CHEATING THE GOVERNMENT: DOES TAXPAYER PERCEPTION MATTER?

*Daniel Jeong-Dae Lee**

Do people cheat because they can get away with it or because they feel that the rules are unfair? This paper addresses this question in the context of tax evasion. Specifically, taxpayer perception is incorporated into a widely used consumption-based method for estimating income tax evasion. Compared to the standard method, which distinguishes taxpayers only by their occupational or income type as a way of measuring their "ability" to misreport income, the refined method introduces taxpayers who may be "able but unwilling" to cheat because they feel fairly treated with respect to public services and as compared to other taxpayers. Applied to a longitudinal data for the Republic of Korea (2007–2015), the standard method yields a uniform tax evasion rate of 13 per cent, but the refined method provides a range of 7 to 25 per cent based on taxpayer perception. This implies that strategies for improving tax compliance must be tailored to different motivations for tax evasion.

---

* School of Development Economics, National Institute of Development Administration, Bangkok, Thailand (email: lee75@un.org).

I. INTRODUCTION

Tax evasion is as old as taxes themselves. It is a way of cheating the government (Cowell, 1990), with adverse effects on social welfare.[1] Assessing the extent and determinants of tax evasion remains a challenge given its hidden nature.

In the economic literature, the benchmark model posits that people cheat if the probability of being caught and punished is low compared to potential monetary savings (Becker, 1968; Allingham and Sandmo, 1972). A popular comparison is between wage earners who are "unable" to cheat because their incomes are subject to withholding or third-party reporting, versus the self-employed who are "able" to cheat because their incomes are harder to detect by tax authorities (Slemrod, 2007; Kleven and others, 2011).

Tax authorities are certainly concerned about enforcement, but they also emphasize "tax morale," which is generally assumed to mean increasing voluntary compliance with tax laws and creating a social norm of compliance (Luttmer and Singhal, 2014). Behavioural models in economics also suggest that tax decisions are influenced by perceptions of fairness and other factors that fall outside the standard, expected utility framework (Bordignon, 1993; Feld and Frey, 2002; Torgler, 2003, among other studies). Not accounting for such factors could result in overattributing evasion to the lack of heavy enforcement.

To contribute to this debate on whether taxpayers are "unable" or "unwilling" to cheat, the following hypothesis is examined: *Ability to cheat matters only if one is willing to cheat*. If this were true, taxpayers with identical occupational or income characteristics might exhibit very different compliance behaviours.

To empirically test this hypothesis in the absence of reliable data on tax evasion, the consumption-based method for estimating income tax evasion is extended. This method was pioneered by Pissarides and Weber (1989) and has been widely applied by (Lyssiotou, Pashardes and Stengos, 2004; Johansson, 2005; Hurst, Li and Pugsley, 2014; Kim, Gibson and Chung, 2017; Kukk and Staehr, 2017; Engström and Hagen, 2017, among other studies). It is based on measuring "excess" consumption among the self-employed as evidence of undeclared income, compared to the consumption of wage earners, which serves as the benchmark.

---

[1]   If widespread, tax evasion could constrain the provision of necessary public services. In addition, if higher tax rates were applied to a narrow tax base to make up for the revenue shortfall because of evasion, there could be adverse welfare effects. Moreover, if concentrated among certain segments of society, tax evasion could undermine trust and social cohesion.

The contribution from the present paper is to introduce another layer of heterogeneity among taxpayers based on their perception of fairness with respect to public services and compared to other taxpayers. Applying the refined method to longitudinal household survey data for the Republic of Korea, a higher rate of tax evasion is estimated for the self-employed who feel unfairly treated, while a lower rate is found for those who feel fairly treated, compared to what is predicted using the standard method. The Republic of Korea is used as an illustration given its high degree of self-employment relative to income level, but the refined method is replicable in other countries.

This paper is organized as follows. In section II, the conceptual framework is laid out. Section III contains an explanation of the consumption-based method for estimating income tax evasion. Section IV includes a discussion of the data used. Section V provides an overview of the estimation results and some robustness checks. Section VI concludes.

II. CONCEPTUAL FRAMEWORK

In this section, the conditions for tax evasion are examined, as shown in the benchmark model in the literature and a simple extension, which has been proposed to incorporate taxpayer perception. This is followed by the presentation of a problem tree approach to determine which taxpayers are both "able" and "willing" to cheat.

Conditions for tax evasion

The standard framework for considering a taxpayer's choice of whether and how much to misreport income is a deterrence model first formulated by Allingham and Sandmo (1972), who adapted the model of the economics of crime of Becker (1968). Under this model, people cheat if the probability of being caught and punished is low compared to potential tax savings. The expected utility function is given as:

$$E(U) = (1 - p)U(W - \theta X) + pU(W - \theta X - \pi(W - X)) \tag{1}$$

where W and X are the actual income and the reported income, respectively. Tax is levied at a constant rate θ on X. Importantly, the two states are separated by p, the probability of detection; and π is the penalty rate. Letting Y and Z stand for the net income without and with detection, respectively, the first-order condition for an interior solution gives $\frac{U'(Z)}{U'(Y)} = \frac{(1-p)\theta}{p(\pi-\theta)}$, which implies that a higher probability of detection discourages tax evasion. Taking the derivative of the expected utility when $W = X$, the condition for evasion is $p\pi < \theta$, namely, the ability to escape detection matters.

To this model, a parameter, which is applied regardless of detection, is added to account for factors other than the threat of punishment, which may influence taxpayer behaviour. Specifically, taxpayers may be less inclined to cheat if they view the tax system as fair and are satisfied with the public services they receive. In contrast, if they feel that they are unfairly treated, they may perceive some tax evasion is justified (Bordignon, 1993; Barth, Cappelen and Ognedal, 2013). The expected utility function becomes:

$$E(U) = (1 - p)U(Y) + pU(Z) - C(E - E^*) \qquad (2)$$

where the additional parameter is a function of evaded income, $E = W - X$, and what the taxpayer perceives to be "justified" evasion, $E^* \geq 0$. In equation 2, the condition for tax evasion not only depends on the probability of detection, but also on taxpayer perception, such that optimal evasion is lower than what is predicted under equation 1 if $E > E^*$, but higher if $E < E^*$.

Problem tree approach

As decisions related to tax evasion become more complex, a problem tree approach is suitable to determine tax evasion outcomes based on key parameters contained in equation 2. The steps are shown in figure 1.

Figure 1. Grouping taxpayers by income and perception
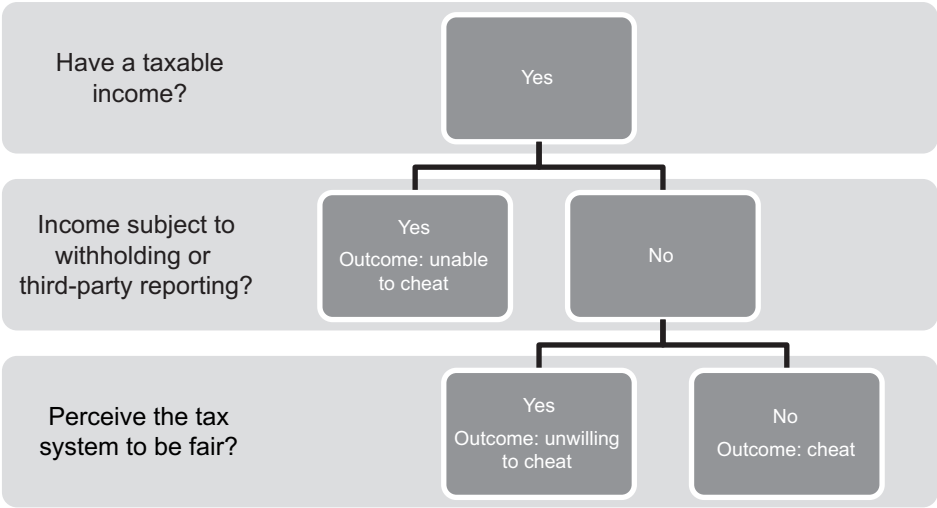
Based on the restricted sample of households with a taxable income, taxpayers are differentiated by whether their income is subject to withholding or third-party reporting (W_T) or simply self-reported (W_S). The probability of detection would be close to 1 if the taxpayer were to declare an amount below W_T as such discrepancies are easy to detect, whereas reporting below W_S may go undetected.[2] As taxpayers typically draw on various types of income, a certain threshold on the ratio $\dfrac{W_S}{W_S + W_T}$ is applied to separate those who are "able" versus "unable" to cheat.

In practice, employees who rely mostly on labour income fall under the former, while the self-employed who depend on business income are under the latter. Whether capital income from interest, dividends and rent qualifies as W_T or W_S depends on the extent to which the country's financial sector is subject to third-party reporting, including at the international level, without which offshore accounts could facilitate tax evasion.

After the sample is restricted to taxpayers that are "able" to cheat, they are differentiated based on their perception of fairness, which is related to the "justified" tax evasion parameter E^* in equation 2. Accurately measuring taxpayer perception is challenging. For instance, for the World Values Surveys, respondents are asked whether tax evasion is "never, sometimes, or always justified", but it is difficult to confirm whether they would act in such a manner.[3] In other cases, respondents are asked whether they would fully declare their income if misreporting were to go undetected, but it is questionable why anyone would reveal such information if there is even a small chance that it might be used against them.

An alternative approach is to construct a latent variable based on taxpayer responses to questions, such as the following: How does one's tax burden compare to others' earning similar income levels?; Are the wealthy paying enough taxes?; and Is the quality of public services commensurate with your tax payment? Given that "fairness" is likely to reflect multiple dimensions, such as horizontal and vertical equity and reciprocity, structural equation modelling is used to pull these dimensions into a latent variable and then apply a certain threshold to distinguish between those "unwilling" and those "willing" to cheat.[4] Such grouping of taxpayers by perception along with income

---

[2]　Currently, many countries make extensive use of withholding and third-party reporting, under which the audit rate alone is a poor proxy for the probability of detection. For instance, firms remit the majority of tax revenue to the government, including through withholding taxes owed by employees. The financial sector, including banks, insurers and pension funds, also reports taxable income earned by individuals to the government. Such information could be compared to what is declared by the taxpayer. See also Kleven and others (2011).

[3]　World Values Surveys have been used in empirical studies on tax morale, such as Torgler (2003).

[4]　Structural equation modelling has emerged as a useful tool in other social sciences, including education and psychology. In the economic literature, Schneider and Enste (2000) applied structural equation modelling to estimate the extent of the shadow economy based on country-level variables.

serves as a basis for incorporating taxpayer perception into the estimation of tax evasion, as discussed in the following section.
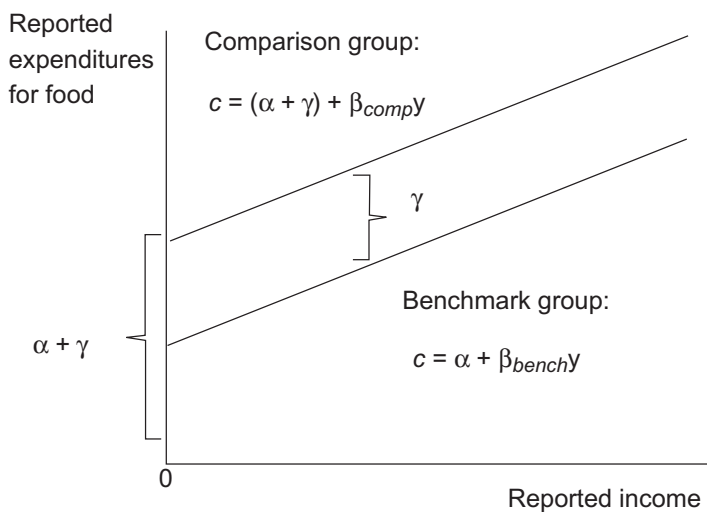
III. CONSUMPTION-BASED METHOD

This section provides an explanation of the consumption-based method for estimating income tax evasion, which has been widely applied in the literature (Pissarides and Weber, 1989; Lyssiotou, Pashardes and Stengos, 2004; Johansson, 2005; Hurst, Li and Pugsley, 2014; Kim, Gibson and Chung, 2017; Kukk and Staehr, 2017; Engström and Hagen, 2017, among other studies). Compared to other methods, such as the currency demand approach, this one is based on household income and expenditure surveys, which are widely available in most countries and can be used to incorporate micro-level information, such as taxpayer perception. It is also more replicable compared to special audit programmes or experiments with actual taxpayers that have been deployed in few countries (see, for instance, United States, Internal Revenue Service, 2016; and Kleven and others, 2011), but could be costly and raise issues of legality in other countries.

Basic approach and assumptions

In figure 2, the consumption-based method is illustrated. The figure shows two log-linear Engel curves, one for wage earners (benchmark group) and the other for the self-employed (comparison group). Letting c stand for log food expenditures and y log

Figure 2. Engel curve showing consumption-income relationship

Reported expenditures for food

Comparison group:

$c = (\alpha + \gamma) + \beta_{comp}y$

γ

Benchmark group:

$c = \alpha + \beta_{bench}y$

$\alpha + \gamma$

0

Reported income

disposable income, β denotes the elasticity of consumption with respect to income. The intercept difference γ between the two curves measures the "excess consumption" of the self-employed, which is used to estimate the undeclared fraction of income: 1 – exp $(-\gamma/\beta)$.

This method relies on three central assumptions. First, the elasticity of consumption with respect to income, β, is equal for the two groups, as illustrated by the curves having the same slope. Second, neither group systemically misreports spending on certain items, such as food or non-durables. Food is widely used in the literature, as it is mundane enough for individuals not to be afraid of reporting truthfully and purchased regularly, and accordingly, less subject to unintended misreporting. An exception to this is Lyssiotou, Pashardes and Stengos (2004), who use non-durable expenditures to account for demand heterogeneity. Third, wage earners report their true incomes, while the self-employed systematically underreport their income by a constant factor.[5] One additional assumption is that people misreport their income in surveys to the same degree that they misreport it to the tax authorities (Hurst, Li and Pugsley, 2014).

Accounting for transitory income

Given that household consumption is not influenced by current income, but instead by a more permanent measure of income, estimating the Engel curve using the current income would result in a measurement error. Specifically, transitory income fluctuations would attenuate the estimate of the income elasticity (Wooldridge, 2009), which, in turn, would result in an overestimation of income misreporting among the self-employed. For household i, let c_i stand for food expenditure, y_i^p the permanent income, and X_i a vector of variables affecting consumption. D_{it} is a dummy taking the value 1 for the self-employed. A log-linear Engel curve is given as:

$$lnc_i = \gamma D_{it} + \beta lny_i^P + \theta X_i + \varepsilon_i \tag{3}$$

Let g_i stand for a random variable showing the degree of transitory income variation such that $y_i = g_i y_i^P$. The mean of g_i is assumed to be the same for both groups, but the variance may differ. Substitution gives an Engel curve in terms of the current income, with the transitory element pushed to the error term:

$$lnc_i = \gamma D_{it} + \beta lny_i + \theta X_i + (\varepsilon_i - \beta lng) \tag{4}$$

---

[5] If wage earners also underreport their income, the method will only provide a lower bound estimate of tax evasion.

Pissarides and Weber (1989) and other early studies dealt with this through instrumental variable (IV) techniques, but more recent studies (Hurst, Li and Pugsley, 2014; Kim, Gibson and Chung, 2017; Engström and Hagen, 2017) have exploited panel data and constructed multi-year average income measures. In this paper, the latter approach is followed. With average income, positive and negative variations of transitory income cancel each other over time, especially for the self-employed, who typically have higher income variation from year to year. This also implies that the covariance between the degree of underreporting and the degree of transitory income variation disappears over time. The Engel curve is now written as:

$$lnc_{it} = \alpha + \gamma D_{it} + \beta \overline{lny_{it}} + \theta X_{it} + \mu_t + \varepsilon_{it} \tag{5}$$

where subscript t denotes year, $\overline{lny_{it}}$ is the mean value of reported incomes over time for the same household i, μ_t is the time effect, and ε_{it} the cumulative effects of unobserved determinants over time.

Incorporating taxpayer perception

In previous studies related to this subject, income tax evasion was estimated based on a single comparison group, that is, the self-employed. However, as discussed in section II, such an approach could overlook factors other than tax savings from misreporting, which may influence taxpayer behaviour. To test the hypothesis that *ability matters only if one is willing to cheat*, one could take the interaction term approach. Let $D_{A\,it}$ and $D_{W\,it}$ denote dummy variables taking the value 1 for the "able" and the "willing" to cheat, respectively, and $D_{W\,it} * D_{A\,it}$ be the interaction term. The log-linear Engel curve for estimation becomes:

$$lnc_{it} = \alpha + \gamma D_{A\,it} + \delta D_{W\,it} + \omega D_{W\,it} * D_{A\,it} + \beta \overline{lny_{it}} + \theta X_{it} + \mu_t + \varepsilon_{it} \tag{6}$$

where the parameter of interest is ω. With the interaction term, the partial effect of ability to cheat on consumption would be $\gamma + \omega D_{W\,it}$.

One question, however, is whether the benchmark group should be the "unable and unwilling" to cheat as in equation 6 or simply the "unable" to cheat. In the former case, the "unable" to also cheat should be allowed in a way that is inconsistent with the main assumptions discussed earlier. Accordingly, in the present case, two comparison groups are explicitly introduced, "able but unwilling" and "able and willing" to cheat, as shown by $D_{AUW\,it}$ and $D_{AW\,it}$, respectively:

$$lnc_{it} = \alpha + \gamma D_{AW\,it} + \delta D_{AUW\,it} + \beta \overline{lny_{it}} + \theta X_{it} + \mu_t + \varepsilon_{it} \tag{7}$$

If the concerned hypothesis were true, δ would be insignificant, namely the "able but unwilling" to cheat would exhibit a similar behaviour to the benchmark group, those that are unable to cheat.

IV. DATA AND SAMPLE RESTRICTION

Compared to other members of the Organization for Economic Coordination and Development (OECD), the Republic of Korea has a relatively high degree of self-employment, which, accordingly, presents a relevant case for testing this method. The panel data of the Republic of Korea, drawn from the National Survey on Tax and Benefit, contains information on approximately 5,000 households across nine years (2007–2015), including the occupations, incomes, assets, expenditures, taxes and social security, as well as gender, age and educational attainment of family members. Importantly, the most recent round of the survey for 2015 contains questions on how the respondents perceive the tax system and the factors that affect their taxpaying decision.

Following the steps outlined in figure 2, households with income levels below the exemption threshold and households with zero annual expenditure are eliminated. To maintain some degree of homogeneity, only households whose primary income earner is between the ages of 20 and 70 are retained.

The next step is to differentiate taxpayers by their occupational or income type. Following previous studies, the "able" to cheat is defined as households that derive at least 25 per cent of their total income from business income, and the "unable" to cheat are households that attain less than 1 per cent of their total income from business income.

Then, a latent variable "willing to cheat" is constructed based on responses to the questions regarding vertical and horizontal equity and satisfaction with public services, discussed in section II. The responses are scaled from 1 to 5; the scale is reversed when necessary for consistency. These variables are standardized and combined into a latent variable using structural equation modelling. As shown in figure 3, the latent variable "willing" to cheat is most closely associated with horizontal equity and public services. The top and bottom quartiles, according to this measure, are defined as the "unwilling" and "willing" to cheat, respectively, and the two middle quartiles are dropped from the sample.

In accordance with previous studies, the following control variables are considered:

(a) *Family size*: more mouths to feed or clothe, thus likely to have a positive sign.

(b) *Capital city dummy*: price levels are higher, thus likely to have a positive sign.

(c) *Home ownership*: without the need for rental fees, general spending may be higher, thus likely to have a positive sign.

(d) *Age of primary income earner*: likely to spend more with age, but at a diminishing rate; thus, it is likely to have a positive sign, but the quadratic would have a negative sign.

Figure 3. Constructing a latent variable based on taxpayer perception
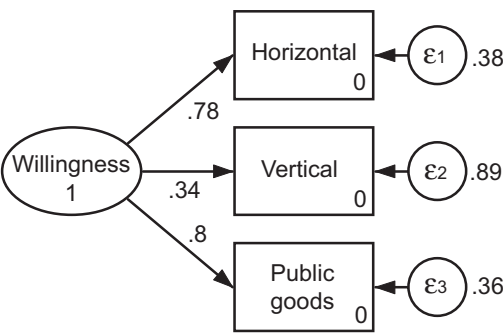
Note: The circle denotes the unobserved latent variable, and the squares denote the observed variables, which are based on responses to questions regarding horizontal and vertical equity and satisfaction with public goods and services. The latent variable is constructed based on a maximum likelihood estimation.

(e) *Education of primary income earner*: higher educational attainment, through higher income, may result in more spending, thus a positive sign.

(f) *Hours worked*: working more hours may increase income and spending, but long working hours may also indicate low wages; thus, the expected sign is unclear.

Table 1 presents the descriptive statistics for the benchmark group "unable" to cheat (17,125 observations), and the two comparison groups "able, but unwilling" (4,474 observations) and "able and willing" to cheat (2,142 observations). Approximately 25 per cent of the sample is self-employed (defined as respondents whose business income share is 25 per cent or above),[6] and the rest of the sample consists of wage earners (respondents whose business income share is 1 per cent or below). Those with business income share above 1 per cent but below 25 per cent are dropped. Based on summary statistics, there are no marked differences across the three groups, but the "able and willing" to cheat tend to be located in the capital city and work longer hours, perhaps to meet the higher cost of living.

---

[6] The 25 per cent threshold is applied in Pissarides and Weber (1989) and subsequent studies. A robustness test with respect to this assumption by also applying slightly lower or higher thresholds does not change the main findings of the present paper.

Table 1. Descriptive statistics

	Unable (wage earner)	Able but unwilling (self-employed)	Able and willing (self-employed)
Log of income	8.5582 (0.4357)	8.4850 (0.4160)	8.5793 (0.4733)
Log of food expenditures	6.7493 (0.4911)	6.6891 (0.5162)	6.8673 (0.5584)
Family size	3.8102 (1.0054)	3.8355 (1.1314)	3.7005 (0.9031)
Capital city	0.2071 (0.4053)	0.2047 (0.4035)	0.3086 (0.4620)
Age	47.0318 (9.1368)	47.7295 (9.0039)	47.1328 (8.9109)
Education	2.4918 (0.6469)	2.3596 (0.6932)	2.5125 (0.6220)
Hours worked	42.1849 (8.9804)	46.3978 (15.5131)	49.1839 (13.0722)
Observations	17 125	4 475	2 142

Note: Income is the three-year average income. Capital city dummy denotes 1 if the household is based in Seoul and 0 otherwise. Age and education refer to those of the household head, or the primary income earner. Educational attainment is on a scale of 1 to 3, where 1 is lower-secondary or below, 2 is upper-secondary and 3 is tertiary. Working hour is the average per week.

V. ESTIMATION RESULTS

This section contains a report of the main estimation results and provides some robustness checks. Full results are shown in annex I.

Main results – taxpayer perception in the food Engel curve

Log food consumption is regressed on log disposable income and a set of control variables, as in equations 5 and 7. Table 2 shows the estimates of key parameters, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$, and the corresponding estimated amount of income underreporting, denoted as $1 - \hat{k}$. There are 23,946 observations. The estimated elasticity of consumption with respect to income, $\hat{\beta}$, is stable across different specifications, with a 1 per cent increase in income resulting in an increase of just under 0.4 per cent in food expenditure based on the ordinary least squares (OLS) estimator.

Compared to the "unable" to cheat, the "able" to cheat consume about 6 per cent more, as shown in column 2. The estimated tax evasion rate is 13 per cent assuming that people cheat if they can. However, given that some taxpayers may be "able but unwilling" to cheat, taxpayer perception is incorporated in column 3 based on the latent variable approach. While $\hat{\gamma}$ is positive and statistically significant, it is relatively small compared to $\hat{\delta}$ for the "able and willing" to cheat who consumes 11 per cent more than the benchmark group. This translates into a tax evasion rate of 25 per cent, or double that of the "average" evasion rate for the "able" to cheat.

Columns 4 and 5 show the instrumental variable (IV) results. In accordance with Hurst, Li and Pugsley (2014), education is used as instrument for income, as it would affect food consumption only through the income effect.[7] While it passes the endogeneity tests,[8] $\hat{\beta}$ is implausibly large. This is similar to the findings of Engström and Hagen (2017), who compare the performance of several instruments, including education, housing and capital income, and conclude that the use of the IV approach is inferior to a simple OLS based on multi-year average income. Given the large $\hat{\beta}$, the estimated tax evasion rates are lower than when using OLS. Nevertheless, columns 4 and 5 exhibit the same trend as columns 2 and 3.

Full estimation results in annex I show that the control variables have the expected sign. While family size, capital city dummy, age and education are statistically significant, home ownership dummy and hours worked are not. The year dummies are included with 2008 as the base year; the first and last years in the sample, 2007 and 2015, do not appear because of the use of three-year average income.

To determine whether taxpayer perception emerges as a significant factor under other specifications, two robustness checks are conducted.

Taxpayer perception under demand heterogeneity

Following Lyssiotou, Pashardes and Stengos (2004), to account for demand heterogeneity among households, estimation results with log of non-durable expenditures as a dependent variable are shown in table 3. $\hat{\beta}$ is stable across different specifications, with a 1 per cent increase in income resulting in an increase of approximately 0.3 per cent in non-durable expenditures based on the OLS estimator.

---

[7]   While used in the literature, education is not liked to be a perfect instrument. For instance, education may affect how people are conscious about their nutrition and accordingly their food consumption pattern.

[8]   The null hypothesis of the Durbin and Wu-Hausman tests is that the variable under consideration can be treated as exogenous. If the test statistics are highly significant, the null can be rejected and treated as endogenous.

Table 2. Estimation results: log of food expenditures as dependent variable

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0554*** (0.00582)		0.0922*** (0.00698)	
able but unwilling $\hat{\gamma}$			0.0261*** (0.00680)		0.0819*** (0.00865)
able and willing $\hat{\delta}$			0.113*** (0.00851)		0.111*** (0.00928)
log(income) $\hat{\beta}$	0.387*** (0.007)	0.393*** (0.00734)	0.387*** (0.00737)	0.866*** (0.0295)	0.858*** (0.0305)
Observations	23 946	23 946	23 946	23 946	23 946
R-squared	0.282	0.284	0.287	0.144	0.149

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. OLS, ordinary least squares; IV, instrumental variable.

	$\hat{k}$	$1 - \hat{k}$
OLS		
able	0.869	0.131
able but unwilling	0.935	0.065
able and willing	0.747	0.253
IV		
able	0.899	0.101
able but unwilling	0.909	0.091
able and willing	0.879	0.121

Note: $\hat{k}$ is the general term for the fraction of the true income, which is declared.

Compared to the "unable" to cheat, the "able" to cheat consume approximately 7 per cent more, as shown in column 2. This gives an "average" tax evasion rate of 19 per cent.

With taxpayer perception in column 3, essentially the same trend is established as in the case of the food Engel curve. Compared to the "unable" to cheat, the "able but unwilling" consume approximately 4 per cent more. On the other end of the spectrum, the "able and willing" consume approximately 13 per cent more compared to the same benchmark group. This gives an estimated tax evasion rate of 33 per cent, three times than that of the "able but unwilling."

Columns 4 and 5 show the IV results. Again, $\hat{\beta}$ is relatively large but within a more reasonable range. It may be because education is a better instrument for income in the case of non-durable expenditures, compared to food expenditures. For instance, households with more educated parents are likely to spend more on children's education, although they may not necessarily spend more on food. The estimated tax evasion rate is slightly lower than in the case of OLS, at 14 per cent, on average, and with a range of 11 to 20 per cent, depending on taxpayer perception.

Table 3. Estimation results: log of non-durable expenditures as dependent variable

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0693***		0.0909***	
		(0.00475)		(0.00543)	
able but unwilling $\hat{\gamma}$			0.0381***		0.0700***
			(0.00552)		(0.00671)
able and willing $\hat{\delta}$			0.131***		0.130***
			(0.00697)		(0.00722)
log(income) $\hat{\beta}$	0.321***	0.329***	0.323***	0.608***	0.593***
	(0.006)	(0.00588)	(0.00588)	(0.0231)	(0.0239)
Observations	23 913	23 913	23 913	23 913	23 913
R-squared	0.318	0.324	0.328	0.254	0.263

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. OLS, ordinary least squares; IV, instrumental variable.

	$\hat{k}$	$1-\hat{k}$
OLS		
able	0.810	0.190
able but unwilling	0.889	0.111
able and willing	0.667	0.333
IV		
able	0.861	0.139
able but unwilling	0.889	0.111
able and willing	0.803	0.197

Note: $\hat{k}$ is the general term for the fraction of the true income which is declared.

Taxpayer perception based on an alternative measure

As noted in section II, there are different ways to measure taxpayer perception. Compared to the main result, which is based on the latent variable approach, here an alternative measure based on the following hypothetical situation is used: "If you have rent income which the tax authorities cannot observe, would you declare the full amount?" Those who indicate "Yes, all of the income" are defined as "unwilling" to cheat, while those who indicate "There is no need to declare if the amount is small" or "Declare none in any case" are defined as the "willing" to cheat. The sample size is almost double as compared to earlier specifications, as no observations are dropped to define willingness.

Table 4. Estimation results: alternative measure of willingness to cheat

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0462*** (0.00408)		0.0866*** (0.00492)	
able but unwilling $\hat{\gamma}$			0.0390*** (0.00539)		0.0745*** (0.00613)
able and willing $\hat{\delta}$			0.0518*** (0.00524)		0.0938*** (0.00609)
log(income) $\hat{\beta}$	0.372*** (0.005)	0.377*** (0.00509)	0.377*** (0.00509)	0.839*** (0.0211)	0.840*** (0.0211)
Observations	49 300	49 300	49 300	49 300	49 300
R-squared	0.284	0.286	0.286	0.150	0.150

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. OLS, ordinary least squares; IV, instrumental variable.

	$\hat{k}$	$1 - \hat{k}$
OLS		
able	0.885	0.115
able but unwilling	0.902	0.098
able and willing	0.872	0.128
IV		
able	0.902	0.098
able but unwilling	0.915	0.085
able and willing	0.894	0.106

Note: $\hat{k}$ is the general term for the fraction of the true income which is declared.

Table 4 shows the estimation results for food Engel curve. $\hat{\beta}$, is stable across different specifications, with a 1 per cent increase in income resulting in just below a 0.4 per cent increase in food expenditures based on the OLS estimator. With taxpayer perception in column 3, essentially the same trend is established as in the latent variable approach, although the willingness to cheat now explains a much smaller portion of the overall variation. The difference between the "unable" and the "able" to cheat is much more significant compared to perception differences within the "able" to cheat. This seems to confirm the suspicion why anyone would reveal information which might be used against them.

XI. CONCLUSION

In this paper, the question of why people cheat in the context of income tax evasion is examined. The standard expected utility model of tax evasion posits that people cheat if the chances of getting caught and punished are quite low. Given that this information asymmetry between taxpayer and the tax authority plays a key role, many countries have introduced third-party reporting schemes to complement the traditional audit in increasing the probability of detection. Such efforts have improved tax compliance in many cases, but only up to a certain extent. The remaining "residual" may be more difficult to explain and requires a deeper understanding of the human behaviour.

The possibility that people are "able" and yet "unwilling" to cheat the government because they feel fairly treated is explored in this paper. While this is a plausible argument, empirically testing it is not so straightforward. Using the consumption-based method for estimating income tax evasion, this paper provides a way to tackle the issue. Based on an illustrative case of the Republic of Korea, taxpayer perception matters, and in some cases monetary returns may play a relatively minor role in determining tax evasion, unlike in the case of the standard, expected utility model.

The policy implication is that, for curbing tax evasion, voluntary compliance measures and appropriate changes to a tax law may be just as necessary as third-party reporting and other enforcement measures. An optimal strategy for improving compliance may be to target the extensive and the intensive margins: foster social and cultural norms for compliance, while making cheating more difficult for those who are persistently inclined to cheat.

ANNEX I. FULL ESTIMATION RESULTS

Table A.1. Estimation results: log of food expenditures as dependent variable

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0554*** (0.00582)		0.0922*** (0.00698)	
able but unwilling			0.0261*** (0.00680)		0.0819*** (0.00865)
able and willing			0.113*** (0.00851)		0.111*** (0.00928)
log(income)	0.387*** (0.007)	0.393*** (0.00734)	0.387*** (0.00737)	0.866*** (0.0295)	0.858*** (0.0305)
family size	0.076*** (0.004)	0.0761*** (0.00357)	0.0766*** (0.00355)	0.0543*** (0.00383)	0.0547*** (0.00384)
capital city	0.215*** (0.007)	0.210*** (0.00690)	0.206*** (0.00688)	0.179*** (0.00818)	0.178*** (0.00809)
home ownership	0.010* (0.006)	0.0104* (0.00591)	0.0101* (0.00590)	-0.0372*** (0.00704)	-0.0368*** (0.00705)
age	0.034*** (0.003)	0.0327*** (0.00312)	0.0332*** (0.00314)	0.00620 (0.00378)	0.00669* (0.00382)
age squared	-0.000*** (0.000)	-0.000335*** (3.34e-05)	-0.000338*** (3.35e-05)	-0.000105*** (3.94e-05)	-0.000108*** (3.96e-05)
education	0.081*** (0.005)	0.0855*** (0.00512)	0.0825*** (0.00513)		
hours worked	-0.001*** (0.000)	-0.00160*** (0.000274)	-0.00174*** (0.000274)	-0.00107*** (0.000301)	-0.00112*** (0.000304)
2009.year	0.000 (0.011)	-2.97e-05 (0.0110)	-0.000688 (0.0110)	-0.0128 (0.0120)	-0.0129 (0.0120)
2010.year	0.069*** (0.011)	0.0686*** (0.0108)	0.0686*** (0.0108)	0.0486*** (0.0120)	0.0489*** (0.0120)
2011.year	0.066*** (0.011)	0.0661*** (0.0108)	0.0659*** (0.0108)	0.0162 (0.0123)	0.0168 (0.0123)
2012.year	0.168*** (0.011)	0.169*** (0.0108)	0.169*** (0.0108)	0.120*** (0.0123)	0.121*** (0.0123)
2013.year	0.160*** (0.011)	0.163*** (0.0108)	0.162*** (0.0108)	0.106*** (0.0125)	0.107*** (0.0125)
2014.year	0.197*** (0.011)	0.199*** (0.0109)	0.199*** (0.0109)	0.134*** (0.0127)	0.135*** (0.0127)
Constant	2.060*** (0.084)	2.028*** (0.0835)	2.074*** (0.0836)	-0.965*** (0.212)	-0.913*** (0.219)
Observations	23 946	23 946	23 946	23 946	23 946
R-squared	0.282	0.284	0.287	0.144	0.149

Note: OLS, ordinary least squares; IV, instrumental variable.

**Table A.2. Estimation results: log of non-durable expenditures
as dependent variable**

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0693*** (0.00475)		0.0909*** (0.00543)	
able but unwilling			0.0381*** (0.00552)		0.0700*** (0.00671)
able and willing			0.131*** (0.00697)		0.130*** (0.00722)
log(income)	0.321*** (0.006)	0.329*** (0.00588)	0.323*** (0.00588)	0.608*** (0.0231)	0.593*** (0.0239)
family size	0.092*** (0.003)	0.0914*** (0.00281)	0.0919*** (0.00279)	0.0785*** (0.00298)	0.0794*** (0.00298)
capital city	0.183*** (0.006)	0.178*** (0.00565)	0.173*** (0.00561)	0.159*** (0.00633)	0.157*** (0.00624)
home ownership	-0.010** (0.005)	-0.00960* (0.00491)	-0.00998** (0.00489)	-0.0377*** (0.00559)	-0.0368*** (0.00558)
age	0.048*** (0.003)	0.0471*** (0.00262)	0.0477*** (0.00263)	0.0315*** (0.00300)	0.0325*** (0.00303)
age squared	-0.000*** (0.000)	-0.000466*** (2.80e-05)	-0.000469*** (2.81e-05)	-0.000329*** (3.14e-05)	-0.000337*** (3.15e-05)
education	0.045*** (0.004)	0.0505*** (0.00415)	0.0472*** (0.00415)		
hours worked	-0.000 (0.000)	-0.000763*** (0.000231)	-0.000907*** (0.000230)	-0.000453* (0.000245)	-0.000558** (0.000245)
2009.year	0.006 (0.009)	0.00564 (0.00885)	0.00487 (0.00883)	-0.00172 (0.00938)	-0.00191 (0.00932)
2010.year	0.055*** (0.009)	0.0548*** (0.00874)	0.0547*** (0.00873)	0.0432*** (0.00937)	0.0436*** (0.00931)
2011.year	0.074*** (0.009)	0.0741*** (0.00875)	0.0738*** (0.00874)	0.0449*** (0.00958)	0.0459*** (0.00954)
2012.year	0.133*** (0.009)	0.134*** (0.00874)	0.134*** (0.00872)	0.105*** (0.00962)	0.106*** (0.00957)
2013.year	0.130*** (0.009)	0.133*** (0.00873)	0.133*** (0.00872)	0.0998*** (0.00974)	0.101*** (0.00971)
2014.year	0.151*** (0.009)	0.154*** (0.00885)	0.154*** (0.00884)	0.116*** (0.00997)	0.117*** (0.00995)
Constant	2.701*** (0.071)	2.661*** (0.0704)	2.710*** (0.0702)	0.895*** (0.168)	1.000*** (0.172)
Observations	23 913	23 913	23 913	23 913	23 913
R-squared	0.318	0.324	0.328	0.254	0.263

Note: OLS, ordinary least squares; IV, instrumental variable.

Table A.3. Estimation results: alternative measure of willingness to cheat

	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) IV
able		0.0462*** (0.00408)		0.0866*** (0.00492)	
able but unwilling			0.0390*** (0.00539)		0.0745*** (0.00613)
able and willing			0.0518*** (0.00524)		0.0938*** (0.00609)
log(income)	0.372*** (0.005)	0.377*** (0.00509)	0.377*** (0.00509)	0.839*** (0.0211)	0.840*** (0.0211)
family size	0.079*** (0.002)	0.0792*** (0.00229)	0.0790*** (0.00229)	0.0522*** (0.00258)	0.0519*** (0.00259)
capital city	0.187*** (0.005)	0.186*** (0.00475)	0.187*** (0.00475)	0.143*** (0.00588)	0.143*** (0.00588)
home ownership	0.007 (0.004)	0.00685* (0.00410)	0.00686* (0.00411)	-0.0366*** (0.00487)	-0.0368*** (0.00487)
age	0.041*** (0.002)	0.0392*** (0.00219)	0.0395*** (0.00218)	0.0172*** (0.00262)	0.0177*** (0.00260)
age squared	-0.000*** (0.000)	-0.000415*** (2.33e-05)	-0.000417*** (2.32e-05)	-0.000224*** (2.73e-05)	-0.000227*** (2.72e-05)
education	0.078*** (0.004)	0.0810*** (0.00356)	0.0812*** (0.00356)		
hours worked	-0.000 (0.000)	-0.000526*** (0.000181)	-0.000520*** (0.000181)	0.000459** (0.000207)	0.000486** (0.000208)
2009.year	0.013* (0.007)	0.0124* (0.00748)	0.0124* (0.00748)	0.000763 (0.00815)	0.000746 (0.00815)
2010.year	0.068*** (0.007)	0.0678*** (0.00748)	0.0676*** (0.00748)	0.0442*** (0.00828)	0.0437*** (0.00829)
2011.year	0.070*** (0.007)	0.0707*** (0.00741)	0.0705*** (0.00741)	0.0170** (0.00856)	0.0164* (0.00857)
2012.year	0.171*** (0.007)	0.172*** (0.00743)	0.171*** (0.00743)	0.112*** (0.00867)	0.111*** (0.00868)
2013.year	0.163*** (0.007)	0.165*** (0.00741)	0.164*** (0.00741)	0.0961*** (0.00889)	0.0950*** (0.00890)
2014.year	0.201*** (0.007)	0.203*** (0.00741)	0.203*** (0.00742)	0.124*** (0.00911)	0.123*** (0.00912)
Constant	1.999*** (0.059)	1.986*** (0.0591)	1.982*** (0.0591)	-1.042*** (0.154)	-1.057*** (0.154)
Observations	49 300	49 300	49 300	49 300	49 300
R-squared	0.284	0.286	0.286	0.150	0.150

Note: OLS, ordinary least squares; IV, instrumental variable.

ANNEX II. ENGEL CURVE ESTIMATION
IN PISSARIDES AND WEBER (1989)

For household i, let C_i stand for the expenditure share of food and Y_i^P the permanent income that influences consumption decisions. Z_i is a vector of household characteristics.

$$lnC_i = Z_i\alpha + \beta lnY_i^P + \varepsilon_i$$

The true income Y_i is a function of the reported income Y_i' and the permanent income

$$Y_i = k_iY_i'$$

$$Y_i = p_iY_i^P,$$

where $k_i \geq 1$ and p_i are random variables showing the degree of underreporting and the degree of transitory income variation, respectively. Note that $k_i = 1$ for employees. The distinction between permanent and transitory income may seem redundant, but it could be important, as it is likely that the covariance of underreporting and transitory income is not zero. If an individual has had an exceptionally good year in self-employment income, he may be less inclined to declare the full income, as it may arouse the curiosity of the tax authority. The mean of p_i is assumed to be the same for both groups, although the variance may differ. k_i and p_i are assumed to be log normal,

$$lnk_i = \mu_k + v_i$$

$$lnp_i = \mu_p + u_i$$

with errors that have zero mean and constant variances σ_v^2 and σ_u^2. Substitution gives

$$lnC_i = Z_i\alpha + \beta lnY_i' - \beta lnp + \beta lnk_i + \varepsilon_i$$

$$lnC_i = Z_i\alpha + \beta lnY_i' + \beta(\mu_k - \mu_p) + \beta(v_i - u_i) + \varepsilon_i$$

Let SE_i be a dummy taking the value one if the household is self-employed, and zero if not, and SE and EE stand for the two groups. Then,

$$lnC_{ij} = Z_i\alpha_j + \beta_j lnY_i' - \gamma_j SE_i + \eta_i$$

where $\gamma = \beta\left[(\mu_{kSE} - \mu_{kEE}) - (\mu_{pSE} - \mu_{pEE})\right] = \beta[\mu_{kSE} + \frac{1}{2}(\sigma_{uSE}^2 - \sigma_{uEE}^2)]$

This provides a rough estimate of underreporting, as $lnk = \gamma/\beta$. The mean of the underreporting component can be derived as

$$ln\bar{k} = \mu_{kSE} + \frac{1}{2}[\sigma_{vSE}^2 - (\sigma_{uSE}^2 - \sigma_{uEE}^2)]$$

Pissarides and Weber (1989) introduce a reduce form equation out of concern for endogeneity, but also to gain an independent estimate of the variances of underreporting and transitory income.

$$lnY_i' = Z_i\delta_1 + X_i\delta_2 + \tau_i$$

where X_i is a set of identifying instruments. Here, the error term consists of deviations of actual from permanent income and actual from reported income as well as the unexplained variation in permanent income. The residual variances for the two groups are related by

$$\sigma_{\tau SE}^2 - \sigma_{\tau EE}^2 = \sigma_{vSE}^2 + (\sigma_{uSE}^2 - \sigma_{uEE}^2) - 2cov(uv)_{SE}$$

The lower (σ_{vSE}^2 = 0) and upper ($\sigma_{uSE}^2 = \sigma_{uEE}^2$) bounds give an interval, whose mid-point is reported in Pissarides and Weber (1989).

$$ln\bar{k} \in [\frac{\gamma}{\beta} - \frac{1}{2}(\sigma_{\tau SE}^2 - \sigma_{\tau EE}^2) + cov(uv)_{SE} , \ \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{\tau SE}^2 - \sigma_{\tau EE}^2) + cov(uv)_{SE}]$$

REFERENCES

Allingham, M.G., and A. Sandmo (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, vol. 1, Nos. 3-4, pp. 323-338.

Barth, Erling, Alexander W. Cappelen, and Tone Ognedal (2013). Fair tax evasion. *Nordic Journal of Political Economy*, vol. 38, No. 3, article 3.

Becker, G. (1968). Crime and punishment: an economic approach. *Journal of Political Economy*, vol. 76, No. 2, pp. 169-217.

Bordignon, M. (1993). A fairness approach to income tax evasion. *Journal of Public Economics*, vol. 52, No. 3, pp. 345-362.

Cowell, F.A. (1990). *Cheating the Government: The Economics of Evasion*. Cambridge, MA: MIT Press.

Engström, P., and J. Hagen (2017). Income underreporting among the self-employed: a permanent income approach. *European Economic Review*, vol. 92, pp. 92-109.

Feld, L.P., and B.S. Frey (2002). Trust breeds trust: how taxpayers are treated. *Economics of Governance*, vol. 3, No. 2, pp. 87-99.

Hurst, E., G. Li, and B. Pugsley (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics*, vol. 96, No. 1, pp. 19-33.

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy*, vol. 31, pp. 99-109.

Kim, B., J. Gibson, and C. Chung (2017). Using panel data to estimate income under-reporting by the self-employed. *The Manchester School*, vol. 85, No. 1, pp. 41-64.

Kleven, H.J., and others (2011). Unwilling or unable to cheat? Evidence from a randomized tax audit in Denmark. *Econometrica*, vol. 79, No. 3, pp. 651-692.

Korea Institute of Public Finance (2018). *National Survey of Tax and Benefit*. (In Korean). Available at panel.kipf.re.kr.

Kukk, M., and K. Staehr (2017). Identification of households prone to income underreporting: employment status or reported business income? *Public Finance Review*, vol. 45, No. 5, pp. 599-627.

Luttmer, E.F.P., and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives*, vol. 28, No. 4, pp. 149-168.

Lyssiotou, Panayiota, Panos Pashardes, and Thanasis Stengos (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal*, vol. 114, No. 497, pp. 622-640.

Pissarides, C.A., and G. Weber (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, vol. 39, No. 1, pp. 17-32.

Schneider, F., and D.H. Enste (2000). Shadow economies: size, causes, and consequences. *Journal of Economic Literature*, vol. 38, No. 1, pp. 77-114.

Slemrod, J. (2007). Cheating ourselves: the economics of tax evasion. *Journal of Economic Perspectives*, vol. 21, No. 1, pp. 25-48.

Torgler, B. (2003). Tax morale and institutions. CREMA Working Paper Series, 2003-09. Basel, Switzerland: Center for Research in Economics, Management and the Arts.

United States, Internal Revenue Service (2016). Tax gap estimates for tax years 2008-2010. Available at www.irs.gov/newsroom/the-tax-gap.

Wooldridge, J.M. (2009). *Introductory Econometrics: A Modern Approach*, 4th ed. Nashville, Tennessee: South-Western.