Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2018-25 | March 12, 2018 | http://www.economics-ejournal.org/economics/discussionpapers/2018-25

Can heterogeneity in reporting behavior explain the gender gap in self-assessed health status?

Dilek Basar and Mehmet A. Soytas

Abstract
This paper explains the gender differences in self-assessed health status by providing a theoretical identification mechanism through a dynamic structural model which allows for heterogeneity in discount factors of individuals. Theoretical predictions are empirically tested and estimation results support the structural model implications. The authors conclude that accounting for heterogeneity in individual discount factors explains a significant portion of the gender gap in self-assessed health status.

JEL I10 I12 I14 J16
Keywords Gender; self-assessed health status; discount factor heterogeneity; dynamic structural model; ordinal generalized linear model estimation

Authors
Dilek Basar, ✉ Department of Economics, Hacettepe University, Ankara, Turkey, dbasar@hacettepe.edu.tr
Mehmet A. Soytas, Faculty of Business, Ozyegin University, Istanbul, Turkey

Citation Dilek Basar and Mehmet A. Soytas (2018). Can heterogeneity in reporting behavior explain the gender gap in self-assessed health status? Economics Discussion Papers, No 2018-25, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2018-25

Received February 12, 2018 **Accepted as Economics Discussion Paper** March 7, 2018
Published March 12, 2018

1 Introduction

Although gender inequality in many aspects of society has started to be decreased, it is well-known that gender differentials in health are widespread (Waldron, 1997)[35]. Furthermore, it has been emphasized that there is a paradox for the relationship between gender and health. To be specific, women generally report poorer health than men on self-reported measures of health although they live longer than men (Ross and Bird, 1994)[27]. Men also have more severe chronic conditions than women do particularly at earlier age. In a similar vein, it has been proved that men adopt more unhealthy behaviors such as lack of physical activity, poor diet, the use of tobacco and alcohol, inadequate preventive care and risky behaviour (Courtenay, 2000)[9]. Gender differences in health related quality of life measures could be because of biological factors and/or other non-health related factors such as social, economic and demographic factors. In this context, gender differences in employment conditions and health behaviours may explain gender gap in perceived health status. Lower labour force participation rates of women, lower average income, and more work distress can be counted as the labour market related reasons why women are more likely than men to report poor health (Ross and Bird, 1994)[27]. As a consequence of social stratification and social gender perception, gender-based division of labour may lead to women poor perceived health status.

In this respect, it is important to investigate the factors underlying the gender paradox in self-assessed health (SAH) which is one of the most widely used subjective instruments of health especially in the analysis of equity and equality in health, factors affecting health status and utilization of health care services (Bago d'Uva et al., 2008)[3]. Although SAH cannot be regarded as a perfect proxy for health status due to its subjective nature, numerous studies have used SAH as an outcome variable in health care research (see, e.g.,Contoyannis et al., 2004[7]; Humphries and Van Doorslaer, 2000[17]; Van Doorslaer et al., 1997[34]; Trannoy et al., 2010[33]).

From a methodological point of view, SAH is prone to measurement error since it does not reflect the true health status, which may lead to heterogeneity in the reporting of health (Hernandez-Queveda et al., 2005)[16]. Individuals may report different levels on a categorical scale even if they have the same true health status because of different cut-off points. A growing number of studies have focused on the measurement error problems related to SAH measures (see, e.g., Crossley and Kennedy, 2002[11]; Jurges, 2007[20]; Brown et al., 2010[6]). The reporting bias problem can be dealt with using more objective health measures. One branch of the existing literature has used vignettes approach to address reporting heterogeneity (see, e.g., Rice et al., 2011[26]; Salomon et al., 2004[28]; Bago d'Uva et al., 2008)[3]) whereas another branch of the existing literature has used a generalized ordered probit or a hierarchical ordered probit model or both to correct for reporting bias (see, e.g., Lindeboom and van Doorslaer, 2004[22]; Jurges, 2007[20]). Notwithstanding these methodological concerns, most of the studies have highlighted that SAH is a valid measure that predicts objective health status such as mortality (Idler and Benyamini, 1997[19]; Contoyannis et al., 2004[7]).

The existing literature has focused on accounting for reporting heterogeneity but without providing a theoretical model. This study is the first attempt to investigate the gender gap in SAH by providing a theoretical model and by testing empirically the theoretical predictions in the context of adjusting for the heterogeneity in reporting behaviour. Although gender inequality is more common in developing world, most of the studies have investigated gender differences in reporting SAH for developed countries (see, e.g., Schulz et al., 1994[30]; Svedberg et al., 2001[31]). In this respect, this study also extends the existing literature by focusing on a developing country, Turkey, where gender differences in health are more pronounced.

The theoretical model used in this study proposes that SAH is a proxy for the respondents' perception of total utility derived from their health, which depends upon individual discount rates determining both current and expected future health levels. This approach enables us to distinguish current valuation functions for two individuals even if they have the same true unobserved health today. In this context, it is possible to argue that the gender gap in SAH may result from the differences in discount rates of women and men.

Although it is widely accepted in the existing literature that "women are more risk averse than men," the findings can be argued to be mixed (Nelson, 2014)[24]. This important argument generally tested in finance and behavioural experiments such as gambling (see, e.g., Ida and Goto, 2009[18]; Crosan and Greezy, 2009[10]). With respect to risky health behaviours, most of the studies employ smoking and obesity as proxies for individual discount rates (see, e.g., Borghans and Golsteyn, 2006[5]; Khwaja et al., 2007[21]; Scharff and Viscusi, 2011[29]; Harrison et al., 2010[15]; Ida and Goto, 2009[18]). Among these studies, Ida and Goto, 2009[18] and Scharff and Viscusi, 2011[29] found that smokers are more likely to have a higher time discount rate as compared to non-smokers (i.e., smokers were less future-oriented than non-smokers). This finding is not surprising because the strong relationship between time preference and smoking is well documented with the rational addition model proposed by Becker and Murphy, 1988[4]. From this perspective, when it comes to the relationship between gender, time preference and smoking trilateral relationship, Harrison et al., 2010[15] found that men smokers have significantly higher discount rates than men non-smokers whereas a similar relationship is not statistically significant for women.

In this context, this study investigates the role of heterogeneity in reporting behaviour in explaining the gender gap in SAH in Turkey using smoking as a proxy for individual discount rate. With this aim, Turkish National Health Survey for 2012 was used to estimate parameters obtained from the theoretical model.

The rest of the paper is structured as follows; Section 2 introduces the theoretical model and the identification equations while section 3 describes the identification and estimation methodology. In section 4, we outline the estimation results and the key findings. Section 5 concludes.

2 Model

Anderson (1984)[1] discusses the inadequacy of the ordered logit/probit models with the key insight that the models available for ordered categorical response variables are not wide enough to cover the range of problems that might arise in practice. This is particularly true for the micro-level data, and starting with Terza (1985)[32], many authors have paid attention on the fact that the models are not capable enough to account for the individual level heterogeneity which is very likely to be a major concern in micro-level studies. The presence of individual level heterogeneity carefully taken into the model environment is likely to be the most important distinction of social science applications of ordered choice models. This is almost inevitable in an application with SAH, where the reporting of the health is pure subjective in contrast to an outcome obtained through a medical examination of the individual. On the other hand, despite the lack of rigid objectivity, SAH questions are almost available in all main surveys and this makes them very convenient and readily available. Existence of many individual level controls for income, education, age, race and family structure in the aforementioned surveys makes the researchers use SAH as the one of the main modelling tools for health research.

In the model proposed here, a structural behavioral model is assumed for the individual. Individual derives utility from her current health status and she responds to changes in her life-cycle health through a discount rate. Namely, an individual calculates her current value of health as a sum of current and discounted future health benefits. In this context, we set up the current value of an individual's health as follows:

$$V(x_{it}) = u(x_{it}) + E_t \left\{ \sum_{s=t+1}^{T} \beta_i^{s-t} u(x_{is}) \right\} \tag{1}$$

where x_{it} is the vector of observed characteristics of the individual i at time t. $u(x_{it})$ is the time t health utility function and β_i is discount factor of the individual. The value of the total health benefits enjoyed by the individual is represented by $V(x_{it})$. We propose a linear health utility function, which depends on observable individual characteristics vector x_{it}, the vector of marginal utility coefficients for the observed characteristics γ' and an unobserved component of the utility ε_{it}.

$$u(x_{it}) = \gamma' x_{it} + \varepsilon_{it} \tag{2}$$

Using the linear utility function and replacing it into (1), we get the following expression for the value of health:

$$V(x_{it}) = \gamma' x_{it} + \varepsilon_{it} + E_t \left\{ \sum_{s=t+1}^{T} \beta_i^{s-t} \left(\gamma' x_{is} + \varepsilon_{is}\right) \right\} \tag{3}$$

The state vector x_{it} can be decomposed into three kind of observable sub-state vectors which are denoted by x_{1it}, x_{2it}, x_{3it}. Each of this sub-vectors has a special life-cycle formation and any observable individual characteristics, or demographics is assumed to fit into one of these categories of vectors. To illustrate the use of the sub-state vectors, without loss of generality it is simply assumed that each of the vectors x_{1it}, x_{2it}, x_{3it} just contains one variable. The definition of x_{it} is also extended by adding 1 as the first element. Therefore x_{it} vector now has four elements, a constant (1), education (x_{1it}), age (x_{2it}) and income (x_{3it}), which affect the health outcomes. The four variable case of equation (3) is expressed as follows:

$$\begin{aligned} V(x_{it}) &= \gamma_0+\gamma_1 x_{1i}+\gamma_2 x_{2it}+\gamma_3 x_{3it}+\varepsilon_{it} \\ &\quad +E_t\left\{\sum_{s=t+1}^{T}\beta_i^{s-t}\left[\gamma_0+\gamma_1 x_{1it}+\gamma_2 x_{2is}+\gamma_3 x_{3is}+\varepsilon_{is}\right] \mid x_{it}\right\} \end{aligned}$$

Education, income and age are widely used characteristics and demographics in the existing literature which are expected to be closely related to individual health status (see, e.g., Auster et. al. 1969[2]; Grossman, 1972 [14]; Or, 2000[25]; Contoyannis and Jones, 2004[8]). Each of these variables, however are different in terms of their life-cycle trajectories, especially from the perspective of a time t individual's expectation point of view. For the sake of derivation, we assume an infinite time horizon: i.e. $T=\infty$, but this assumption is not critical for the validity of the results and can be relaxed easily with a finite T. The first kind of variable is individual's education level (x_{1it}). This variable is fixed over the life-cycle of the individual[1], or for our purposes it implies $E_t[x_{1is}|x_{1it}]=x_{1it}$, for $s>t$. The second kind of variable is age (x_{2it}). For any future time s, age of an individual can be written as follows: $x_{2is}=x_{2it}+s-t$. Therefore age follows a deterministic trend. The third kind is income (x_{3it}), and it is assumed that future income of an individual depends on current income with the following trajectory: $x_{3is}=x_{3it}(1+\lambda(s-t))$, where λ is the annual growth in real income over years[2]. With these definitions, the value function of an individual can be expressed as follows:

$$\begin{aligned} V(x_{it}) &= \gamma_0+\gamma_1 x_{1i}+\gamma_2 x_{2it}+\gamma_3 x_{3it}+\varepsilon_{it} \\ &\quad +\gamma_0\sum_{s=t+1}^{\infty}\beta_i^{s-t}+\left(\gamma_1\sum_{s=t+1}^{\infty}\beta_i^{s-t}\right)E_t\{x_{1i}|x_{it}\}+\gamma_2 E_t\left\{\sum_{s=t+1}^{\infty}\beta_i^{s-t}(x_{2it}+s-t)\right\} \\ &\quad +\gamma_3 E_t\left\{\sum_{s=t+1}^{\infty}\beta_i^{s-t}(x_{3it}(1+\lambda(s-t)))|x_{it}\right\}+E_t\left\{\sum_{s=t+1}^{\infty}\beta_i^{s-t}\varepsilon_{is}|x_{it}\right\} \end{aligned}$$

The above expression is simplified by using the following facts:

$$\sum_{s=t+1}^{\infty}\beta_i^{s-t}=\frac{\beta_i}{1-\beta_i},\ \sum_{s=t+1}^{T}\beta_i^{s-t}(s-t)=\frac{\beta_i}{(1-\beta_i)^2}$$

[1] One can think of education as an indicator for the skill level of the individual for earlier ages. In the empirical application, sample contains individuals beyond the education age, i.e. age 25 and further.

[2] We follow Zabalza (1979) for the evolution of an individual's income.

Defining $\rho_i = \frac{\beta_i}{1-\beta_i}$ and using $E_t\{x_{1it}\} = x_{1i}$, $E_t\{x_{2it}\} = x_{2it}$ and $E_t\{x_{3it}\} = x_{3it}$, we write:

$$\begin{aligned} V(x_{it}) &= \gamma_0 + \gamma_1 x_{1it} + \gamma_2 x_{2it} + \gamma_3 x_{3it} + \varepsilon_{it} \\ &+ \gamma_0\rho_i + \gamma_1\rho_i x_{1it} + \gamma_2\rho_i x_{2it} + \gamma_2\rho_i(1+\rho_i) \\ &+ \gamma_3\rho_i x_{3it} + \gamma_3\lambda\rho_i(1+\rho_i)x_{3it} + \sum_{s=t+1}^{\infty}\beta_i^{s-t}E_t\{\varepsilon_{is}|x_{it}\} \end{aligned} \tag{4}$$

In this context, it is assumed that the unobserved component of the utility is independent across the individuals and over time with a zero mean conditional on observable state variables x_{it}, $(E[\varepsilon_{it}|x_{it}] = 0$, $E[\varepsilon_{it}\varepsilon_{i't}|x_{it}] = 0$, $E[\varepsilon_{it}\varepsilon_{is}] = 0)$. This assumption sets the last term in (4) to 0[3].

3 Estimation

We want to explore the equation in (4) in order to estimate the coefficients $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3, \lambda\}$. If we had assumed a constant discount rate β (so constant ρ) for all individuals, equation (4) would reduce to:

$$V(x_{it}) = \alpha_0 + \alpha_1 x_{1it} + \alpha_2 x_{2it} + \alpha_3 x_{3it} + \varepsilon_{it} \tag{5}$$

In this expression, the parameters correspond to: $\alpha_0 = (\gamma_0 + \gamma_0\rho) + \gamma_2\rho(1+\rho)$, $\alpha_1 = (\gamma_1 + \gamma_1\rho)$, $\alpha_2 = (\gamma_2 + \gamma_2\rho)$ and $\alpha_3 = \gamma_3 + \gamma_3\rho + \gamma_3\lambda\rho(1+\rho)$. Econometrically we can not separately identify the utility parameters $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ and the parameters ρ and λ. In fact the estimation equation in (5) is observationally equivalent to a static utility model that is estimated widely using an ordered probit (logit) model in the literature. Therefore the dynamic life-cycle model presented in this paper[4] has the potential to identify the period utility parameters and the parameter λ if we allow for heterogeneity in the discount rate β. For instance with ρ_i as a variable, we can write an estimable version of the equation in (4) as follows:

$$V(z_{it}) = \gamma_0 z_{0it} + \gamma_1 z_{1it} + \gamma_2 z_{2it} + \gamma_3 z_{3it} + (\gamma_3\lambda)z_{4it} + \varepsilon_{it} \tag{6}$$

[3] This assumption is likely to be a strong assumption. Especially given the fact that health schocks not captured by the observable characteristics are very often persistent. An important condition such as cardiovascular disease or hypertension is expected to prevail for the rest of the life once it is diagnosied for the individual. Therefore a more realistic assumption might be to model the unobserved component as an AR process. However this, though being important, is not modelled explicitly in the paper. The model with the persistence unobserved component makes the identification and estimation more complex and it is a topic of another paper (see Kose and Soytas, 2015)[23].

[4] Obviously, this is true with the proposed functional form of the utility and the specific type of variables and their life-cycle trajectories. Relaxing the linear utility or assuming other stochastic trends for the variables can possibly produce partial/full identification. However this will also bring other issues including taking the expectations of the nonlinear functions of the future values of the variables.

where $z_{0it} = 1 + \rho_i$, $z_{1it} = x_{1it} + \rho_i x_{1it}$, $z_{2it} = \rho_i + \rho_i^2 + x_{2it} + \rho_i x_{2it}$, $z_{3it} = x_{3it} + \rho_i x_{3it}$, $z_{4it} = \rho_i x_{3it} + \rho_i^2 x_{3it}$ and $z_{it} = (z_{0it}, z_{1it}, z_{2it}, z_{3it}, z_{4it})'$. Therefore if we have a measure of the discount rates of individuals (ρ_i), the variation in this can help us identify all of the coefficients of the theoretical model proposed in the Model Section. We generally can observe education, income and the age of an individual from a given survey sample, but still the estimation of the model requires direct measures of health valuations $V(z_{it})$ given in (6). However, in most data sets, we only observe self-assessed health outcomes (SAH) rather than a continuous measure for health benefits/levels. Therefore, in order to make the model in (6) operational, we need to define a structure for the unobserved component ε_{it} and set the relation of the health benefits/levels to the self-assessed health outcomes. This unobserved component is crucial for the model estimation and theoretically it also contributes for the differences in self-assessed health statuses.

Health benefits of an individual cannot be observed. However, being a function of health benefits, self-assessed health statuses are observed as a discrete variable. We define the relationship between the valuation function and discretely reported SAH categories as follows:

$$SAH_i = \left\{ \begin{array}{ccc} 1 & if & -\infty < V(z_{it}) \leq c_1 \\ 2 & if & c_1 < V(z_{it}) \leq c_2 \\ 3 & if & c_2 < V(z_{it}) \leq c_3 \\ 4 & if & c_3 < V(z_{it}) \leq c_4 \\ 5 & if & c_4 < V(z_{it}) \leq \infty \end{array} \right\} \quad (7)$$

Unlike general models using SAH measure in the literature, we hypothesize that SAH reflects not only current health but also life-time health benefits. In fact, having the SAH measure as an indicator for the current health, most models accumulate effects of life-cycle expectations into the model through parameters and/or the unobserved component. The estimation of the health model with SAH, in this framework, depends on the empirical assumptions made about the relation of the SAH categories and the model. Further, the unobserved error term of the model determines the estimation model type used in the literature. For instance, a normally distributed error term will lead to an ordered probit whereas a logistic distribution assumption will drive an ordered logit model. In most cases, a threshold structure combined with the distributional assumption on ε_{it} is the only way to assign a latent health status to discrete SAH outcomes.

When we consider all these facts, it is possible to argue that this paper's contribution to the existing literature is twofold. First it proposes a theoretical framework that explores the dynamic structure in the health valuation of the individuals. Most models in the literature does not depend on a robust underlying theory for the data generating process, but assume equation (5) as the starting point for the analysis. As shown above, equation (5) is consistent with a life-cycle model as well as a static model, and without further structure, the data is not capable of making the distinction. This is important since, the coefficients estimated with (5) when the life-cycle model is the true model will be contaminated with the effects of various channels; i.e. both current period and future observable variables' effect will be reflected in the coefficients. Therefore first, the theoretical model allows us to have a mechanism to think about the health benefits consistent with the rational economic agent's behavior. Secondly, and perhaps more importantly, the heterogeneous discount rates ρ_i developed in the model can be used to estimate the model parameters. This will allow us to test the model against the static model and use the various predictions from the model to answer some of the policy relevant issues. One of which is the paradox that women generally report poorer health than men on self-reported measures of health although they live longer than men (Ross and Bird, 1994)[27]. Individuals may report different levels on a categorical scale as SAH even if they have the same true health status because of different cut-off points. This is investigated widely in the health literature using SAH, however most of the papers attacked this problem using a statistical approach mainly to identify the key variables that may lead to the differences in the reporting behaviour. In this paper, we propose a behavioural model about why the reporting behaviour might change across individuals. To be specific, we test that whether the reason women generally report poorer health can be related to the heterogeneity in the discounting in gender. With the estimation equation (6) induced from the model, first we need to develop the empirical measurement for the discount factor to answer this question. That is what we do next.

3.1 Accounting for Individual Subjective Discount Rates

Theoretical model with heterogeneity in individual discount rates enables us identify utility parameters. Ideally we would replace ρ_i for each individual in z_{it} in equation (6) and estimate the model, for instance using an ordered probit specification. However in general we don't have a measure for individual discount factor and it is by itself a measure to be derived using the sample information at hand. Since discount rates are shown (Harrison et al., 2010[15]) to vary significantly with respect to socioeconomic conditions and demographics, we assume the following functional form of the relationship between individual discount rates and socioeconomic variables.

$$\rho_i = \theta' w_i + u_i \tag{8}$$

where θ is a $L \times 1$ parameter vector and w_i is a $L \times 1$ vector of containing socioeconomic variables for an individual i. The main hypothesis of this paper implies that coefficients, γ, identified in this framework should be significantly different from the coefficients identified without future looking structure. In the latter case, the coefficients we can identify are $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$. In this regard, the essential contribution of this paper is the inclusion of stochastic individual specific discount rate. We use our theoretical results to show that the gender gap in self-rated health statuses, a puzzle in the health economics literature, can be partly explained by taking the differences in stochastic discounting into account.

3.2 Identification

It should be clear that the underlying heterogeneity in the individual discount rates interacting with the different types of characteristics and demographics variables has a wide potential for the identification. However, this theoretically novelty comes at a cost, depending on the type of the variables used. The health benefits/levels can only be assigned to the self-assessed health outcomes in the estimation. This creates an ordered outcomes model for the health valuations and main implication of this structure is the requirement to derive the final estimation equation. This final estimation equation differs depending on which types of characteristics and demographics variables used, i.e. $1, x_{1i}, x_{2i}, x_{3i},$. The model for instance can be estimated using only a constant (or in other words 1 as a variable only), which is the case when the period health utility is fixed for everyone for every period, but only the subjective discount rate is different[5]. In this case the simplest model becomes:

$$V(x_{it}) = \gamma_0 + \gamma_0\theta' w_i + \gamma_0 u_i + \varepsilon_{it} \tag{9}$$

This model with an assumption about the distribution of $\gamma_0 u_i + \varepsilon_{it}$ and the relationship between the valuation function and discretely reported SAH categories defined as in (7) identifies the coefficients, γ_0 and θ.

If we include only the variables of type 1 (x_{1i}) in addition to a constant in x_{it}, the valuation functions can be written as:

$$V(x_{it}) = \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \gamma_1 x_{1it} u_i + \gamma_0 u_i + \varepsilon_{it} \tag{10}$$

As can be seen in equation (10), the valuation function now includes interaction terms in observed and unobserved variables, therefore it requires additional assumptions for the estimation (i.e. the distribution of the composite error $\gamma_1 x_{1it} u_i + \gamma_0 u_i + \varepsilon_{it}$, and the correlation between u_i and ε_{it}). However the coefficients are still identified, the final estimation form may not be trivial to estimate with a plain ordered probit for instance. We will show this in several steps below and derive an estimation equation for this model. The SAH score of an individual allows us to write the model as follows. Let $\eta_{it} = (\gamma_1 x_{1it} u_i + \gamma_0 u_i + \varepsilon_{it})$:

$$SAH_i = \left\{ \begin{array}{lll} 1 & if & -\infty < \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \eta_{it} \leq c_1 \\ 2 & if & c_1 < \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \eta_{it} \leq c_2 \\ 3 & if & c_2 < \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \eta_{it} \leq c_3 \\ 4 & if & c_3 < \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \eta_{it} \leq c_4 \\ 5 & if & c_4 < \gamma_0 + \gamma_0\theta' w_i + \gamma_1 x_{1i} + \gamma_1\theta' w_i x_{1it} + \eta_{it} \leq \infty \end{array} \right\} \tag{11}$$

[5] This will be an unrealistic model of course, but it is instructive to show the identification power of the structure.

3.3 A naive estimator with bootstrap

As the above discussion on identification shows, with our general structure of x_{it}, the estimation of the model can be nontrivial, although identification remains straightforward. The details of the general case in terms of its relation to the SAH outcomes are given Kose and Soytas, 2015[23]. Here we divert to a more simpler approach, and try to answer the question of interest in a direct way. It can be less efficient econometrically, but if it proves the importance of discount rates in reporting behaviour, then estimating the cumbersome general model efficiently gains more meaning.

Let $\hat{\rho}_i$ denotes the estimated counterpart of ρ_i from equation (8). Using this as a measure, we can obtain $\hat{z}_{0it}= 1+\hat{\rho}_i$, $\hat{z}_{1it}= x_{1it}+\hat{\rho}_i x_{1it}$, $\hat{z}_{2it}= \hat{\rho}_i+\hat{\rho}_i^2+x_{2it}+\hat{\rho}_i x_{2it}$, $\hat{z}_{3it}= x_{3it}+\hat{\rho}_i x_{3it}$, $\hat{z}_{4it}= \hat{\rho}_i x_{3it}+\hat{\rho}_i^2 x_{3it}$, and replace them in (6), to obtain:

$$V(\hat{z}_{it}) = \gamma_0\hat{z}_{1it} + \gamma_1\hat{z}_{1it} + \gamma_2\hat{z}_{2it} + \gamma_3\hat{z}_{3it} + (\gamma_3\lambda)\hat{z}_{4it} + \varepsilon_{it} \qquad (12)$$

and the related threshold structure:

$$SAH_i = \left\{ \begin{array}{lll} 1 & if & -\infty < V(\hat{z}_{it}) \leq c_1 \\ 2 & if & c_1 < V(\hat{z}_{it}) \leq c_2 \\ 3 & if & c_2 < V(\hat{z}_{it}) \leq c_3 \\ 4 & if & c_3 < V(\hat{z}_{it}) \leq c_4 \\ 5 & if & c_4 < V(\hat{z}_{it}) \leq \infty \end{array} \right\} \qquad (13)$$

Obviously, using estimated variables from a first stage will make the standard errors of the coefficients incorrect. We propose a bootstrap procedure to correct for this bias.

4 Estimation Results

4.1 Data Description

We employ data from the Turkish National Health Survey for 2012. The self-assessed health status in the survey is coded as an ordinal variable. The corresponding question for SAH reads: "What is the status of your health?"

$$SAH_i = \left\{ \begin{array}{ll} 1 & \text{if individual reports "Very bad"} \\ 2 & \text{if individual reports "Bad"} \\ 3 & \text{if individual reports "Fair"} \\ 4 & \text{if individual reports "Good"} \\ 5 & \text{if individual reports "Very Good"} \end{array} \right\}$$

The female and male differs in their reporting of SAH in the data.

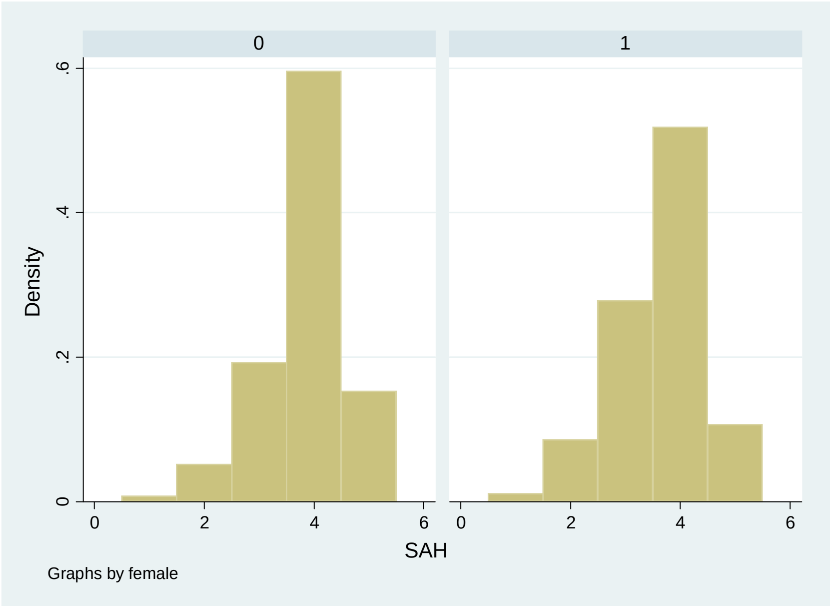

Figure 1: Male and Female SAH outcomes

We see females report less of the ""Very Good"" and "Good" outcomes and report more of the "Fair" and "Bad" outcomes compared to their male counterparts. This result can be an actual indicator of women having less health (or feeling less healthy) than men. However it might also be an indicator of a difference in reporting behavior. The SAH is a subjective outcome obtained by the self reporting of the individual about his/her own health. Therefore any variation in the anticipation of the question, have the potential to produce a spurious outcome as in Figure 1, even though there might have been no difference in actual healths of the individuals. To overcome this reporting issue and potentially ruling it out as an explanation requires individual level heterogeneity to be accounted for in the analysis. The best way to proceed in this direction would be to contrast the SAH outcomes of the individuals with some objective measure of health of theirs and check the consistency of the results. Unfortunately most surveys that ask the SAH question are generally not rich enough in other questions (or medical records) that allows the research to follow this straightforward path. The 2012 Turkish National Health Survey is not an exception. It contains the SAH but does not have a direct objective health record. However unlike many other surveys not aiming the health outcomes as the primary focus, it is relatively extended in the supporting questions about the health outcomes.

4.2 Equation for the discount factor

We use the smoking bahaviour of individuals as a factor to identify the individual discount factors. We control for individual characteristics and gender in the equation. The estimated equation takes the following form:

$$\rho_i = \theta_0 + \theta_1 Gender_i + \theta_2 educ_i + \theta_3 age_i + \theta_4 income_i + u_i \tag{14}$$

where ρ_i is a proxied by a variable constructed to capture the smoking intensity[6]. The predicted values from this equation, $\hat{\rho}_i$ will be used as a proxy for $\hat{\rho}_i$. Obviously we make the following assumption here: gender, education, age and income capture the variation in the smoking behavior that is related to the discounting behavior. Therefore the error term u_i is the individual heterogeneity in smoking, and is not related to the discounting behavior. This assumption might be strong, but previous research identifies those factors as the important factors in discounting behavior. Moreover it is the functional form based on the structural model which gives the identification in our methodology, so it is not required to augment the discounting equation with extra variables that are not present in the main utility equation. However those extensions can be important, at least in terms efficiency and left as a topic for future research.

4.3 Equation for the utility parameters

We make the following assumptions for the model error terms in equation (12) :$E[\varepsilon_{it}] = 0$, $E[\varepsilon_{it}|x_{it}] = 0$, $E[\varepsilon_{it}\varepsilon_{i't}] = 0$, $E[\varepsilon_{it}\varepsilon_{is}] = 0$. The unobserved component of the utility is independent across the individuals and over time with a zero mean conditional on observables The same covariates are used in the estimation as described in the discount equation. The predicted values from the discount factor equation $(\hat{\rho}_i)$ are used to construct the $\hat{z}_{0it} = 1 + \hat{\rho}_i$, $\hat{z}_{1it} = x_{1it} + \hat{\rho}_i x_{1it}$, $\hat{z}_{2it} = \hat{\rho}_i + \hat{\rho}_i^2 + x_{2it} + \hat{\rho}_i x_{2it}$, $\hat{z}_{3it} = x_{3it} + \hat{\rho}_i x_{3it}$, $\hat{z}_{4it} = \hat{\rho}_i x_{3it} + \hat{\rho}_i^2 x_{3it}$. The valuation of the individual conditional on the state variable $\hat{z}_{it}$ is given as:

$$V(\hat{z}_{it}) = \gamma_0 \hat{z}_{1it} + \gamma_1 \hat{z}_{1it} + \gamma_2 \hat{z}_{2it} + \gamma_3 \hat{z}_{3it} + (\gamma_3 \lambda) \hat{z}_{4it} + \varepsilon_{it} \tag{15}$$

4.4 Parameter estimates and results

An important focus of this study is on the role of heterogeneity in reporting behaviour in explaining gender paradox relating to SAH by claiming that an individual calculates her current value of health considering both current and discounted future health benefits. Therefore, rather than assuming a constant discount rate for all individuals, the model is extended to include stochastic individual specific discount rate. The results of the model that accounts for the discount rate proxied by smoking intensity are presented in Table 1 (Model 1 and Model 2)[7].

[6] The variable is constructed as a continuous measure. It takes a value of 5 for smoking regularly, 4 for smoking sometimes, 2 for previously smoked but quiting, and 1 for never smoking,

[7] The difference between Model 1 and 2 is that Model 1 includes less interaction terms whereas Model 2 includes all interaction terms.

However, for comparison purposes, the results of the ordered probit model, which is the base model not considering future looking structure, are also presented in Table 1.

The most pronounced and arguably important finding relates to the sign of the female coefficient. To be specific, the negative sign of female coefficient in the base model, meaning that females are less likely to report good health as compared to males, turns to be positive for the discount rate specifications where there is no reporting bias. Further, the coefficients are smaller in magnitude in the discount rate specifications. The marginal effects presented in Table 2 also indicate that females reported very good health level are more likely have a higher discount rate as compared to males which means that they are less-future oriented. These findings verified our hypothesis that the reason for females to report generally poorer health compared to their male counterparts is related to the heterogeneity in individual discount factors. In other words, accounting for heterogeneity in individual discount factors in the analysis provides an important explanation for the gender gap in SAH, which is the essential contribution of this study to the existing literature.

On the other hand, the results also provide implications relating to the association between individual discount rates and socioeconomic variables. For example, the marginal effects of education variable indicate that individuals reporting good and very good health are more likely to have higher levels of education while controlling for the individual discount rates. Moreover, the magnitude of the marginal effects becomes larger for the higher levels of educational attainment. This positive association between health and education supports the premise which argues that individuals with higher levels of education behave more cautiously against major health threats and constraints such as the deleterious effects of cigarette, alcohol, fat consumption or the importance of early diagnosis of diseases. Therefore, they may be seen as effective producers of health as they infer the results of every health input they use to produce health (Grossman, 1972)[14]. In addition to Grossman approach, Fuchs, 1982[13] developed a different approach relating to the path of the effect of education on health. According to Fuchs' time preference approach, an individual spending a greater proportion of her time for education will also allocate more resources to health. Therefore, the higher level of health status is a result of time preference rather than having higher educational attainment.

As is well documented in the existing literature, the association between health and income can be both negative and positive. First, income is a factor directly affecting the health status of the individual by providing better housing and living conditions. On the other hand, higher levels of income may also affect the lifestyle of the individuals and have a negative impact on health. Auster et al., 1969[2] state that, with an increase in their income levels, individuals shift their lifestyles where there is less exercise, more deleterious consumption patterns and more stress. Therefore, it is emphasized in the existing literature that there is a non-linear relationship between income and health. For example, Fuchs, 1974[12] point outs that the minimum level of income is important for health and, there is not a high correlation between income and health especially in developed countries when this income level is exceeded. In this context, the results of the model indicate a positive relationship between SAH and income for a developing country while accounting for individual discount rates in the analysis. The marginal effects of the model also indicate that the positive relationship between income and health is most apparent among individuals belonging to the richest income quintile.

As a final covariant, increase in age has a negative relationship with reporting good and very good health status as is expected. Individuals are expecting more health problems in the future and, therefore, report worse health status outcomes since we argue that the current value of health is calculated by an individual considering both current and discounted future health benefits.

5 Conclusion

The aim of this study is to explore the role of heterogeneity in reporting behaviour in explaining gender paradox for SAH by providing a theoretical identification mechanism. This mechanism argues that SAH is a proxy for the individuals' perception of total utility derived from their current and expected future health levels. The findings confirm that an important part of the gender gap in SAH is explained when heterogeneity in individual discount factors are controlled for in the analysis. More specifically, we find a clear evidence of heterogeneity in individual discount factors since the negative sign of female coefficient in the base model turns to be positive in the discount rate specifications.

From an empirical perspective, this study also suggests that the existing findings assuming a constant discount rate for all individuals could be quite different if individual specific discount rates were considered. Therefore, it can be argued that the results of the studies which do not adjust for the heterogeneity in reporting behaviour in the analysis should be treated with caution.

Table 1: Parameter Estimates

	Base Model	Model 1	Model 2		Base Model	Model 1	Model 2
						Cut-offs	
Constant			3.707		-2.680	-2.510	1.879
			(0.496)		(0.057)	(0.164)	(0.698)
Female	-0.286	0.126	0.158		-1.638	-1.463	2.921
	(0.017)	(0.121)	(0.125)		(0.048)	(0.161)	(0.694)
Education					-0.469	-0.287	4.088
incomplete	0.174	0.329	0.074		(0.046)	(0.160)	(0.696)
	(0.049)	(0.200)	(0.059)		1.444	1.632	5.999
primary	0.305	0.437	0.086		(0.047)	(0.160)	(0.697)
	(0.033)	(0.151)	(0.078)				
secondary	0.409	0.424	0.063				
	(0.041)	(0.190)	(0.108)				
highschool	0.514	0.801	0.189				
	(0.039)	(0.218)	(0.090)				
tertiary	0.648	0.980	0.417				
	(0.041)	(0.243)	(0.055)				
Age							
age 35-44	-0.294	-0.362	-0.233				
	(0.021)	(0.143)	(0.015)				
age 45-54	-0.577	-0.527	-0.373				
	(0.023)	(0.137)	(0.025)				
age 55-64	-0.892	-0.842	-0.485				
	(0.026)	(0.140)	(0.073)				
Income							
poor	0.171	-0.064	0.228				
	(0.039)	(0.150)	(0.067)				
medium	0.209	0.160	0.194				
	(0.035)	(0.138)	(0.062)				
rich	0.218	0.036	0.182				
	(0.036)	(0.149)	(0.061)				
veryrich	0.337	0.013	0.331				
	(0.036)	(0.152)	(0.065)				
lambda 1			-0.217				
			(0.182)				
lambda 2			-0.007				
			(0.183)				
lambda 3			-0.028				
			(0.178)				
lambda 4			-0.197				
			(0.176)				
Wald test		2889.8	3067.0	2528.1			
Pseudo R2		0.072	0.075	0.071			
Log likelihood		-19547	-19487	-19567			
Number of obs		19,308	19,308	19,308			

Note: Base Model represents the standard ordered logit estimation using the variables for the education, age and income as the explanatory variables. Model 1 includes the interaction terms of the education, age and income variables. Model 2 includes the structural model induced interaction terms of the education, age and income variables. The standard errors in Model 1 and 2 are calculated by Bootstrap by re-sampling 500 times.

Table 2: Marginal Effects

Self Reported Health	Very bad	Bad	Fair	Good	Very good
Probability	0.003	0.041	0.251	0.620	0.085
Female	-0.001	-0.013	-0.040	0.030	0.025
	(0.001)	(0.011)	(0.033)	(0.025)	(0.020)
Education					
incomplete	-0.001	-0.006	-0.019	0.014	0.011
	(0.000)	(0.004)	(0.013)	(0.010)	(0.008)
primary	-0.001	-0.007	-0.022	0.016	0.013
	(0.0006)	(0.0062)	(0.0185)	(0.0140)	(0.0114)
secondary	-0.001	-0.005	-0.016	0.012	0.010
	(0.001)	(0.009)	(0.027)	(0.020)	(0.016)
highschool	-0.002	-0.016	-0.048	0.036	0.029
	(0.0008)	(0.0075)	(0.0225)	(0.0170)	(0.0138)
tertiary	-0.004	-0.035	-0.105	0.079	0.065
	(0.0006)	(0.0045)	(0.0134)	(0.0102)	(0.0082)
Age					
ageG2	0.002	0.020	0.059	-0.044	-0.036
	(0.0002)	(0.0015)	(0.0043)	(0.0033)	(0.0026)
ageG3	0.003	0.031	0.094	-0.071	-0.058
	(0.0004)	(0.0023)	(0.0065)	(0.0051)	(0.0040)
ageG4	0.004	0.041	0.122	-0.092	-0.075
	(0.0008)	(0.0063)	(0.0188)	(0.0143)	(0.0116)
Income					
poor	-0.002	-0.019	-0.057	0.043	0.035
	(0.0006)	(0.0055)	(0.0164)	(0.0124)	(0.0101)
medium	-0.002	-0.016	-0.049	0.037	0.030
	(0.0005)	(0.0049)	(0.0147)	(0.0111)	(0.0090)
rich	-0.002	-0.015	-0.046	0.035	0.028
	(0.0005)	(0.0051)	(0.0154)	(0.0116)	(0.0095)
veryrich	-0.003	-0.028	-0.083	0.063	0.051
	(0.0006)	(0.0051)	(0.0152)	(0.0115)	(0.0094)

Note: The marginal effects implied by Model 2 are presented. Very Good, Good, Fair, Bad and Very Bad are the discrete ordered outcomes corresponding to the Self Assessed Health (SAH) status of the individual.

References

[1] Anderson J., "Regression and Ordered Categorical Variables," *Journal of the Royal Statistical Society*, Series B (Methodological), 1984, 46(1): 30.

[2] Auster, R., I. Leveson and D. Sarachek, "The Production of Health, An Exploratory Study," *The Journal of Human Resources,* 1969, 4:411-436.

[3] Bago d'Uva, T., E. van Doorslaer, M. Lindeboom, O. O'Donnell and S. Chatterji, "Does Reporting Heterogeneity Bias the Measurement of Health Disparities," *Health Economics,* 2008, 17(3): 351-375.

[4] Becker, G. and K.M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy,* 1988, 96(4): 675-700.

[5] Borghans L. and B. H. H. Golsteyn, "Time Discounting and the Body Mass Index: Evidence from the Netherlands," *Economics and Human Biology,* 2006, 4(1): 39-61.

[6] Brown, S., J. Roberts and K. Taylor, "Reservation Wages, Labour Market Participation and Health," *Journal of the Royal Statistical Society,* Series A,2010, 173(3): 501-529.

[7] Contoyannis, P., A.M. Jones and N. Rice, "The Dynamics of Health in the British Household Panel Survey, *Journal of Applied Econometrics*, 2004, 19: 473-503.

[8] Contoyannis, P., A.M. Jones, "Socio-economic Status, Health and Lifestyle," *Journal of Health Economics*, 2004, 23:965-995.

[9] Courtenay, W.H.,"Behavioral factors associated with disease, injury and death among men: Evidence and implications for prevention," *Journal of Men's Stduies*,2000, 9(1): 81-142.

[10] Crosan, R. and U. Greezy, "Gender Differences in Preferences," *Journal of Economic Literature*,2009, 47(2): 448-474.

[11] Crossley, T.F. and S. Keenedy, "The Reliability of Self-Assessed Health Status," *Journal of Health Economics*, 2002, 21(4): 643–658.

[12] Fuchs, V., "Who Shall Live?," 1974, Basic Books: New York.

[13] Fuchs, V., "Time Preference and Health: An Exploratory Study," 1982, In *Economic Aspects of Health*, Fuchs V (ed.). University of Chicago Press: Chicago.

[14] Grossman, M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy,* 1972, 80: 223-255.

[15] Harrison, G. W., M. I. Lau and E. E. Rutström, "Individual Discount Rates and Smoking: Evidence from a Field Experiment in Denmark," *Journal of Health Economics,* 2010, 29: 708–717.

[16] Hernandez-Queveda, C., A.M. Jones and N. Rice N„ "Reporting Bias and Heterogeneity in Self-Assessed Health. Evidence from the British Household Panel Survey," 2005, University of York, HEDG Working Paper, No. 0504.

[17] Humphries, K. and E. van Doorslaer, "Income-related Inequalities in Health in Canada," *Social Science and Medicine,* 2000, 50: 663-671.

[18] Ida, T. and R. Goto, "Interdependency Among Addictive Behaviours and Time/Risk Preferences: Discrete Choice Model Analysis of Smoking, Drinking, and Gambling," *Journal of Economic Psychology,*, 2009, 30: 608–621.

[19] Idler, E.L. and Y. Benyamini, "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies," *Journal of Health and Social Behaviour,*, 1997, 38: 21-37.

[20] Jurges, H., "True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health," *Health Economics,* 2007, 16(2): 163-178.

[21] Khwaja, A., D. Silverman and F. Sloan, "Time Preference, Time Discounting, and Smoking Decisions," *Journal of Health Economics,* 2007, 26(5): 927–949.

[22] Lindeboom, M. and van Doorslaer, E. "Cut-point Shift and Index Shift in Self-reported Health," *Journal of Health Economics,* 2004, 23: 1083-1099.

[23] Kose T., M. A. Soytas, "Does Heterogeneity in Individual Discount Rates Explain the Gender Gap in Self-Assessed Health Status?", 2016, Ozyegin University Working Paper.

[24] Nelson, J.A., "Are Women Really More Risk Averse than Men? A Re-Analysis of the Literature Using Expanded Methods," *Journal of Economic Surveys*, 2014, 29(3): 566-585.

[25] Or, Z., "Determinants of Health Outcomes in Industrialised Countries: A Pooled, Cross-Country, Time-Series Analysis," *OECD Economic Studies*, 2000, 30:54-77.

[26] Rice, N., S. Robone and P. Smith, "Analysis of the Validity of the Vignette Approach to Correct for Heterogeneity in Reporting Health System Responsiveness," *The European Journal of Health Economics*, 2011, 12(2): 41–162.

[27] Ross, C.E. and C.E. Bird, "Sex Stratification and Health Lifestyle: Consequences for Men's and Women's Perceived Health," *Journal of Health and Social Behavior*, 1994, 35: 161–178.

[28] Salomon, J.A., A. Tandon and C.J.L. Murray, "Comparability of Self-Rated Health: Cross Sectional Multi-Country Survey using Anchoring Vignettes," *BMJ*,2004, 328(7434): 258.

[29] Scharff, R.L. and W.K. Viscusi, "Heterogeneous Rates of Time Preference and the Decision to Smoke," *Economic Inquiry*, 2011, 49(4): 959–972.

[30] Schulz, R., M. Mittelmark, R. Kronmal, J.F. Polak, C.H. Hirsch, P. German and J. Beakwala, "Predictors of Perceived Health Status in Elderly Men and Women," *Journal of Aging and Health*, 1994, 6(4): 419–447.

[31] Svedberg, P., P. Lichtenstein and N.L. Pedersen, "Age and Sex Differences in Genetic and Environmental Factors for Self-Rated Health: A Twin Study," *Journal of Gerontology*,2001, 56B: 171-178.

[32] Terza, J., "Ordered Probit: A Generalization," Communications in Statistics - A. Theory and Methods, 1985, 14: 1-11.

[33] Trannoy, A., S. Tubeuf, F. Jusot and M. Devaux, "Inequality of Opportunities in Health in France: A First Pass," *Health Economics*, 2010, 19(8): 921-938.

[34] Van Doorslaer, E., A. Wagstaff and H. Bleichrodt, "Income-Related Inequalities in Health: Some International Comparisons," *Journal of Health Economics*, 1997, 16: 93-112.

[35] Waldron, I., " What do we Know about Causes of Sex Differences in Mortality? A Review of the Literature," 1997, In *The Sociology of Health and Illness*, Canrad P (ed.). New York: St. Martin's Press.

[36] Zabalza, A., "A Note on the Estimation of Subjective Rates of Discount from Labour Supply Functions," *Economica* New Series, 1979, 46(182): 197-202.