35.53% women sixty-five or younger had received annual Pap smear screening (CDC 2015).

Doctor's or health professional's intervention is one of the important factors for the screening decisions. Several studies showed that quality interaction between health professionals and patients is a critical factor in increasing the awareness of the value of preventive cancer screening and increasing the rates at which women complied with screening guidelines (Fisher 1984; Schofield et al. 1994; Taylor et al. 2003; Ferrante et al. 2007). Some studies found, though, that physicians still recommend a shorter-than-recommended interval for Pap smear tests (Meissner et al. 2010).

HPV vaccination is another factor affecting decisions about Pap smear screening. Researchers have investigated the potential ex-ante moral hazard induced by HPV vaccination and the false perception that it is fully effective. Whether or not HPV vaccination induces moral hazard is controversial. Some studies reported that HPV vaccination would make women reluctant to take Pap screening and delayed screening intervals (Ferris et al. 2012; Budd et al. 2014). However, another study found no empirical evidence of ex-ante moral hazard from HPV vaccination (Moghtaderi and Dor 2016).

I argue that HPV vaccination increases women's awareness of the value of preventive care for cervical cancer. The vaccination indirectly provides relevant information about HPV and screening care. Their awareness around HPV, are more likely to remain up-to-date about guidelines around Pap screening and the vaccinated women are more willing to undergo Pap testing (Price et al. 2011). However, further studies are needed to investigate this effect.

The purpose of this study is to investigate which health-related interventions significantly affect misuse of cervical cancer screening and the Pap smear screening decision among women younger than 30. I consider three types of misuse of screening: (1) Pap smear screening for women aged under 21 (too-early screening), (2) HPV testing for women aged under 30 (unnecessary HPV test), and (3) annual Pap smear screening (overuse of screening), based on the USPSTF's recommendations for cervical cancer screening. I empirically evaluate the effect of several health-related interventions on the misuse of cervical cancer screening by the binary and multinomial logistic regression model. I also discuss which interventions are the most effective to increase the probability of recommended use of cervical cancer screening[2].

3.2 Literature Review

Frequent screening for cervical cancer increases healthcare costs. Several studies evaluated the financial burden of annual and triennial screening strategies. Sawaya et al. (2015) summarize the incremental cost-effectiveness ratio (ICER)[3] of screening for cervical cancer in terms of quality-adjusted-life-year (QALY)[4] by screening intervals. The ICER of the

[2]The decisions for cervical cancer screening is made by women. Doctors or health professionals provide the information of cervical cancer and a proper screening method only.

[3]The incremental cost-effectiveness ratio (ICER) is a monetary measure to evaluate the cost-effectiveness of an intervention relevant to the healthcare analysis. The ratio of the difference in two healthcare costs to the difference in the corresponding effects. If the ICER of an intervention above the other intervention, the former is less cost-effective than the latter. The cost is usually measured by quality-adjusted life year (QALY).

[4]The quality-adjusted life year (QALY) is a monetary value of healthcare interventions and the corresponding quality of life lived. One QALY represents one year in perfect health status that a person

annual screening for cervical cancer would be over $500 000 per QALY gained. However, the ICER of the screening every 3 to 5 years would be less than $100 000 per QALY gained. It shows that the annual screening is not cost-effective comparing other longer interval screening methods. Berger, Schroy, and Dinh (2016) show the similar result by performing another test for cervical cancer, Multitarget stool DNA (mt-sDNA) test. They estimate the ICER and found that the annual, 3-year, and 5-year mt-sDNA screening costs would be $20,178, $11,313, and $7388 per QALY, respectively, compared with no screening. Akker-van Marle et al. (2002) also show that the average ICER of common cancer screening tests every 6 years would be $15,500 per QALY gained while $23,900 per QALY gained when annual cancer screening, comparing to a no screening strategy. Thus, they conclude that frequent screening increases the costs of screening and follow-up treatments but small gains in avoiding cancer cases.

Another financial burden of the misuse of the screening for cervical cancer is the misuse of health providers and the corresponding patient time. Yarnall et al. (2003) evaluate the amount of time required for a primary care physician to provide recommended preventive services. From their evaluation, they conclude that 1,773 hours of a physician's annual time, or 7.4 hours per working day, is needed for the provision of preventive services to fully satisfy the USPSTF recommendations. Thus, there would be a huge opportunity cost of physicians' compliance with preventive services recommendations.

The efficiency of annual Pap smear screening is uncertain. Habbema et al. (2012) compare cervical cancer screening intensity and cervical cancer mortality trends in the United States (annual screening) and the Netherlands (triennial screening). This study shows that the cervical cancer screening system in the Netherlands would be as effective as the U.S. system. Thus, annual Pap smear screening would not effective but highly costly comparing to triennial screening strategy based on this study.

obtains by accepting healthcare interventions.

Frequent cancer screening could induce psychological problem, in particular, when women have positive test result. Ideström, Milsom, and Andersson-Ellström (2003) show that about 30% of female respondents felt stressed and anxious between being informed of the result of the first Pap smear test and subsequent further investigation. This finding is correlated with less satisfaction with follow-up and a negative influence on self-esteem. The anxiety from positive test result lasts for twelve weeks. Drolet et al. (2015) conduct a survey in Canada and find that 35% of women have clinically meaningful anxiety at 12 weeks and the corresponding QALY lost following an abnormal result were between 0.007 and 0.009. Also, the following treatment including colposcopy after the positive test from a Pap smear decreases sexual activity. Gath et al. (1995) conduct a survey in the UK and find that 43 % of the study cohort have a decreased frequency of sexual intercourse. Ideström, Milsom, and Andersson-Ellström (2003) show a similar result that 8% of female participants in their survey report a remaining negative influence on sexuality and their experience of sexual intercourse.

The harm to women's health is another problem in frequent screening. Health professionals do not recommend frequent screening, in particular, young women, because of the false-positive test result. Although the abnormalities of a Pap smear test in the young women's cervix could go away if left alone, doctors try to remove them. This procedure they used can injure the cervix and induces problem later when they become pregnant including premature birth (Grady 2009). Also, a cohort study shows the existence of self-reported side-effect after a cervical screening. Korfage et al. (2012) report that 8 and 5% of women experience lower abdominal pain, vaginal bleeding, discharge, or urinary problems for 2-3 and 4-7 days, respectively, following the Pap smear.

Unnecessary use of HPV test also induces problem in increasing medical cost in society. Lee, Berkowitz, and Saraiya (2011) argue that healthcare providers reported inappropriate uses of HPV testing, and it may lead to unnecessary follow-up. They examine the 2006

National Ambulatory Medical Care Survey and National Hospital Ambulatory Medical Care Survey and find that the non-recommended HPV test for women younger than 30 increased medical costs without added benefits. Liverani (2015) also argue that inappropriate testing for HPV types on healthy subjects increases costs without benefit and potentially results in overtreatment.

Several empirical studies have explored the factors that affect the misuse of the Pap smear screening. Gerend et al. (2017) find from an online survey in 2014 that hesitancy about the longer screening interval would stem from concern about developing cancer between screenings. Although over two-thirds are willing to undergo less frequent screening, 65% of respondents express discomfort or unsure with the longer screening interval. Also, Almeida et al. (2013) find that the overuse of screening was associated with younger age, more medical visits, contraceptive management visits, and gynecology provider specialty while the underuse is associated with older age, fewer medical visits, and increased comorbidity.

The existing studies also explain that health professional's recommendations would be effective to improve both underuse and overuse of cervical cancer screening including Pap smear. Coughlin et al. (2005) investigate the role of physician recommendations in underuse of Pap screening. They examine the NHIS conducted in and find that lack of a physician recommendation contributes to underuse of Pap screening by many eligible women. In contrast, Meissner et al. (2010) investigate the role of physician recommendations in overuse of Pap screening. They examine the National Survey of Primary Care Physicians and Health Information Trends Survey and evaluate women's willingness to follow a 3-year Pap test interval. From the result, they find that women would like to accept a 3-year interval for Pap tests, although most primary care physicians want to continue to recommend shorter intervals.

3.3 Methods

3.3.1 Data and Measures

I use the NHIS conducted in 2015 (CDC 2015). The main purpose of the NHIS is to monitor the health of the U.S. population. It includes a broad range of health topics, as well as many demographic and socioeconomic characteristics.

I examine two relevant modules: the Sample Adult (SA) module, and the Sample Adult Cancer Control (SACC) module. The SA module investigates health-related issues including physical health condition and cognition. The SACC module, which assesses respondents' knowledge and attitudes related to cancer, cancer-related health behaviors, and cancer screening and risk assessment, is administered every five years, most recently in 2015.

I consider three types of dependent variables relating to the misuse of cervical cancer screenings based on the USPSTF guideline. First, as the USPSTF recommends against screening for cervical cancer in women younger than age 21 years, I consider a variable that represented whether or not women young than age 21 years had a Pap smear (called Model I). This variable is respondents' response to the question, "Have you EVER HAD a Pap smear or Pap test?" I dichotomize as 1 if they respond "Yes," representing too-early screening for young women aged 18-20. I also dichotomize as 0 if they do "No," representing compliance with the recommendation for the women.

Second, as the USPSTF recommends against screening for cervical cancer with HPV testing, alone or in combination with cytology, in women younger than age 30 years, I consider a variable that represents whether women younger than age 30 years received HPV test along with a Pap smear or a Pap smear alone (called Model II). For this variable I combine the dependent variable in Model I with respondents' response to the question, "An HPV test is sometimes given with the Pap test for cervical cancer screening. Did you have an

HPV test with your most recent Pap?" I dichotomize as 2 if they respond "Yes," representing unnecessary HPV test for young women aged 21-29. I dichotomize as 1 if they do "No," representing compliance with the recommendation for the women. I also dichotomize as 0 if they have never received a Pap test for the women.

Lastly, as the USPSTF recommends screening for cervical cancer in women age 21 to 65 years with cytology (Pap smear) every 3 years, I consider a variable that represented whether or not women younger than age 30 years have a Pap smear annually (called Model III). For this variable, I combine this variable in Model I with respondents' response to the question, "How many Pap tests have you had in the LAST 6 YEARS?" I dichotomize as 2 if they respond 6 in last six years, representing overuse of screening (annual Pap). I dichotomize as 1 if they do less than 6 in last six years, representing compliance with the recommendation. I also dichotomize as 0 if they have never received a Pap test. Table 3.2 shows the summary of dependent variables by models.

[Table 3.2 about here]

I consider three key health-related interventions. The first intervention is a doctor's recommendation for the most recent Pap smear screening. This intervention is respondents' response to the question, "Was your most recent Pap test recommended by a doctor or other health professional?" I dichotomize as 1 if they respond "Yes," 2 if they do "Did not see a doctor in the last 12 months," and 0 if they did "No." The second intervention is whether or not respondents have heard of HPV before, and this variable is the respondents' response to the questions, "Have you ever heard of HPV? HPV stands for human papillomavirus." I dichotomize as 1 if they respond yes, 0 if not. The last intervention is an indicator of HPV vaccination, the respondents' response to the question, "Have you ever received an HPV shot or vaccine?" I dichotomize as 1 if they respond yes, 0 if not.

I also consider two other health-related variables. The first variable is an indicator of the

flu shot, the respondents' response to the question, "Have you had a flu shot? A flu shot is usually given in the fall and protects against influenza for the flu season." I dichotomize as 1 if they respond yes, 0 if not. The other variable is an indicator of staying in bed due to illness during the past twelve months. This variable is the respondents' response to the question, "During the past 12 months, about how many days did illness or injury keep you in bed more than half of the day?" I dichotomize as 1 if they keep in bed at least half of the day, 0 if not.

Demographic factors including age, race (white or non-white), marital status (married or single), and employment status (employed or not) are also included in this model.

I consider a subsample for each dependent variable. For Model I, I consider the portion of the female sample aged 18–20 only, consisting of 525 respondents. I then remove 29 missing and "Don't Know/Refused to Answer" (DK/RF) observations in the dependent variable and 35 missing and DK/RF observations (31 from HPV vaccination and 4 from others) in independent variables, resulting in a final subsample for Model I of 461. For Model II, I consider the portion of the female sample aged 21-29 only, and the number of the subsample is reduced to 2,474. I also drop 418 DK/RF observations in the dependent variable (253 from HPV test and 165 from Pap smear) and 78 missing and DK/RF observations (70 from HPV vaccination and 8 from others) in the independent variable, and the final number of the subsample for Model II was 1,978. For Model III, I also consider the portion of the female sample aged 21-29 only, and the number of the subsample is reduced to 2,309 excluding DK/RF responses in the Pap smear indicator. The dependent variable for this model is the respondents' responses to the frequency question, "How many Pap tests have you had in the last 6 years?" I drop 172 observations that have more than six Pap test in the last 6 years, as these are abnormal frequencies for average-risk women. I also consider the average-risk women based on the USPSTF recommendations. For this purpose, I exclude the high-risk populations by using the answer to the question, "Have you had a [fill1: Pap/Pap or HPV]

test in the LAST 3 YEARS where the results were NOT normal?" I discard 274 observations of "Yes" or DK/RF answer and the number of the subsample is reduced to 1,863 [5]. I also drop 87 missing and DK/RF observations in independent variables (76 from HPV vaccination and 11 from others) and finally have 1,776 for Model III. Table 3.3, 3.4, and 3.5 show the descriptive statistics of the variables used in this study by dependent variables.

[Table 3.3, 3.4, and 3.5 about here]

3.3.2 Econometric Model

Let U_{ij} and p_{ij} represent the latent utility of choosing alternative $j \in \{1, \cdots, J\}$ to agent $i \in \{1, \cdots, n\}$ and the corresponding probability, respectively. Then, the latent utility is assumed to be:

$$U_{ij} = \boldsymbol{x}_i' \boldsymbol{\beta}_j + \varepsilon_{ij} \tag{3.1}$$

where x_i is a vector of covariates for agent i, and ε_{ij} is a error term for agent i's choice of alternative j. Also, the corresponding log-odds and probability are written by:

$$\log \frac{p_{ij}}{p_{iJ}} = \boldsymbol{x}_i' \boldsymbol{\beta}_j + \varepsilon_{ij} \tag{3.2}$$

and

$$p_{ij} = \frac{\exp(\boldsymbol{x}_i' \boldsymbol{\beta}_j)}{\sum_{m=1}^{J} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_m)} \tag{3.3}$$

[5] The National Cancer Institute recommends more frequent screening for women who are HIV positive, have a weakened immune system, were exposed before birth to a medicine called diethylstilbestrol (DES), which was once prescribed to pregnant women, had a recent abnormal Pap test or biopsy result, have had cervical cancer (National Cancer Institute 2017). This institute also recommend that women who have abnormal screening results may need to have a follow-up Pap test in 6 months or a year. Based on these recommendations, an annual Pap smear would be required for women who have experienced abnormal test results from cervical cancer screening methods recently (National Cancer Institute 2016). In addition, Sirovich and Welch (2004) distinguish between samples with and without a history of abnormal Pap smears in their study of Pap smear screening intervals. Thus, it is reasonable to regard women who have experienced any abnormal test result from screening methods as a high-risk population for this study.

As an unattainable part of this situation is to identify the order of preferences, I choose one category as the reference choice and set its coefficient to zero. Thus, the probability specified in (3.3) is rewritten by:

$$p_{ij} = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{m=2}^{J} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_m)}, \ j \in \{1, \cdots, J-1\} \tag{3.4}$$

The odds ratio represents the odds of the outcome occurring with a predictor compared to the odds of the outcome occurring without a predictor. Mathematically, the odds ratio of a covariate $x_p \in \{0, 1\}$ is written by:

$$OR_{x_p} = \frac{p_{j|x_p=1}}{1 - p_{j|x_p=1}} \tag{3.5}$$

Based on the definition of the odds ratio, I can interpret the estimation result of the odds ratio of a covariate; if the odds ratio is greater than one, it indicates that the preference is less likely to be in the target preference as a predictor increases than in the reference preference.

The relative risk represents the ratio of the probability of choosing one outcome over that of choosing the baseline outcome. Thus, the relative risk ratio represents how the risk of the outcome falling in the baseline compared to the risk of the outcome falling in the target changes when a variable changes. It yields the change in the relative risk for a unit change in the independent variable. Mathematically, the relative risk ratio of a covariate $x_p = \{0, 1\}$ is written by:

$$RRR_{x_p} = \frac{p_{j|x_p=1}/p_{1|x_p=1}}{p_{j|x_p=0}/p_{1|x_p=0}} \tag{3.6}$$

Based on the definition of the relative risk ratio, I can interpret the estimation result of the relative risk ratio of a covariate; if the relative risk ratio is greater than one, it indicates that the preference is less likely to be in the target preference as a predictor increases than in the

reference preference.

I estimate these specified models by employing the binomial and multinomial logistic regressions. I obtain the coefficient estimates and the corresponding odds ratio/relative risk ratio by the maximum likelihood estimation (StataCorp 2013).

3.4 Estimation Results

Table 3.6 and 3.7 show the estimation results including the estimates and the corresponding odds ratios/relative risk ratios of the predictors from the regression model, respectively. Among these results, I focus on the odds ratios/relative risk ratios of health-related interventions in this section.

[Table 3.6 and 3.7 about here]

First, the doctor's recommendation of the Pap smear screening is a significant factor in encouraging women to receive Pap smears at the proper usage. The relative risk ratios from Model II shows that the relative probability of receiving a Pap smear only would be 8.20 (=1/0.122) times higher than that of receiving no Pap smear test for young women aged 21-30 if their doctor or other health professional recommended the screening. Similarly, the relative risk ratios from Model III indicate that the relative probability of a proper frequency of the screening would be 5.47 (=1/0.183) times higher than that of receiving no Pap test within six years for them if their doctors or other health professionals recommend the screening.

However, the doctor's recommendation is also a significant factor in inducing the misuse of the screening including HPV testing along with a Pap smear and the annual Pap smear screening. The odds ratio from Model I shows that the odds of receiving the too-early Pap smear are about 37 times higher than that of no too-early Pap smear for young women

aged 18-20 if their doctors or other health professionals recommend the screening. Also, the relative risk ratios from Model II show that the relative probability of preferring HPV testing along with a Pap smear to a Pap smear alone would increase by a factor of 1.35 for young women aged 21-29 if their doctors or other health professionals recommend the screening. Further, the relative risk ratios from Model III indicate that the relative probability of the annual Pap smear that is not recommended for young women aged 21-29 over a proper frequency of the screening would increase by a factor of 1.35 for them if their doctors or other health professionals recommend the screening.

Second, hearing of HPV is a significant factor in encouraging respondents to receive Pap smears at the proper frequency of the screening, but not other compliance with the recommendation. The relative risk ratios from Model III show that the relative probability of the proper frequency of the screening would be 2.28 (=1/0.439) times higher than that of receiving no Pap smear test for young women aged 21-30 if they have ever heard of HPV. However, it is not a significant factor in compliance with Pap only for young women aged 21-29 and no screening for young women aged 18-20. Similarly, this intervention does not influence all types of the misuse of cervical cancer screenings considered in this study. It significantly affects the decision to receive HPV testing along with a Pap smear and overuse of annual Pap smear, but not too-early screening for young women. The relative risk ratios from Model II show that the relative probability of receiving HPV testing along with a Pap smear would be 5.62 times higher than that of receiving a Pap smear alone for young women aged 21-30 if their doctors or other health professionals recommend the screening. Further, the relative risk ratios from Model III show that the relative probability of receiving annual Pap smear screening would be 1.63 times higher than that of a proper interval of screening for the young women. However, it is not a significant factor in inducing the too-early screening and the overuse of the screenings.

Moreover, HPV vaccination is a significant factor in encouraging women to receive Pap

smears at the proper usage. From Model II, I observe that the relative probability of receiving a Pap smear only would be 1.96 (=1/0.511) higher than that of receiving no Pap smear test for young women aged 21-30 if they have received HPV vaccination. Similarly, I observe from Model III that the relative probability of a proper frequency of the screening would be 2.18 (=1/0.458) times higher than that of receiving no Pap test within six years for them. However, it induces the unnecessary use of HPV for young women as the relative probability of receiving HPV test would be 1.73 times higher than that of a Pap smear only for the women.

3.5 Discussion

Based on these results, I discuss the effects of doctor recommendations, information about HPV, and HPV vaccination, on the misuse of cervical cancer screening for young women including too-early Pap smear, HPV test, and annual Pap smear that are not recommended.

First, my empirical analysis describes that doctor's recommendation plays a significant role in encouraging women to receive a Pap smear in both proper and improper ways. If a woman has heard a recommendation of having a Pap smear, she is more willing to receive it, even misuse it. Based on this result, I discuss the role of doctor recommendation. Doctors and other health professionals could provide correct information and appropriate expertise on screening method for cervical cancer screening. In this case, women lean on the information and are more likely to receive the screening including a Pap smear than women who do not have their recommendation are. However, doctors and other health professionals fail to follow the guideline of the expert groups that recommended when to start screening, annual screening, and HPV testing guideline for young women aged 30 or less. In this case, young women are more likely to misuse the screening methods for cervical cancer, and it induces financial and psychological problem individually and in society.

Moreover, HPV vaccination helps women who have never taken a Pap smear to receive it. The result shows that HPV vaccination would decrease the probability of receiving no Pap smear and increase the probability of receiving a Pap smear although it could induce young women's unnecessary HPV testing.

The effect of HPV vaccination on the prevention behaviors of cervical cancer is controversial. Sopracordevole et al. (2013) analyze students in a secondary school in a northeastern Italian city and find that HPV vaccinated teenaged girls do not have more knowledge of HPV and prevention methods than HPV unvaccinated teenaged girls. In contrast, Price et al. (2011) empirically analyze the effect of HPV vaccination on the willingness to receive a Pap smear for women aged 18-74 and find that women who received HPV vaccination possess the high knowledge and intention to participate in a Pap smear. My empirical analysis supports the argument that HPV vaccination plays an important role in enlightening women to the significance of proper preventive care for HPV and cervical cancer. If a woman receives HPV vaccine, it is likely to make her consider the danger of cervical cancer and find relevant preventive care to reduce the likelihood of cervical cancer. In such a case, HPV vaccination could play an important role in increasing the proper frequency of Pap smears to avoid cancer.

With regard to potential ex-ante moral hazard related to HPV vaccination, I observe that the odds ratio of receiving HPV in Model I is less than 1. I also observed that the relative risk ratios of receiving HPV to "No Pap" in Model II and III are less than 1. This indicates that HPV vaccination does not cause any ex-ante moral hazard; there is no evidence that the HPV vaccination decreases the tendency to receive Pap smears. This finding supports the argument of Moghtaderi and Dor (2016) who examine the short-run effect of HPV vaccination for cervical cancer on participation in the Pap test and find no evidence of ex-ante moral hazard in the short-run.

Lastly, I find from my empirical analysis that all of the health interventions I considered

in this study encourage women younger than age 30 years to choose HPV test along with a Pap smear. My results are consistent with previous studies that indicate doctors and other health providers still recommend HPV testing to women younger than age 30 years old, and they also ask HPV test along with a Pap smear because of their concern of cervical cancer and convenience of HPV test, a pattern inconsistent with USPSTF guidelines. Despite this guideline, it is not strongly prohibited in practice among young women. Thus, expert group and public health authority should consider the misuse of HPV test among women younger than age 30 years.

3.6 Conclusions

This study investigated the effects of three health-related interventions, recommendation by a doctor, HPV vaccination, and informal information on HPV, on the misuse of cervical cancer screening and the Pap smear screening decision among women younger than 30. The results show that doctor's or other health professional's recommendation plays a significant role in not only receiving Pap smears at the proper usage but also induces unnecessary HPV testing and too-early and overuse of annual Pap smear screening. Also, they show that HPV vaccination encourages compliance with the recommendations including Pap smear only and proper frequency among women aged 21-29 while it also induces misuse of screening including unnecessary HPV testing. In addition, hearing of HPV increases the likelihood of Pap smears at a proper frequency while it induces unnecessary HPV test and the overuse of screening.

To prevent young women from the misuse of cervical cancer screening, expert groups and public health authorities should provide precise and correct information about the appropriate cancer screening. Education program with such information in high school and college would be a proper policy to achieve this goal. Further, doctors and other health professionals should follow the guideline of cervical cancer screening and practice it

correctly for young women.

References

Akker-van Marle, M Elske van den, Ballegooijen, Marjolein van, Oortmarssen, Gerrit J van, Boer, Rob, and Habbema, J Dik F (2002). "Cost-effectiveness of cervical cancer screening: comparison of screening policies". *Journal of the National cancer institute* 94.3, pp. 193–204.

Almeida, Cristina M, Rodriguez, Michael A, Skootsky, Samuel, Pregler, Janet, Steers, Neil, and Wenger, Neil S (2013). "Cervical cancer screening overuse and underuse: patient and physician factors." *The American Journal of Managed Care* 19.6, pp. 482–489.

American Cancer Society (Jan. 2017a). *What Are the Key Statistics About Cervical Cancer?* https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html.

— (2017b). *What Are the Risk Factors for Cervical Cancer?* https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html (last visited: 3/13/2018).

Berger, Barry M, Schroy, Paul C, and Dinh, Tuan A (2016). "Screening for colorectal cancer using a multitarget stool DNA test: modeling the effect of the intertest interval on clinical effectiveness". *Clinical Colorectal Cancer* 15.3, e65–e74.

Budd, Alison C, Brotherton, Julia M L, Gertig, Dorota M, Chau, Theresa, Drennan, Kelly T, and Saville, Marion (2014). "Cervical screening rates for women vaccinated against human papillomavirus". *The Medical Journal of Australia* 201.5, pp. 279–282.

CDC (2015). *Flu Vaccination Coverage, United States, 2015-16 Influenza Season.* https://www.cdc.gov/flu/fluvaxview/coverage-1516estimates.htm (last visited: 6/14/2017).

Centers for Disease Control and Prevention (n.d.). *Screening Recommendations and Considerations.* https://www.cdc.gov/cancer/knowledge/provider-education/cervical/recommendations.htm.

Chen, Han-Yang, Kessler, Courtenay L, Mori, Naoyo, and Chauhan, Suneet P (2012). "Cervical cancer screening in the United States, 1993–2010: characteristics of women who are never screened". *Journal of women's health* 21.11, pp. 1132–1138.

Coughlin, Steven S, Breslau, Erica S, Thompson, Trevor, and Benard, Vicki B (2005). "Physician recommendation for Papanicolaou testing among US women, 2000". *Cancer Epidemiology and Prevention Biomarkers* 14.5, pp. 1143–1148.

Daley, Ellen, Perrin, Kay, Vamos, Cheryl, Hernandez, Natalie, Anstey, Erica, Baker, Elizabeth, Kolar, Stephanie, and Ebbert, Judith (2013). "Confusion about Pap smears: lack of knowledge among high-risk women". *Journal of women's health* 22.1, pp. 67–74.

Drolet, Mélanie, Bénard, Élodie, Boily, Marie-Claude, Ali, Hammad, Baandrup, Louise, Bauer, Heidi, Beddows, Simon, Brisson, Jacques, Brotherton, Julia ML, Cummings, Teresa, et al. (2015). "Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis". *The Lancet infectious diseases* 15.5, pp. 565–580.

Ferrante, Jeanne M, Chen, Ping-Hsin, Crabtree, Benjamin F, and Wartenberg, Daniel (2007). "Cancer screening in women: body mass index and adherence to physician recommendations". *American Journal of Preventive Medicine* 32.6, pp. 525–531.

Ferris, Daron G, Waller, Jennifer, Dickinson, Ashley, McCracken, Courtney, and Goebel, Angela (2012). "Impact of pap test compliance and cervical cancer screening intervals on human papillomavirus vaccine acceptance". *Journal of Lower Genital Tract Disease* 16.1, pp. 39–44.

Fisher, Sue (1984). "Doctor-patient communication: a social and micro-political performance". *Sociology of Health & Illness* 6.1, pp. 1–29.

Gath, DH, Hallam, N, Mynors-Wallis, L, Day, A, and Bond, SA (1995). "Emotional reactions in women attending a UK colposcopy clinic." *Journal of Epidemiology & Community Health* 49.1, pp. 79–83.

Gerend, Mary A, Shepherd, Melissa A, Kaltz, Emily A, Davis, Whitney J, and Shepherd, Janet E (2017). "Understanding women's hesitancy to undergo less frequent cervical cancer screening". *Preventive Medicine* 95, pp. 96–102.

Grady, Denise (2009). "Guidelines push back age for cervical cancer tests". *The New York Times website. Available at: http://www. nytimes. com/2009/11/20/health/20pap. html.*

Habbema, Dik, De, KOK, INGE, MCM, and BROWN, MARTIN L (2012). "Cervical cancer screening in the United States and the Netherlands: a tale of two countries". *The Milbank Quarterly* 90.1, pp. 5–37.

Ideström, Monica, Milsom, Ian, and Andersson-Ellström, Agneta (2003). "Women's experience of coping with a positive Pap smear: a register-based study of women with two consecutive Pap smears reported as CIN 1". *Acta obstetricia et gynecologica Scandinavica* 82.8, pp. 756–761.

Korfage, IJ, Ballegooijen, Marjolein van, Wauben, Brendy, Looman, CWN, Habbema, JDF, and Essink-Bot, M-L (2012). "Having a Pap smear, quality of life before and after cervical screening: a questionnaire study". *BJOG: An International Journal of Obstetrics & Gynaecology* 119.8, pp. 936–944.

Lee, Jennifer Wai-Yin, Berkowitz, Zahava, and Saraiya, Mona (2011). "Low-risk human papillomavirus testing and other nonrecommended human papillomavirus testing practices among US health care providers". *Obstetrics & Gynecology* 118.1, pp. 4–13.

Liverani, Carlo Antonio (2015). "Reconsidering primary HPV testing in cervical cancer screening". *Journal of Life Sciences* 9, pp. 83–89.

Massad, L Stewart, Einstein, Mark, Myers, Evan, Wheeler, Cosette M, Wentzensen, Nicolas, and Solomon, Diane (2009). "The impact of human papillomavirus vaccination on cervical cancer prevention efforts". *Gynecologic Oncology* 114.2, pp. 360–364.

Meissner, Helen I, Tiro, Jasmin A, Yabroff, K Robin, Haggstrom, David A, and Coughlin, Steven S (2010). "Too much of a good thing? Physician practices and patient willingness for less frequent Pap test screening intervals". *Medical care*, pp. 249–259.

Moghtaderi, Ali and Dor, Avi (2016). *Immunization and Moral Hazard: The HPV Vaccine and Uptake of Cancer Screening*. Tech. rep. National Bureau of Economic Research.

Moyer, Virginia A (2012). "Screening for cervical cancer: US Preventive Services Task Force recommendation statement". *Annals of Internal Medicine* 156.12, pp. 880–891.

National Cancer Institute (2016). *The American Cancer Society Guidelines for the Prevention and Early Detection of Cervical Cancer*. https://www.cancer.org/cancer/cervical-cancer/prevention-and-early-detection/cervical-cancer-screening-guidelines.html (last visited: 3/13/2018).

— (2017). *Understanding Cervical Changes: A Health Guide for Women*. https://www.cancer.gov/types/cervical/understanding-cervical-changes (last visited: 3/13/2018).

Price, Rebecca Anhang, Koshiol, Jill, Kobrin, Sarah, and Tiro, Jasmin A (2011). "Knowledge and intention to participate in cervical cancer screening after the human papillomavirus vaccine". *Vaccine* 29.25, pp. 4238–4243.

Saslow, Debbie, Runowicz, Carolyn D, Solomon, Diane, Moscicki, Anna-Barbara, Smith, Robert A, Eyre, Harmon J, and Cohen, Carmel (2002). "American Cancer Society guideline for the early detection of cervical neoplasia and cancer". *CA: a cancer journal for clinicians* 52.6, pp. 342–362.

Sawaya, George F, Kulasingam, Shalini, Denberg, Thomas D, and Qaseem, Amir (2015). "Cervical cancer screening in average-risk women: best practice advice from the Clinical

Guidelines Committee of the American College of Physicians". *Annals of Internal Medicine* 162.12, pp. 851–859.

Schofield, Margot J, Sansonfisher, Rob, Halpin, Stephen, and Redman, Selina (1994). "Notification and Follow-Up of Pap Test-Results: Current Practice and Women' s Preferences". *Preventive Medicine* 23.3, pp. 276–283.

Siegel, Rebecca L., Miller, Kimberly D., and Jemal, Ahmedin (2017). "Cancer statistics, 2017". *CA: A Cancer Journal for Clinicians* 67.1, pp. 7–30. ISSN: 1542-4863. DOI: 10.3322/caac.21387. URL: http://dx.doi.org/10.3322/caac.21387.

Sirovich, Brenda E and Welch, H Gilbert (2004). "The frequency of Pap smear screening in the United States". *Journal of General Internal Medicine* 19.3, pp. 243–250.

Sopracordevole, Francesco, Cigolot, Federica, Mancioli, Francesca, Agarossi, Alberto, Boselli, Fausto, and Ciavattini, Andrea (2013). "Knowledge of HPV infection and vaccination among vaccinated and unvaccinated teenaged girls". *International Journal of Gynecology & Obstetrics* 122.1, pp. 48–51.

StataCorp (2013). *Stata 13 Base Reference Manual: mlogit.* https://www.stata.com/manuals13/rmlogit.pdf (last visited: 3/13/2018).

Taylor, Vicky, Lessler, Daniel, Mertens, Kathy, Tu, Shin-Ping, Hart, Alton, Chan, Nadine, Shu, Jianfen, and Thompson, Beti (2003). "Colorectal cancer screening among African Americans: the importance of physician recommendation." *Journal of the National Medical Association* 95.9, p. 806.

Watson, Meg, Benard, Vicki, King, Jessica, Crawford, Anatasha, and Saraiya, Mona (2017). "National assessment of HPV and Pap tests: Changes in cervical cancer screening, National Health Interview Survey". *Preventive Medicine* 100, pp. 243–247.

White, Arica (2017). "Cancer screening test use–United States, 2015". *Morbidity and Mortality Weekly Report* 66.

Yarnall, Kimberly SH, Pollak, Kathryn I, Østbye, Truls, Krause, Katrina M, and Michener, J Lloyd (2003). "Primary care: is there enough time for prevention?" *American Journal of Public Health* 93.4, pp. 635–641.

Tables

Table 3.1: Recommendation summary of cervical cancer screening by age group

Age group	**Screening recommendation**	
	Pap smear	HPV test
18 – 20 years	Not recommended	Not recommended
21 – 29 years	Every 3 years	Not recommended
30 – 65 years	Every 3 years	Every 5 years
66 or more years	Not recommended†	Not recommended†

† The USPSTF recommends against screening for cervical cancer in women older than age 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer.

Table 3.2: Classification of dependent variables by models

Model	**Age**	**Preventive behaviors**		
		Underuse of screening	Compliance with recommendations	Misuse of screening
I	**18-20**	-	No screening	Too-early screening
II	**21-29**	No screening	Pap smear only	Unnecessary HPV test
III	**21-29**	No screening	Proper frequency	Annual Pap

DK/RF: Don't know or refuse to answer

Table 3.3: Descriptive statistics of total sample and subsample of model I

Variable	**Subsample of Model I**		
	No Pap	Too-early Pap	Total
Observation	285	176	461
(Percent observation)	(0.618)	(0.382)	
Demographics			
Average Age	18.95	19.28	19.08
Percent Non-white	0.249	0.318	0.275
Percent Married	0.028	0.097	0.054
Percent Employed	0.414	0.597	0.484
Health Variables			
Percent Flu shot	0.305	0.301	0.304
Percent In bed due to illness	0.432	0.472	0.447
Interventions			
Doctor recommended			
Percent Yes	0.025	0.511	0.210
Percent Didn't See a doctor	0.060	0.034	0.050
Percent Heard about HPV	0.698	0.750	0.718
Percent HPV vaccination	0.491	0.455	0.477

Table 3.4: Descriptive statistics of subsample of model II

Variable	Subsample of Model II			
	No Pap	Pap only	Pap with HPV	Total
Observation	353	817	808	1,978
(Percent observation)	(0.179)	(0.413)	(0.408)	
Demographics				
Average Age	23.81	25.39	25.53	25.16
Percent Non-white	0.323	0.256	0.259	0.269
Percent Married	0.153	0.351	0.282	0.288
Percent Employed	0.629	0.645	0.713	0.670
Health Variables				
Percent Flu shot	0.207	0.339	0.394	0.338
Percent In bed due to illness	0.377	0.425	0.453	0.428
Interventions				
Doctor recommended				
Percent Yes	0.110	0.523	0.614	0.486
Percent Didn't See a doctor	0.130	0.042	0.022	0.050
Percent Heard about HPV	0.640	0.756	0.957	0.817
Percent HPV vaccination	0.204	0.301	0.496	0.363

Table 3.5: Descriptive statistics of subsample of model III

Variable	Subsample of Model III			
	No Pap	Not annually	Annually	Total
Observation	375	953	448	1,776
(Percent observation)	(0.211)	(0.537)	(0.252)	
Demographics				
Average Age	23.95	25.04	26.04	25.06
Percent Non-white	0.317	0.254	0.228	0.261
Percent Married	0.168	0.304	0.386	0.296
Percent Employed	0.629	0.679	0.674	0.667
Health Variables				
Percent Flu shot	0.208	0.337	0.391	0.323
Percent In bed due to illness	0.381	0.441	0.440	0.428
Interventions				
Doctor recommended				
Percent Yes	0.139	0.503	0.607	0.452
Percent Didn't See a doctor	0.131	0.048	0.004	0.055
Percent Heard about HPV	0.640	0.846	0.902	0.816
Percent HPV vaccination	0.205	0.381	0.393	0.347

Table 3.6: Estimates and standard errors by models

Variable	Model I	Model II†		Model III††	
	Too-early pap	No Pap	Pap with HPV	No Pap	Annually
Demographics					
Age	0.520**	-0.262**	0.061**	-0.197**	0.172**
	(0.160)	(0.031)	(0.023)	(0.029)	(0.024)
Non-white	0.533	0.163	0.082	0.140	-0.128
	(0.281)	(0.165)	(0.122)	(0.157)	(0.142)
Married	1.115	-0.779**	-0.307*	-0.628**	0.156
	(0.581)	(0.182)	(0.122)	(0.174)	(0.130)
Employed	0.666**	-0.024	0.128	-0.092	-0.103
	(0.253)	(0.154)	(0.117)	(0.148)	(0.131)
Health Variables					
Flu shot	-0.262	-0.195	0.144	-0.315*	0.091
	(0.286)	(0.168)	(0.111)	(0.160)	(0.125)
In bed due to illness	0.128	-0.076	-0.012	-0.092	-0.059
	(0.257)	(0.149)	(0.106)	(0.141)	(0.119)
Interventions					
Doctor recommended					
Yes	3.880**	-2.105**	0.300**	-1.697**	0.302*
	(0.425)	(0.195)	(0.105)	(0.176)	(0.122)
Didn't See a doctor	0.058	0.561*	-0.313	0.487*	-2.295**
	(0.501)	(0.271)	(0.319)	(0.245)	(0.728)
Heard HPV	0.290	-0.312	1.726**	-0.823**	0.491*
	(0.333)	(0.172)	(0.199)	(0.167)	(0.195)
HPV vaccination	-0.358	-0.671**	0.550**	-0.781**	0.163
	(0.287)	(0.178)	(0.114)	(0.165)	(0.131)

**$p<0.01$, *$p<0.05$, † Reference category is "Pap smear only," †† Reference category is "Proper frequency"

Table 3.7: Odds ratio/relative risk ratios and standard errors by models

Variable	Model I	Model II†		Model III††	
	Too-early pap	No Pap	Pap with HPV	No Pap	Annually
Demographics					
Age	1.571**	0.769**	1.063**	0.821**	1.188**
	(0.305)	(0.024)	(0.024)	(0.024)	(0.029)
Non-white	1.704	1.177	1.086	1.150	0.880
	(0.547)	(0.195)	(0.132)	(0.180)	(0.125)
Married	3.048	0.459**	0.735*	0.534**	1.169
	(2.056)	(0.083)	(0.089)	(0.093)	(0.152)
Employed	1.946 **	0.977	1.137	0.912	0.902
	(0.250)	(0.150)	(0.133)	(0.135)	(0.118)
Health Variables					
Flu shot	0.769	0.763	1.155	0.730*	1.095
	(0.250)	(0.128)	(0.128)	(0.117)	(0.137)
In bed due to illness	1.136	0.927	0.988	0.912	0.943
	(0.309)	(0.138)	(0.105)	(0.128)	(0.112)
Interventions					
Doctor recommended					
Yes	48.428**	0.122**	1.350**	0.183**	1.352*
	(15.758)	(0.024)	(0.146)	(0.032)	(0.165)
Didn't See a doctor	1.060	1.752*	0.731	1.627*	0.101**
	(0.486)	(0.475)	(0.242)	(0.399)	(0.073)
Heard HPV	1.337	0.732	5.619**	0.439**	1.634*
	(0.581)	(0.122)	(1.118)	(0.073)	(0.319)
HPV vaccination	0.699	0.511**	1.734**	0.458**	1.177
	(0.214)	(0.091)	(0.198)	(0.076)	(0.154)

**$p<0.01$, *$p<0.05$, † Reference category is "Pap smear only," †† Reference category is "Proper frequency"