# Distributional Implications of Joint Tax Evasion

Odd E. Nygård, Joel Slemrod and Thor O. Thoresen



Odd E Nygård, Joel Slemrod, Thor O Thoresen, Distributional Implications of Joint Tax Evasion, The Economic Journal, Volume 129, Issue 620, May 2019, Pages 1894–1923, https://doi.org/10.1111/ecoj.12619

# Distributional Implications of Joint Tax Evasion*

# Short title: Distribution and Joint Tax Evasion

Odd E. Nygård, Joel Slemrod, and Thor O. Thoresen

January 9, 2018

**Abstract**

Both buyers and sellers of goods and services may benefit from letting their economic transactions go unrecorded for tax purposes. The supplier reduces his tax burden by underreporting income and sales, whereas the consumer may gain from buying a non-taxed lower-priced

*Corresponding author: Joel Slemrod, Stephen M. Ross School of Business and Department of Economics, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109. Email: jslemrod@umich.edu. Financial support from the Norwegian Research Council is gratefully acknowledged. Thanks are due to Elin Halvorsen and Bård Lian for helpful advice and assistance on data. We are grateful for comments by seminar participants of the Oslo Fiscal Studies Staff Seminar (Oslo, October, 2014), the Norwegian Research Forum on Taxation (Bergen, September, 2015), Nordic Workshop in Tax Policy and Public Economics (Oslo, November, 2015), the 9th Norwegian-German Seminar on Public Economics (Munich, November, 2015), and the 72nd Annual Congress of the IIPF (Lake Tahoe, Nevada, August, 2016).

product. The distributional implications of such joint tax evasion depend on the amounts evaded, on where the evaders on both sides of the market are found in the income distribution, and on how the financial gain is split. Our empirical investigations show that the tax-evasion-controlled estimate of income inequality in Norway exhibits more income dispersion than official estimates.

Keywords: tax evasion, income inequality, expenditure approach

JEL codes: D31, D63, H26

Who gains more from tax evasion, the rich or the poor? This is a complicated question, because tax evasion generates numerous effects on factor and commodity prices (Kesselman, 1989; Slemrod, 2007; Alm and Finlay, 2013). Even when ignoring general equilibrium effects and discussing first-order distributional effects only, there are considerable complications to identifying the distributional effects, as we often lack reliable information on tax evasion. In the present paper we address how to account for the fact that much tax evasion behaviour involves the participation of more than one taxpayer, and estimate the distributional consequences using data from Norway.

The point of departure here is that tax evasion often takes place in the interaction between buyers and sellers of services and commodities to households, referred to as collusive, or joint, tax evasion (Abraham *et al.*, 2016). For example, the building and construction industry sector is one of the most tax evasion-ridden industries in Norway (KRISINO, 2011). Under a mutual agreement between sellers and buyers of goods and services, transactions or parts of transactions are often not reported to the tax authorities.

The distributional aspects of tax evasion have received little attention in the literature. Two exceptions are Bishop *et al.* (2000) and Johns and Slemrod (2010), who both use micro data to address the issue of how measures of tax redistribution and income inequality are altered by accounting for tax evasion. A novelty of the present study is that we discuss distributional effects of tax evasion from a market perspective, i.e., tax evasion that results from suppliers and buyers of goods and services deciding jointly to let their transactions go unrecorded for tax purposes, which means that both sides of the market may gain financially.

Measures of the financial gains from tax evasion on the part of the producers are obtained using the so-called expenditure approach method (Pissarides and Weber, 1989), which is an "indirect method" of identifying evasion behaviour. See surveys of the literature in Andreoni *et al.* (1998), Schneider and Enste (2000), Slemrod and Yitzhaki (2002), Torgler (2007), Slemrod and Weber (2012), and Alm (2012). In an indirect-method approach, evasion is not measured directly, but indirectly, via measurable traces of true income.

Several groups of wage earners have limited scope for tax evasion (because effective third-party reporting of income is a standard procedure), and they therefore represent a convincing benchmark. We modify the original Pissarides and Weber framework somewhat by letting the food consumption and income of the benchmark group be compared with two other groups: not only the self-employed, but also wage earners who may be involved in tax evasion on the supply side (as employed craftsmen). As in Pissarides and Weber (1989), we assume that there is a common slope in the Engel curves for food, but allow the intercepts to differ in the three groups. In this way, using data from the Survey of Consumer Expenditure (Holmøy and

Lillegård, 2014), we can estimate the amount by which reported income must be scaled up in order to obtain true income levels for tax evaders,[1] and allow for variations in the degree of underreporting along the income range.

With respect to the buyers' side, we rely on a "direct method" for measuring tax evasion, using Norwegian survey data on purchases of illegal services (TNS Gallup, 2009; Opinion, 2006, 2014, 2016). We estimate a probability of being involved in transactions not reported to the tax authorities, on the basis of characteristics such as income and education. Non-reported supply-side income is used in turn to derive the amount of tax evaded by buyers of the services.

By combining empirical evidence obtained from the supply side and the demand side of the market, we can address two aspects of the distributional effects of tax evasion: what the tax-evasion-controlled distribution of income looks like, and how the tax savings from not reporting income and transactions are distributed. A measure of income inequality that controls for tax evasion accounts for the unreported income of suppliers and the benefits to consumers of paying a lower price. We obtain tax-evasion-adjusted incomes for suppliers by applying estimates of the expenditure approach directly. To calculate additional income for consumers, we use the estimated evasion probabilities, along with assumptions about how prices in the hidden

[1] The so-called "expenditure approach", set forth by Pissarides and Weber (1989) and exemplified by application to British data, has sparked tax evasion examinations in several other countries, but estimates for Norway have so far not been provided. Tax evasion estimates for other countries by this method include Schuetze (2002) for Canada, Johansson (2005) for Finland, Engström and Holmlund (2009) and Engström and Hagen (2017) for Sweden, Martinez-Lopez (2013) for Spain, Paulus (2015) for Estonia, Feldman and Slemrod (2007) and Hurst *et al.* (2014) for the U.S, and Kim *et al.* (2017) for Korea and Russia.

market deviate from prices in the regular market. The overall effect depends on how both suppliers and consumers are positioned in the distribution of income. We are then able to discuss how the "hidden-economy-controlled" income distribution compares with the official one: is it less or more equal? If a recorded income distribution exhibits high inequality simply because many tax evaders are found at the low end of the income distribution because of low reported income, this has profound implications for redistribution policies.

The distribution of the reductions in tax burdens of suppliers and consumers due to tax evasion is, of course, related to these adjustments in income. However, measuring distributional effects becomes more complicated when we address reductions in tax burden (tax savings), instead of income. Results here depend even more than those from an income distribution perspective on how suppliers and consumers divide the economic gains, which we do not observe. For example, we may have a situation where the suppliers' returns are squeezed to the extent that the consumers retain all of the financial advantage, and the distribution of the financial gain is determined by the demand side alone. We show that the tax-evasion-controlled income distribution is more unequal than the distribution not taking account of evasion, and the effective tax progressivity is less than indicated by official figures.

Although the present study does not provide definitive evidence about all the elements involved in a robust depiction of the distributional effects of joint tax evasion, we show how empirical evidence can be used to provide empirical illustrations of effects, given the conceptual foundations. The paper is organised as follows. Section 1 summarises some of the main perspectives on tax evasion, and collusive tax evasion in particular. In Section 2 we probe deeper into the theoretical background to our empirical investigations, while Section 3 presents the empirical approaches to obtaining measures of economic gain for the supply and demand side, respectively, and estimated results. The overall effects on the distribution of economic well-being are summarised in Section 4, and Section 5 provides a conclusion for the paper.

# 1 Preliminaries

We first consider the supplier. As is standard (Allingham and Sandmo, 1972; Yitzhaki, 1974; Andreoni *et al.*, 1998), the agent has an (exogenously given) income level, and faces a tax rate. Then he decides how much income he will report to the tax authorities, comparing the expected utility of being detected and paying a penalty for tax evasion to the expected utility of being able to keep the evaded tax. The same expected utility reasoning can be used to explain behaviour on the demand side (Cremer and Gahvari, 1993), as exogenously given disposable income can be used to buy commodities or services when there are two types of possible transactions, regular and hidden.

Although we shall proceed from this standard framework, we acknowledge that these simple models do not provide a complete description of everyone's decision-making. One key criticism is that some taxpayers are probably not motivated by narrow self-interest alone, but instead act as a member of a group, influenced by norms, customs, reciprocity, and patriotism. Further, notions such as shame, guilt and morality arguably also influence decisions in some circumstances. Others argue that the expected utility model does not provide a satisfactory description of peoples' perception of risk, i.e. they seem to attach too much weight to low-probability events, which has resulted in contributions applying prospect theory (Dahmi and al-Nowaihi, 2007) and the rank-dependent expected utility model (Eide *et al.*, 2011). We believe, however, that the Allingham-Sandmo deterrence model explains the essential reasoning underlying the theoretical framework.

There are other studies that challenge the predominant perspective in the literature that tax evasion is an interaction between a single economic agent and the government. For example, previous studies have elaborated on collusive tax cheating between employees and employer, see Yaniv (1988; 1992) and Kleven *et al.* (2016). Boadway *et al.* (2002) construct a model in which tax evasion requires the collaboration of at least two taxpayers. Using a game theory approach, they describe how sanctions against tax evasion may lead to a direct increase in the expected cost of a transaction in the illegal sector, but may also increase the ability of an agent to commit to cooperating in tax evasion, and may therefore lead to more tax evasion. Similarly, Chang and Lai (2004) model collaborative tax evasion between a seller and his customer as a game, and incorporate a social norm into such collusive tax-evading activities. More prevalent tax evasion undermines social norms; penalties may induce more collaboration and may therefore lead to increased tax evasion if tax evasion is already widespread, explained by a snowballing effect (or a critical-mass force). Abraham *et al.* (2016) show that, in a laboratory setting, the tax compliance norm has a stronger negative effect on the magnitude of collusive tax evasion than on independent tax evasion. Ognedal (2016), however, focuses on honesty being a competitive disadvantage and "tax morale" representing a poor substitute for sanctions in markets. Honesty reduces cheating, but the output may be less efficiently produced and less efficiently allocated between buyers.

The market-transaction perspective of the present study requires addressing the general equilibrium effects of tax evasion. Thus, the discussion of tax evasion not only accounts for effects working through different sides of the market, but in principle can control for a whole range of reactions by individuals and firms. Persson and Wissén (1984) study, analytically, the conditions under which the actual income distribution is more equal, or more unequal, than the distribution based on reported income. Richer descriptions of the incidence effects of tax evasion can be obtained by employing computable general equilibrium models, as in Alm and Sennoga (2010), who examine how much of the initial benefit of income tax evasion is retained by the evaders and how much is shifted via factor and commodity price changes stemming from mobility.[2]

As the present study analyzes micro data for the supply and demand side, previous studies that use micro data to discuss distributional aspects of tax evasion are relevant. Bishop *et al.* (2000) and Johns and Slemrod (2010) use data from the comprehensive random audit programmes of the Internal Revenue Service (IRS) of the U.S., which allow the researchers to observe income as reported and as adjusted by an audit. Bishop *et al.* (2000) find that including unreported income has only a very small (negative) impact on pre-tax income inequality as measured by either the standard or the extended

[2] See also the discussion in Alm and Finlay (2013).

Gini coefficient. The inclusion of both unreported income and additional taxes owed also has a negligible impact on inequality. Johns and Slemrod (2010) find that accounting for tax noncompliance makes the true income distribution more unequal, but the tax system becomes more progressive. This follows because a given percentage reduction in taxable income corresponds to a particularly high percentage reduction in tax liability for taxpayers with taxable income just above the taxpaying threshold.[3] Kleven *et al.* (2011) also use variation in auditing to identify tax evasion magnitudes and, even though distributional effects are not a main topic of the paper, they report (p. 673) that those with relatively little self-reported income evade more, as a share of self-reported income, than those with relatively high self-reported income. Alstadsæter *et al.* (2017) discuss tax evasion in relation to wealth (instead of income). When using information from random audits and leaks for Sweden, Denmark and Norway, they find that tax evasion clearly increases with wealth – the top 0.01% evade about 30% of their taxes.

Further, Pashardes and Polycarpou (2008) employ an expenditure approach technique, outlined in Lyssiotou *et al.* (2004),[4] and data from Cyprus to estimate tax evasion. Their findings suggest that the income underreporting biases estimates of both inequality and poverty downwards. Tedds (2010) uses an alternative way of implementing the expenditure-based method: parametric restrictions are relaxed and a nonparametric approach to the

[3] Christian (1994) also analyses data from the Taxpayer Compliance Measurement Program to discuss distributional aspects of tax evasion. The study finds that low-income individuals evade more than high-income individuals in the US. In 1988, taxpayers with (auditor-adjusted) incomes over $100,000 on average reported 96.6% of their true incomes to the tax authorities, compared to just 85.9% for those with incomes under $25,000.

[4] Instead of using expenditures on food only, as in Pissarides and Weber (1989), Lyssiotou *et al.* (2004) use information on a whole range of consumer goods.

measurement of income underreporting is explored, thereby reducing the number of assumptions required for estimation. The approach is illustrated by estimating the effect of the Canadian Goods and Services Tax (GST) on income underreporting among the self-employed, and the analysis concludes that the GST increased tax noncompliance by those with larger amounts of self-employment income, whereas tax noncompliance by those with small amounts of self-employment income was not affected.

Finally, we note the results of studies discussing the distributional effects of tax evasion by "discrepancy methods", meaning that data from an income survey are compared to the income reported by income tax returns. Taxpayers may conceal part of their income from the tax authorities, but might consider declaring a higher figure to an anonymous interviewer. Fiorio and D'Amuri (2005), Matsaganis and Flevotomou (2010) and Benedek and Lelkes (2011) use this method on data from Italy, Greece, and Hungary, respectively. Fiorio and D'Amuri (2005) find that the share of unreported income in Italy falls with income, Matsaganis and Flevotomou (2010) suggest that tax evasion generates higher income inequality, more poverty and lower income tax progressivity, which is also in line with the findings in Benedek and Lelkes (2011).

# 2 Theoretical Framework

Before we embark on the empirical investigation, we develop a simple theoretical framework. We assume there are two commodities in the economy: a numeraire good, $c_1$, that cannot be sold in the informal market and a service,

$c_2$, that (partly or entirely) may be sold informally. An individual may be both a supplier and a consumer of good $c_2$, although we will refer to suppliers and consumers as if they are separate individuals.

## *2.1 Supply Side*

Each supplier has a skill level denoted $n$, and a skill type, so that they can supply $c_1$ or $c_2$, but not both. A supplier of $c_2$ decides whether to report the income for tax purposes, remit tax at income tax rate $\tau$, and thereby supply the service formally, or else to not report the income and to supply the service in the informal market.[5,6] In the regular market he gets the price before indirect taxation (before VAT), $p_r$, for the supply of $c_2$, while the price is $p_h$ in the informal market. Let $x_r$ and $x_h$ be the true before-tax income if the supplier is operating in the regular or hidden economy, and let $x_r = np_rl_r$ and $x_h = np_hl_h$, where $l_r$ and $l_h$ denote the optimally chosen labour supply in each sector.

The supplier's expected economic gain from tax evasion is established by computing his expected financial gain from supplying $l_h$ in the hidden market relative to supplying the same amount in the regular market. Let $x_r^* = np_rl_h$ be the (hypothetical) income if the hidden hours of work were instead supplied in the regular market. The expected gain for a supplier of

[5] We assume for the sake of simplicity that both seller and buyer know with certainty whether a transaction will be reported for tax purposes.

[6] We can think of this as a decision in accordance with the Allingham and Sandmo (1972) framework.

participating in the hidden market, $SG$, can then be expressed as

$$\begin{aligned} SG &= (1-\rho)\left(x_h - x_r^*(1-\tau)\right) + \rho(x_h(1-\tau) - x_h\theta - x_r^*(1-\tau)) \\ &= [x_h - x_r^*(1-\tau)] - \rho x_h(\tau+\theta), \end{aligned} \tag{1}$$

where $\theta$ is the penalty levied if the supplier is caught evading tax, as a proportion of unreported income, and $\rho$ is the probability of the fraud being discovered. The term in the square bracket captures the income difference between entering the hidden market and being paid $x_h$ and letting the same hours of work be part of reported pre-tax income, $x_r$. Because there is a risk of being prosecuted and fined, the expected costs of being caught, $\rho x_h(\tau+\theta)$, enter the equation as well. If there is a negligible risk of being discovered, and if the price in the hidden market equals the pre-tax price in the regular market ($p_h = p_r$), such that $x_r^* = x_h$ for the same amount of labour, the gain from tax evasion corresponds to the (hypothetical) income tax ($x_r^*\tau$) on evaded income, i.e., the tax burden if the service was delivered in the regular instead of in the hidden market.

It follows that the relationship between $p_r$ and $p_h$ affects the division of the financial gain between the supply side and the demand side. At the extreme, if $p_r(1-\tau) > p_h$, working in the hidden market generates lower income than in the regular market (for the same working hours). Crucially, we do not observe prices in the two markets, and we therefore do not know how the gain is split between the two sides of the market. In Section 4 we discuss how we proceed to account for this in the distributional analysis.

## *2.2 Demand Side*

Each consumer chooses $c_1$ and $c_2$, and can decide to purchase $c_2$ either in the informal or in the regular market, with the optimal amounts in the two markets denoted $c_h$ and $c_r$, respectively. If he buys services in the hidden market and the fraud is detected by the tax authorities, he pays a penalty,[7] given by a fraction of the tax evaded, denoted $\kappa$. The probability of being caught is denoted $\eta$, and $\pi$ is the tax rate on the service, the VAT. The consumer's expected financial gain, $CG$, from purchasing $c_h$ in the hidden market instead of in the regular market is given by,

$$\begin{aligned} CG &= (1-\eta)\left(c_h p_r(1+\pi) - c_h p_h\right) + \eta(c_h p_r(1+\pi) - c_h p_h(1+\pi) - c_h p_h \kappa\pi) \\ &= \left[c_h p_r(1+\pi) - c_h p_h\right] - \eta c_h p_h \pi(1+\kappa). \end{aligned} \tag{2}$$

The term in the square brackets represents the financial gain on the hidden purchase, whereas the last term describes the penalty.

Thus, the economic gains at the demand side come in the form of different actual price vectors for different consumers, some paying prices below the observable regular price. The true real income can then be seen as, $y^*/P$, where $y^*$ is disposable income and $P$ is a price index.[8] If $y^*/P$ increases with

[7] A recent court case in Norway attests to this type of penalty being imposed on buyers in the informal sector. A buyer of hidden cleaning services was fined NOK 20,000 and sentenced to a 30-day (suspended) prison term. Other customers of the same cleaner were also fined.

[8] To see this, begin by defining real income for an individual as $y^*/P$ , where $P = (p_h c_h + p_r c_r + c_2) / (p_r(1+\pi)c_h + p_r c_r + c_2)$, i.e. $P$ is an individual-specific price index, depending on how much the person buys in the hidden market. This means that the true income is adjusted upwards as long as $p_h < p_r(1+\pi)$ (because $P < 1$). However, we have $P = (p_h c_h + p_r c_r + c_2) / (p_r(1+\pi)c_h + p_r c_r + c_2) = y^*/(y^* + (p_r(1+\pi) - p_h)c_h)$. When $b$, is

income, the (real) true income distribution is more unequal than the official one.

## *2.3 Equilibrium*

To summarise the theoretical outline of tax evasion from a two-sided market perspective, the economy has three markets, one for the $c_1$ good and two, the regular and the informal, for the $c_2$ service. Prices in the regular and hidden market for $c_2$ equilibrate demand and supply so that $S_h(p_r, p_h, \tau, \theta, \rho) = D_h(p_r, p_h, \pi, \kappa, \eta)$ and $S_r(p_r, p_h, \tau, \theta, \rho) = D_r(p_r, p_h, \pi, \kappa, \eta)$, where $S(.)$ and $D(.)$ are aggregate supply and demand, respectively, and where these functions implicitly account for the disposition of collected revenue. The shape of these curves will determine how the prices in the hidden market relate to the price in the regular market, and how the gains are shared between the supply and demand side.

A key component of our approach to tax evasion is that there is an equilibrium condition determining tax evasion on both sides of the hidden market. If we let the observations of supply and demand side be represented by $i$ and $j$, respectively, this can be expressed as

$$\sum_i x_{hi} = \sum_j p_h c_{hj}, \tag{3}$$

i.e., the sum of evaded income equals the sum of hidden consumer payments.[9] However, as we observe $x_{hi}$, and only establish to what extent $c_{hj} > 0$ with

the budget share of the hidden purchase, is given by $b = (c_h(p_r(1+\pi) - p_h))/y^*$, we have $P = 1/(1+b) => 1/P = 1 + b$. Then we can write $y^*/P = y^*(1+b) = y^* + CG$, when the risk of getting caught is excluded.

[9] We suppose that they are subject to regular prices and taxes on factor inputs.

respect to the demand side, we use Equation (3) to obtain aggregate evaded amounts for the demand side, too.[10]

In the next section we turn to the estimation of key parameters for this framework, which are used to describe the overall distributional effects of tax evasion. We estimate the relationship between the true and reported income of the self-employed, and to some extent let it vary according to income. Further, we investigate the characteristics of the demand-side tax evaders. By combining these two sources of information, we describe the overall distributional impact of collusive tax evasion in the informal sector.[11]

# 3 Quantifying Tax Evasion on both Sides of the Market

## *3.1 Identifying the Tax Evaded Income of Suppliers*

The expenditure approach follows from the assumptions that some individuals have the opportunity to underreport, while others do not, and that the groups have similar preferences for a consumption good. It is assumed that

[10] As we shall discuss in Section 4, information on amounts evaded is limited, and we therefore use supply-side estimates to define demand-side quantities. In so doing we also a balance between the evaded amounts of the two sides of the market.

[11] Another type of joint tax evasion takes place in the form of collusive agreements between firms and workers. Firms hire workers under the mutual understanding that neither tax on workers' income nor payroll tax is remitted. As we shall soon explain, we also allow for some groups of wage earners being involved in tax evasion, which to some extent picks up this other type of joint tax evasion. However, we believe that joint tax evasion between firms and workers (at least in Norway) often involves foreign manpower, making it hard to trace in the data in use here. Of course, the focus on collusive tax evasion does not rule out that there are types of noncollusive tax evasion that affect income distribution.

for all parties consumption is determined by permanent disposable income, $y^{pe}$, and a number of individual control variables, $Z'$. When using the log form, we have the following Engel curve relationship, $\ln e = Z'\gamma + \beta \ln y^{pe}$, where $\gamma$ and $\beta$ are parameters.[12] Pissarides and Weber (1989) let $e$ represent expenditure on food and assume that the self-employed are the only group with scope for underreporting. However, the current paper allows some groups of wage-earners to be involved in tax evasion too, thus acknowledging that some wage-earning groups, such as painters or carpenters, may use their "leisure time" to work in the informal economy.[13] Thus, we let the subscript $m$ indicate that there are differences across individuals in the scope for evasion, dividing the population into three types, $m \in [SE, SW, BG]$, the self-employed ($m = SE$), salary workers with scope for tax evasion ($m = SW$), and the benchmark group of salary workers who (by assumption) do not evade tax ($m = BG$).

To back out the true income of the self-employed from the Engel curve relationship, Pissarides and Weber define a proportionality factor, $k$, which defines the relationship between observed income, $y$, and true income, $y^*$,

$$k \equiv \frac{y^*}{y}. \tag{4}$$

Thus, the proportionality factor, $k_{im}$, represents the factor by which the

[12] Thus, we assume a log-linear Engel curve. One alternative is to employ a quadratic form, as argued for by Banks *et al.* (1997), Lyssiotou *et al.* (2004), and Fortin *et al.* (2010).

[13] For example, they may provide paid help to family or acquaintances. As emphasized by Williams (2008), many informal economy buyer-seller interactions are of this type. See also Martinez-Lopez (2013), Dunbar and Fu (2015) and Paulus (2015). Even though we relate these activities to salary workers here, they may more correctly be seen as classical self-employment work.

observed income for individual $i$ belonging to group $m$ must be multiplied in order to obtain true income, given that there are differences across individuals in the scope for evasion. It follows that both $k_{iSE}$ and $k_{iSW}$ can be larger than 1, whereas we assume that there is no tax evasion in the benchmark group ($k_{iBG} = 1$).

Standard applications of the expenditure approach assume that current income fluctuates around permanent income by a factor $g$, expressed as $y^*_{im} = g_{im} y^{pe}_{im}$, and usually assume that the coefficients $\ln g_{im}$ and $\ln k_{im}$ are lognormally distributed around their means, $\ln g_{im} = \mu_{gm} + u_{im}$ and $\ln k_{im} = \mu_{km} + v_{im}$. Then, to establish a link between the Engel curves estimates and the proportionality factor, $k$, a relationship between permanent income and observable income is utilised, $\ln y^{pe}_{im} = \ln y_{im} - (\mu_{gm} - \mu_{km}) - (u_{im} - v_{im})$, which generates the following Engel curve,

$$\ln e_{im} = Z_{im}\gamma + \beta \ln y_{im} - \beta \left(\mu_{gm} - \mu_{km}\right) - \beta \left(u_{im} - v_{im}\right) + \varepsilon_{im}, \qquad (5)$$

where $\varepsilon_{im}$ is a random error term. If, for expositional reasons, we let the Engel curve be adjusted by an indicator variable, $q_i$, which takes the value 1 for the self-employed, $SE$, and 0 for the benchmark salary worker group, $BG$, Equation (5) becomes

$$\begin{aligned} \ln e_{im} &= Z'_{im}\gamma + \beta \ln y_{im} + \beta \left(\mu_{kBG} - \mu_{gBG}\right) \\ &\quad + \beta q_i \left[\left(\mu_{kSE} - \mu_{kBG}\right) + \left(\mu_{gBG} - \mu_{gSE}\right)\right] \\ &\quad - \beta \left(u_{im} - v_{im}\right) + \varepsilon_{im}. \qquad (6) \end{aligned}$$

The mean of $k$ for the self-employed is given by $\ln \overline{k}_{SE} = \mu_{kSE} + \frac{1}{2}\sigma^2_{vSE}$, where $\sigma^2_{vSE}$ is the variance of $v_{im}$ for $m = SE$. Also, as $k_{BG} = 1$ for $m = BG$, and as the mean of $g$ is assumed to be identical in the two groups, $\ln \overline{g}_{SE} = \ln \overline{g}_{BG}$,[14] Equation (6) can be rearranged into the following reduced form, which is the standard empirical specification used to obtain estimates of $k$,

$$\ln e_i = Z_i^{'}\gamma + \beta \ln y_i + \delta q_i + \xi_i. \tag{7}$$

As $\delta = \beta \left[\mu_{kSE} + \frac{1}{2}\left(\sigma^2_{uSE} - \sigma^2_{uBG}\right)\right]$ and $\xi_i = -\ \beta \left(u_{im} - v_{im}\right) + \varepsilon_{im}$, an estimate of the adjustment factor $k$ is given by

$$\overline{k}_{SE} = \exp\left[\mu_{kSE} + \frac{1}{2}\sigma^2_{vSE}\right] = \exp\left[\frac{\widehat{\delta}}{\widehat{\beta}} + \frac{1}{2}\left(\sigma^2_{vSE} + \sigma^2_{uBG} - \sigma^2_{uSE}\right)\right], \tag{8}$$

where $\widehat{\delta}$ and $\widehat{\beta}$ are the estimated parameters of Equation (7). However, as $\sigma^2_{vSE}$, $\sigma^2_{uBG}$, and $\sigma^2_{uSE}$ are usually not known, a standard empirical approach, along the lines of Pissarides and Weber, involves obtaining estimates of the variance of the residuals, $\zeta_{im}$, from an expression, $\ln y_{im} = A_{im}\psi + \zeta_{im}$, where $A_{im}$ includes a set of instruments for permanent income.[15] Then, as shown by Martinez-Lopez (2013) and Kim *et al.* (2017), an estimate of $\overline{k}_{SE}$ is obtained by using estimated coefficients,

$$\overline{k}_{SE} = \exp\left[\frac{\widehat{\delta}}{\widehat{\beta}} \pm \frac{1}{2}\left(\widehat{\sigma}^2_{\zeta SE} - \widehat{\sigma}^2_{\zeta BG}\right)\right]. \tag{9}$$

Our approach to obtaining estimates of $k$ for specific groups of salary workers,

[14] This critical assumption follows Pissarides and Weber.

[15] As spelled out by Pissarides and Weber, $\sigma^2_{\zeta SE} - \sigma^2_{\zeta BG} = \sigma^2_{vSE} + \sigma^2_{uSE} - 2cov(uv)_{SE} - \sigma^2_{uBG}$. They discuss results for both a lower bound case ($\sigma^2_{vSE} = 0$) and an upper bound alternative ($\sigma^2_{uBG} = \sigma^2_{uSE}$). See also Wangen (2005).

$m = SW$, follows the same type of reasoning. In our main specification we estimate Equation (7) directly by using a measure of permanent income for $y_i$, which simplifies Equation (9).[16] We shall return below to the details of how this is done.

Moreover, as a key objective here is to obtain information about how $k$ varies with respect to income, we introduce non-linearities in the measurement of $\delta$. For example, this can be done straighforwardly by introducing a dummy variable denoting high income, $HI_i$, and letting it interact with the variable indicating that there is scope for underreporting, $q_i$, in Equation (7),

$$\ln e_i = Z_i^{'}\gamma + \beta \ln y_i + \delta q_i + \lambda(q_i \times HI_i) + \nu_i, \tag{10}$$

where $\nu_i$ is the error term. Thus, we allow for a differentiation in the self-employment intercept with respect to income level.

## *3.2 Data and Estimation Results for the Expenditure Approach*

Estimates of $k$ are obtained by examining data from the Norwegian Survey of Consumer Expenditure (Holmøy and Lillegård, 2014). We use (pooled) information from the survey for each year from 2003 to 2009, and for 2012. In the period 2003–2009 the surveys were based on random draws of 2,200 individuals. There were no surveys in 2010 and 2011, and the 2012 survey is much larger, with a gross sample of 7,000 individuals. The average response rate across the surveys is approximately 50%.

[16] This means that the expression for $y_{im}^{pe}$ is simplified, $\ln y_{im}^{pe} = \ln y_{im}^{pe(obs)} + \mu_{km} + v_{im}$, which implies that the corresponding expression to Equation (5) is $\ln e_{im} = Z_{im}\gamma + \beta \ln y_{im}^{pe} + \beta\mu_{km} + \beta v_{im} + \varepsilon_{im}$. Thus, also in this specification, $y_{im}^{pe}$ is correlated with the error term, by construction. We will return to the use of instruments shortly.

These data are closely related to the information used by Pissarides and Weber (1989), as the Survey of Consumer Expenditure is based on personal interviews and detailed accounting of household expenditure on food and other consumption items for a period of 14 days. In addition to expenditure, the data include household characteristics, such as income and education, obtained by linking to administrative registers. Information on disposable income, defined as gross income minus tax, is obtained from income tax return data. Food expenditure includes all types of food and non-alcoholic beverages.[17]

As already noted, we employ in the estimation a measure of permanent income which is obtained by taking the average of income over seven years, measured in 2012 prices. Measures of permanent income are linked to the expenditure data by using an income panel dataset for the whole population (Statistics Norway, 2017). As personal ID numbers are unavailable, we merge by using "backward identification" methods, utilising the fact that there are common (identifying) variables in the two datasets, to establish a unique combination of values for the individuals. As we do not have enough information in the datasets to find unique matches for all observations, some observations are lost, but we are able to retain more than 70% of the sample. The sample for which we have observations of permanent income is used in all the estimations. Appendix A contains a discussion of the implications of restricting the estimation to the matched sample. It is argued that this does

[17]However, note that food consumption at restaurants is not included, which may create measurement problems. For example, a self-employed restaurant owner is likely to report very low food purchases compared to a wage earner with similar income. If that is the case, the expenditure method will underestimate the hidden income of the restaurant owner.

not lead to biased estimates, as we find close correspondence between the characteristics of the two datasets.

In line with earlier contributions using the expenditure approach, we exclude those self-employed in agriculture and fisheries. When also conditioning on at least one person in the household being self-employed or wage earner (to exclude pensioner households), we end up with a total sample of approximately 6,200 households. Table 1 (upper panel) provides descriptive statistics for the sample used in the estimation, differentiating between three groups: the benchmark group of non-evading salary workers, the self-employed, and salary workers with (assumed) scope for tax evasion. Households are defined as belonging to the self-employment group as long as one household member is self-employed,[18] and we identify the salary workers with possibilities for tax evasion by field of education: for example, employees with a training as electricians and in building and construction are believed to be able to participate in the hidden economy as well (in their "leisure time"). As we provide estimation results for $k$ when differentiating between high and low income and high and low age, Table 1 also includes information about the sample sizes of these subgroups. Median income (approx. 580,000 NOK)[19] and median age (45) divide the sample into the four groups.

Four sets of estimation results are presented in Table 2, while in Table 3 we show results differentiated by level of income and age. In Table 2 we present the standard expenditure approach results for the self-employed and

[18] In the survey, people self-report their occupation. Note also that some studies, such as Schuetze (2002) and Johansson (2005), present results by the number of self-employed persons in the household.

[19] According to average exchange rates for 2012: 1€ = 7.47 Norwegian kroner (NOK), and 1$ = 5.82 NOK.

for those wage earners assumed to have some scope for evasion, based on the specification in Equation (7). As in most of the previous literature, including Pissarides and Weber (1989), estimates are also obtained by using IV – see Equation (6) for a demonstration of the econometric challenges.[20] Results for specifications in which income is allowed to be represented by both annual income and permanent income are provided, using both ordinary least squares (OLS) and instrumental variables (IV) techniques in the estimation. After some experimenting, we employ car ownership and possession of two or more cars as the excluded instruments.[21] We let income be instrumented both in the annual income variant and the permanent income version, as both specifications involve endogeneity issues. Given that the seven-year income average represents permanent income adequately, Equation (9) is simplified, $\sigma^2_{uBG} = \sigma^2_{uSE}$, and the remaining contribution comes from variance in the self-employment underreporting rate ($\sigma^2_{vSE}$), and $\overline{k}_{SE} = \exp\left[\frac{\widehat{\delta}}{\widehat{\beta}} + \frac{1}{2}\left(\widehat{\sigma}^2_{\zeta SE} - \widehat{\sigma}^2_{\zeta BG}\right)\right]$ is used to obtain estimates of $\overline{k}_{SE}$.[22] Note that $\overline{k}_{SW}$ is calculated in the same manner.

Table 2 reports results for the slope parameter ($\beta$), the difference in the intercepts ($\delta_{SE}$ and $\delta_{BG}$), and $\overline{k}$. In Table B2 in Appendix B we show estimation results for each annual sample of the Survey of Consumer Expenditure.

[20] Engström and Hagen (2017) argue that the attenuation bias generated by erroneously using current income in the regression leads to overestimation of the underreporting.

[21] As in Engström and Hagen (2017), we find that education and house size do not satisfy the exclusion restriction. Using capital income as an instrument gives us approximately the same estimates for $k$ as presented here, but the results of the test statistics (F statistic of the first stage, p-values of the Sargan's overidentification test, and the Wu-Hausman endogeneity test) suggest using car ownership.

[22] In using annual income to calculate $k$, we use the fact that we have already estimated the variances of Equation (9) for permanent income. More precisely, the difference between residual variances, when the residual variances are obtained from estimations using two different income definitions, enter into the calculation of $k$ in the annual income case.

The results clearly illustrate that the single-year samples that we have available for this study are too small to provide reliable results. In fact, the implied evasion rate for 2012, the year for which we have the largest dataset, is not statistically significant. These results illustrate that overall $k$ is estimated with considerable uncertainty. In Appendix B, Table B3, we report the power of the test for the estimates of $k$. As expected, the power is low in many cases. Moreover, the estimates reported in Table B1 (in Appendix B) suggest that the linearity assumption of the Engel function is not rejected, whereas Table B4 shows that the slope estimates ($\beta$) are relatively close across the three groups: the self-employed, wage earners with assumed scope for income underreporting, and other wage earners.

The estimates of $\overline{k}$ are not sensitive to the choice of income measure and the estimation procedure. They range from 1.14 to 1.16 and are statistically significant in three of four specifications, when standard errors are obtained by the delta method. Estimates according to the preferred permanent income specification are 1.16 and 1.15, for the OLS and the IV estimation, respectively.[23] Thus, the OLS and IV estimates are close. To our knowledge, these estimates are the first self-employment tax evasion estimates for Norway based on using the expenditure approach. Although, they are somewhat lower than Engström and Hagen (2017) find for Sweden, we find it reassuring that they are not far from the estimate of the share of underreported income among the self-employed in Finland and Denmark, as reported by Johansson

[23] In a companion paper we use charitable contributions instead of food consumption for identification, and find $\widehat{k}$-values close to the estimates reported here (for much larger samples).

(2005) and Kleven *et al.* (2011).[24]

We find no significant evidence of tax evasion among wage earners who are assumed to have scope for tax evasion. However, point estimates are on the positive side across all four specifications for these salary workers, too, and we shall explore how an effect, as given by the point estimates, may influence the depiction of the distributional effects of underreporting.

Further, some preliminary data investigations revealed that age plays an important role, in addition to income level, as high age goes together with a high estimated value of $k$. Thus, Table 3 shows what happens when we allow $k$ to differ with respect to income level and age. With respect to the self-employed, the estimated value of the $k$ of the low-income/high-age group stands out. This $k$ is clearly statistically significantly different from 1, but is also significantly different from the other estimates.[25] In the distributional analysis that follows, although the other estimates are not statistically different from 1, we let $k$ vary with the point estimates.[26] Note that the results

[24]Both Kleven *et al.* (2011) and Engström and Hagen (2017) report evasion magnitudes measured as the fraction of true income that is underreported, which means that recalculation is needed in order to compare with our estimates of $k$ (the factor by which observed income must be multiplied in order to obtain true income). Note that the comparison to the results of Johansson (2005) (for Finland) refers to the average $k$ for households in which the head is self-employed.

[25]One may conjecture that this follows from a higher marginal propensity to consume among high-age individuals (as for example suggested by the life-cycle hypothesis). In order to investigate this further, we have introduced age dependency in income in the estimation of the Engel function for the wage earners of the benchmark group. However, we do not find significant effects of age.

[26]In the same vein as Tedds (2010), although not applying nonparametric methods as she does, we have also estimated a two-step version of the expenditure approach. In a first step we obtain a food consumption-income relationship for (non-evading) wage earners. This relation is used in a second step to obtain a predicted true income for the self-employed, and the divergences between predicted (true) income and observed income along the income scale describe the distributional effects. As expected, given this method's proximity to the Pissarides and Weber approach, the plot obtained does not provide any additional information to the results already obtained, as in Table 3.

for $\overline{k}_{SE,LI,HA}$ are not consistent with findings reported in Johns and Slemrod (2010) for the U.S., as they find that the ratio of aggregated misreported income to true income generally increases with income.[27]

### *3.3 Buyers in the Hidden Market*

For evidence of tax evasion on the demand side, we use information from four surveys that were carried out in 2006, 2009, 2014, and 2016 to increase knowledge about the informal economy through interviews. Each survey consists of approximately 2000 respondents, and the results are documented in TNS Gallup (2009) and in Opinion (2006; 2014; 2016). The surveys are conducted using two data collection methods. The first survey, in 2006, is based on standard data collection, where a random sample of the population was interviewed. The three other surveys are based on interviewees belonging to web panels. Of course, given the topic of the surveys, there is a clear potential for non-response bias.[28] Here, we pool the information from the four surveys and, after excluding respondents with partially missing information, we are left with approximately 6,300 respondents, which are used in the estimation of the demand side behaviour.

13%, 23%, 11%, and 14% of the respondents, for 2006, 2009, 2014, and 2016, respectively, report that they have bought services and/or goods in the hidden market during the previous two years. These are people who are involved in collusive tax evasion, as defined in the empirical strategy of the

[27]The income concept of Johns and Slemrod (2010) is different from the one used here, however.

[28]Although there is a danger that the non-random response rate contaminates results, note that there is little empirical support for the notion that a low response rate necessarily produces biased estimates (Groves, 2006).

*Table 1: Descriptive Statistics, Supply Side and Demand Side Data*

| | Supply side (expenditure approach) | | | | | |
|---|---|---|---|---|---|---|
| | Non-evaders | | Self-employed | | Evading salary workers | |
| | Mean | Std.dev. | Mean | Std.dev. | Mean | Std.dev. |
| Log net income | 13.27 | 0.44 | 13.22 | 0.52 | 13.29 | 0.35 |
| Log perm. net income | 13.29 | 0.42 | 13.26 | 0.46 | 13.30 | 0.34 |
| Log food expenditures | 10.87 | 0.66 | 10.96 | 0.59 | 10.95 | 0.58 |
| Age | 45.17 | 10.87 | 48.53 | 10.96 | 43.54 | 10.30 |
| Number of adults | 1.83 | 0.55 | 1.93 | 0.49 | 2.05 | 0.55 |
| Number of children | 1.30 | 1.17 | 1.32 | 1.19 | 1.43 | 1.19 |
| House size ($m^2$) | 135.89 | 57.20 | 155.00 | 71.75 | 141.75 | 59.81 |
| High education dummy | .48 | 0.50 | 0.36 | 0.48 | 0.12 | 0.33 |
| Number of cars | 1.34 | 0.73 | 1.43 | 0.85 | 1.53 | 0.76 |
| | | | | | | |
| No of observations | 4978 | | 414 | | 787 | |
| High income, high age | | | 126 | | 221 | |
| Low income, high age | | | 110 | | 113 | |
| High income, low age | | | 102 | | 253 | |
| Low income, low age | | | 76 | | 200 | |

| | Demand side (probit estimation) | |
|---|---|---|
| | Mean | Std.dev. |
| Binary for purchases | 0.12 | 0.33 |
| Log gross income | 13.27 | 0.61 |
| Age | 47.97 | 15.57 |
| Male | 0.52 | 0.50 |
| Self-employed | 0.04 | 0.18 |
| Eastern region dummy | 0.17 | 0.37 |
| | | |
| Observations | 6303 | |

Note: In the supply side sample, age refer to the age of the main income earner.

*Table 2: Estimation Results for the Expenditure Approach. Pooled Consumer Expenditure Data for 2003–2009 and 2012*

| | OLS, annual | IV, annual | OLS, permanent | IV, permanent |
|---|---|---|---|---|
| Slope, $\beta$ | 0.283 (12.44)*** | 0.405 (3.08)*** | 0.343 (13.20)*** | 0.433 (3.13)*** |
| Self-employed, $\delta_{SE}$ | 0.040 (1.57) | 0.051 (1.80)* | 0.043 (1.70)* | 0.051 (1.79)* |
| Salary worker, $\delta_{SW}$ | 0.008 (0.44) | 0.010 (0.54) | 0.012 (0.66) | 0.015 (0.77) |
| Implied $\overline{k}_{SE}$ | 1.15 (2.25) | 1.14 (3.72)* | 1.16 (3.42)* | 1.15 (4.78)** |
| Implied $\overline{k}_{SW}$ | 1.03 (0.25) | 1.03 (0.42) | 1.03 (0.22) | 1.03 (0.31) |
| $R^2$ | 0.388 | 0.398 | 0.392 | 0.390 |
| 1st stage F-statistic | | 69.45 | | 69.98 |
| Sargan ($p - value$) | | 0.371 | | 0.594 |
| Wu-H ($p - value$) | | 0.316 | | 0.476 |
| No of observations | 6,175 | 6,175 | 6,179 | 6,179 |

$p < 0.1^{*}$ $p < 0.05^{**}$ $p < 0.01^{***}$

Notes: The t-statistics and chi-square statistics in parentheses for regression coefficients and $\overline{k}$ values, respectively. Chi-square statistics for $\overline{k}$ are based on the delta method, accounting for variance in $\widehat{\beta}$ and $\widehat{\delta}$, under the null hypothesis that $k = 1$. Regressions include controls for age, age squared, size of house, dummy for higher education, and number of adults and children. Excluded instruments in IV regressions: dummy for owning a car and dummy for owning two or more cars.

*Table 3: Estimation Results for the Expenditure Approach, Including Interactions for Income Level and Age. Pooled Consumer Expenditure Data for 2003–2009 and 2012*

| | OLS, annual | IV, annual | OLS, permanent | IV, permanent |
|---|---|---|---|---|
| Slope, $\beta$ | 0.283 (12.11)*** | 0.394 (3.08)*** | 0.347 (12.91)*** | 0.419 (3.12)*** |
| Self-employed, $\delta_{SE}$ | -0.062 (-0.93) | -0.021 (-0.25) | –0.043 (–0.66) | -0.019 (-0.23) |
| High-inc, SE, $\lambda_{SE,HI}$ | 0.101 (1.21) | 0.048 (0.45) | 0.079 (0.95) | 0.047 (0.46) |
| High-age, SE, $\lambda_{SE,HA}$ | 0.199 (2.50)** | 0.191 (2.35)** | 0.198 (2.51)** | 0.195 (2.45)** |
| High-inc/age, SE, $\lambda_{SE,HI,HA}$ | -0.218 (-2.17)** | -0.208 (-2.04)** | –0.227 (-2.27)** | -0.225 (-2.25)** |
| Salary worker, $\delta_{SW}$ | -0.009 (-0.22) | 0.005 (0.11) | 0.003 (0.07) | 0.013 (0.28) |
| High-inc, SW, $\lambda_{SW,HI}$ | 0.031 (0.64) | 0.007 (0.12) | 0.014 (0.29) | -0.002 (-0.04) |
| High-age, SW, $\lambda_{SW,HA}$ | 0.030 (0.49) | 0.037 (0.60) | 0.034 (0.55) | 0.040 (0.61) |
| High-inc/age, SW, $\lambda_{SW,HI,HA}$ | -0.052 (-0.73) | -0.055 (-0.78) | –0.049 (–0.69) | -0.050 (-0.70) |
| Implied $\overline{k}_{SE,HI,HA}$ | 1.06 (0.17) | 1.01 (0.02) | 1.03 (0.05) | 1.00 (0.00) |
| Implied $\overline{k}_{SE,LI,HA}$ | 1.62 (5.89)** | 1.54 (8.35)*** | 1.60 (8.37)*** | 1.56 (9.83)*** |
| Implied $\overline{k}_{SE,HI,LA}$ | 1.15 (0.52) | 1.07 (0.24) | 1.12 (0.48) | 1.02 (0.05) |
| Implied $\overline{k}_{SE,LI,LA}$ | 0.78 (1.37) | 0.92 (0.14) | 0.89 (0.45) | 0.96 (0.04) |
| Implied $\overline{k}_{SW,HI,HA}$ | 1.00 (0.00) | 0.99 (0.04) | 1.00 (0.00) | 0.99 (0.01) |
| Implied $\overline{k}_{SW,LI,HA}$ | 1.08 (0.21) | 1.12 (0.74) | 1.11 (0.51) | 1.12 (0.95) |
| Implied $\overline{k}_{SW,HI,LA}$ | 1.08 (0.59) | 1.03 (0.17) | 1.04 (0.24) | 1.02 (0.05) |
| Implied $\overline{k}_{SW,LI,LA}$ | 0.97 (0.05) | 1.01 (0.01) | 1.00 (0.00) | 1.02 (0.04) |
| $R^2$ | 0.388 | 0.386 | 0.393 | 0.392 |
| 1st stage F-statistic | | 76.81 | | 78.96 |
| Sargan ($p-value$) | | 0.376 | | 0.593 |
| Wu-H ($p-value$) | | 0.348 | | 0.558 |
| No of observations | 6,175 | 6,175 | 6,179 | 6,179 |

$p < 0.1$* $p < 0.05$** $p < 0.01$*** LI=low income, HI=high income, LA=low age, HA=high age

Notes: The t-statistics and chi-square statistics in parentheses for regression coefficients and $\overline{k}$ values, respectively. Chi-square statistics for $\overline{k}$ are based on the delta method, accounting for variance in $\widehat{\beta}$ and $\widehat{\delta}$, under the null hypothesis that $k = 1$. Regressions include controls for age, age squared, size of house, dummy for higher education, and number of adults and children. Excluded instruments in IV regressions: a dummy for owning a car and a dummy for owning two or more cars.

present study.[29] The lower panel of Table 1 shows descriptive statistics for the sample used in the demand-side estimation, and we see that, overall, approximately 12% have been involved in tax evasion. In three of the surveys (not the 2016 survey), respondents are asked about the amount spent on hidden services, although with respect to relatively wide expenditure intervals. In Table C1 of Appendix C, we have pooled the information from three of the surveys and show how expenditure correlates with household income. The table shows that a clear majority declare that they have spent less than 50,000 Norwegian kroner (NOK) (over the last two years), but the overall picture is that expenditure increases in income.

As seen in Appendix C, the amounts spent by the buyers in the hidden market are reported for wide intervals. As one of the surveys (the 2016 survey) additionally does not include this type of information at all, we instead use the relationship with the supply side to establish demand-side evaded amounts, as already discussed. This means that we make an assumption regarding an Engel curve for hidden expenditure (which will be varied to test for robustness) and use hidden market participation and other characteristics (including income) to place the evaders in the income distribution. By this empirical strategy we also enforce that the amounts evaded on the supply side (obtained by the expenditure approach) equal the hidden payments on the demand side. In the next section we return to the practical implementation of this condition, and what it means in terms of distributional effects.

To obtain information about the determinants for being involved in non-

[29] Of course, that does not rule out the possibility that there might be people on the buyers' side who are innocently involved in tax evasion, i.e., the supplier unilaterally decides not to report the VAT that has been paid.

recorded purchases, we used a pooled dataset consisting of information from all four surveys, and estimated a probit model,

$$\Pr\left(c_{hj} = 1 \mid Z_j\right), \tag{11}$$

where $c_{hj} = 1$ for individual $j$ if he/she reports having paid for services in the hidden market, i.e., have $c_{hj} > 0$, where $Z_j$ symbolises control variables.

The estimation results reported in Table 4 show that income, as well as age, gender and region, are significant explanatory characteristics for the probability of participating in the hidden market. The probability of being in the informal market increases with income, although at a decreasing rate, as indicated by the square term.[30] Further, it decreases with age, in that males have a higher probability than females, and location matters (people in the eastern part of Norway, including in the capital of Oslo, have a higher probability of being involved in hidden transactions). The positive relationship with income suggests that informal markets may contribute to higher "real" income inequality. Of course, this can be explained by the rich being able to purchase goods and services for which there are informal markets, while also having a stronger preference for non-compliance.

To illustrate how these estimates are used to compute tax-evasion-corrected incomes, in Table 5 we show probabilities for three different household income levels, allowing for differences across gender.[31] For example, we see that the probability of entering the informal market increases from 0.11 to

[30] The estimated effect of income on the probability of participating in the informal economy remains positive until income reaches NOK 3.6 million.

[31] Non-significant explanatory variables are not used in this calculation, and the age and the region parameters are set to their average.

*Table 4: Probit Estimation for Buying Services in the Hidden Market. Pooled Data, 2006, 2009, 2014 and 2016*

| | Coefficient |
|---|---|
| Gross household income | 0.0005 (4.05)*** |
| Gross household inc. squared | -1.41x10$^{-7}$ (-2.46)** |
| Age | -0.004 (-2.91)*** |
| Male | 0.112 (2.71)*** |
| Self-employed | 0.069 (0.65) |
| Eastern region dummy | 0.187 (3.60)*** |
| Constant | -1.350 (-14.83)*** |
| Likelihood ratio | 58.37 |
| No. of observations | 6,303 |

$p < 0.1$* $p < 0.05$** $p < 0.01$***

Notes: Income measured in 1,000 Norwegian kroner. z-statistics reported in parentheses

*Table 5: Probability of Buying Services in the Hidden Market for Three Gross Household Income Levels*

| NOK 200,000 | | NOK 800,000 | | NOK 1,400,000 | |
|---|---|---|---|---|---|
| Male | Female | Male | Female | Male | Female |
| 0.108 | 0.080 | 0.190 | 0.161 | 0.249 | 0.214 |

Note: Income measured in 2012 values

0.25 when the income of the household goes up from NOK 0.2 million to NOK 1.4 million.

As the supply-side estimates have been obtained for households, whereas the estimates in Table 4 have been derived at the individual level, we let the individual actions of the demand side represent a "household probability", an issue we return to in the next section.

# 4 Distributional Effects of Tax Evasion

## *4.1 Methodology*

Recall that a central objective of this exercise is to obtain an estimate of the "hidden-economy-controlled" income distribution, and see how it relates to the official one. We obtain an income distribution adjusted for sellers' unreported income by letting the disposable income of each evading household be adjusted by the relevant $k$. As discussed in Section 2, a description of the true income distribution should also reflect that some households face lower prices than they otherwise would because of collusive evasion.[32]

But, as discussed in the Introduction, and as made clear in the outline of the decision-making in Section 2, it is possible to adopt another approach in the discussion of distributional effects, namely by addressing information about the individual tax saved by not reporting income and transactions. This perspective brings the tax incidence challenge of the present analysis to the surface. Recall that we do not observe $p_r$ and $p_h$, which implies that we do not observe how the gains from tax evasion are divided between sellers and buyers. We shall assume, as a point of departure and to fix ideas, that the consumer price in the informal market is equal to the pre-VAT price in the regular market, $p_r = p_h$, and the individual gain is defined by indirect taxation, i.e. VAT. But in reality, for example, we may have a case where suppliers' returns are squeezed to the extent that the purchasers receive all or

[32] Following the earlier literature, in these calculations we ignore the private costs incurred in effecting the evasion, including but not limited to the expected fine due to detection and punishment. For the marginal act of evasion by the marginal evader, this cost would be equal to the expected marginal gain, but on average will be less than the expected gain, so that accounting for these costs would affect the results presented here.

most of the economic advantage, the saved income tax included. Of course, we can have the opposite situation as well. We will return to assumptions regarding the splitting of the economic advantage between the two sides of the market in the sensitivity analysis.

Equation (1) expresses the individual benefit of the evading supplier. Given that $p_r = p_h$, it follows that $\triangle_t = t^* - t$ can be used as a starting point, where $t^*$ is the income tax burden resulting from reporting all income truthfully, where the latter is defined by $x_r^* = x_r + \triangle_y$, where $\triangle_y = y^* - y$. Then Equation (2) defines the advantage of buyers, and when $p_r = p_h$, and given the present VAT rate in Norway, which is 25%, it implies that $(p_r(1+\pi) - p_h) / (p_r(1+\pi)) = 0.2$; the price in the hidden market is 20% below the price in the regular market.[33]

As shown in Section 1, earlier contributions in the literature attest to the possibilities of applying different perspectives on the distributional impact of tax evasion. For example, the results reported in Johns and Slemrod (2010) illustrate that one may obtain results that show that tax evasion increases with income, but which would still imply a more progressive tax schedule. In the following we shall describe the distributional effects from both angles; both as an adjustment to income, given that we are interested in what the distribution of "real" income looks like, and from a perspective where tax evasion results in reduced tax burdens, as highlighted by the discussion in Section 2.

We do this by piecing together the econometric evidence presented in

[33] In a price negotiation situation, the customer may be offered the choice to pay with or without VAT, thereby justifying the assumption of equality of pre-tax prices, at least as a point of departure.

Section 3. In order to translate the empirical findings of the previous section into distributional effects, we utilise the tax-benefit model LOTTE (Aasness *et al.*, 2007). The model is based on data from income tax returns and other administrative registers, which means that it contains information on annual income, family composition, number of children, education, etc. By utilising the fact that there are common variables in the data used to estimate supply side and demand side tax evasion and the tax-benefit model, we impute our estimates into a representative sample of the population. In addition to accounting for evasion behaviour on both sides of the market, this procedure is helpful for keeping track of the balance between supply and demand side amounts, as seen in Equation (3).

When imputing supply side unreported income, we use the OLS estimates for permanent income of Table 3. Note that although all estimates of $k$ are not statistically significantly different from each other ($\overline{k}_{SE,LI,HA}$ is significantly different from the others and from 1), we let $k$ differ with respect to point estimates. This means that income for households containing at least one self-employed person is adjusted according to the four estimates of $k$, depending on age and income. For high age, the income of the below-median-income households is adjusted by 1.60, whereas incomes above the median are adjusted by 1.03. Further, for low age, the figures are 1.12 and 1.00, for income above and below the median, respectively.[34] Correspondingly, an income correction has been allocated to evading salary workers, based on estimates of $k_{SW}$ in Table 3: the adjustments are 1.11 for low

[34]As seen in Table 3, $\overline{k}_{SE,LI,LA}$ is less than 1, but the large standard errors compel us to use 1 in this case.

income/high age and 1.04 for high income/low age.

An important component of the empirical framework is establishing an Engel demand curve for the hidden service. When expenditure is given by $e_h = p_h c_h$, we assume that the hidden economy Engel curve is

$$\ln e_h = (a + \varphi \ln y^*), \tag{12}$$

where $a$ is a constant, $y^*$ is true disposable household income, and $\varphi$ can be interpreted as income elasticity (hidden-good Engel elasticity). Then, when the estimation results of Section 3.3 are used, the expected expenditure on hidden market services of the household represented by buyer $j$ can be written,

$$\Pr(c_{hj} = 1 \mid Z_j) \exp(a + \varphi \ln y_j^*). \tag{13}$$

In the practical implementation, we face (at least) two empirical challenges when translating the empirical evidence into the dataset of the tax-benefit model: We must apply a procedure to assign the gains of hidden consumption to the individuals in the new dataset, and we must find a procedure for going from individual to household level (supply side). First, we assign demand-side individuals to the hidden market group by random draws, based on the characteristics of Table 4.[35] When there is only one adult in the household, the translation into the household level follows straightforwardly from this. In cases where there are two adults in the household, there is a similar joint probability of both being selected, based on the individual

[35] An alternative where all individual incomes are adjusted by average amounts (differentiated by characteristics) generates essentially the same results with respect to overall inequality as the chosen procedure (based on random draws).

probabilities.

Further, note that the size of the evaded amounts on the demand side is obtained by an equilibrium condition, Equation (3). Thus, we have

$$\sum_i x_{hi} = \sum_j \Pr\left(c_{hj} = 1 \mid Z_j\right) \exp(a + \varphi \ln y_j^*), \tag{14}$$

which in turn can be used to obtain an expression for $a$,

$$a = \ln\left(\frac{\sum_i x_{hi}}{\sum_j \Pr\left(c_{hj} = 1 \mid Z\right) y^{*\,\varphi}}\right), \tag{15}$$

for any given value of $\varphi$. As a point of departure, we set $\varphi$ at 1, which means that budget shares of informal consumption, conditional on some positive consumption, are constant along the income scale. This is in accordance with findings described in a report from a Danish expert group (De Økonomiske Råd, 2011).

## *4.2 Effects on Income Inequality and Distribution of Tax Burdens*

We begin, in Table 6, by showing how disposable income is to be adjusted due to the underreporting of sellers' income.[36] The table shows that true post-tax income is on average NOK 8,254 higher than reported income. In total, this corresponds to post-tax income being NOK 19.8 billion higher, if income were correctly measured. The overall picture is that the rate of

[36] We rank the households according to their "true" income. If we used the reported income concept, the tax evaders would be placed lower in the distribution than they actually are, as discussed in Johns and Slemrod (2010).

*Table 6: Distribution of Tax-Evasion-Adjusted Income of Suppliers. Household Post-Tax Income, 2012*

| Decile | Share of self-employed (%) | Reported post-tax income | Supply side tax evasion adjusted post-tax income | Reported income as share of adj. post-tax inc. (%) |
|---|---|---|---|---|
| 1 | 3.74 | 95,921 | 96,204 | 99.71 |
| 2 | 1.93 | 192,589 | 193,001 | 99.79 |
| 3 | 2.77 | 246,985 | 248,125 | 99.54 |
| 4 | 4.03 | 311,494 | 314,113 | 99.17 |
| 5 | 4.54 | 382,843 | 386,777 | 98.98 |
| 6 | 10.12 | 464,627 | 473,088 | 98.21 |
| 7 | 14.97 | 557,818 | 571,301 | 97.64 |
| 8 | 15.17 | 664,432 | 677,744 | 98.04 |
| 9 | 17.68 | 799,260 | 812,586 | 98.36 |
| 10 | 26.41 | 1,229,824 | 1,255,389 | 97.96 |
| All | 10.14 | 494,579 | 502,833 | 98.36 |
| | | | | |
| Gini coeff. | | 0.3642 | 0.3656 | |

Notes: Households ranked by tax-evasion-adjusted post-tax income in 2012. All values refer to decile mean.

underreporting increases with income. As shown in the last column of Table 6, the underreporting rates are higher in decile groups 6 to 10 than in decile groups 1 to 5. Even though we found more tax evasion (as measured by $k$) among both self-employed and wage earners at the low ends of their respective income distributions, we find that overall income inequality is higher in the tax-evasion-adjusted income distribution, as seen in the bottom row of Table 6. A key factor behind this result is the placement of the self-employed in the overall income distribution, as reported in the first column of the table, which shows that there are groups other than the self-employed that dominate the low end of the income distribution.

Next, in Table 7, we turn our attention to the reduced tax burdens among

*Table 7: Distributions of Reduced Tax Burden for Suppliers and Consumers Participating in the Hidden Market*

| Decile | Supplier gain ($SG$) | Supplier gain, income share (%) | Consumer gain ($CG$) | Consumer gain, income share (%) |
|---|---|---|---|---|
| 1 | 75 | 0.10 | 164 | 0.23 |
| 2 | 137 | 0.07 | 297 | 0.16 |
| 3 | 383 | 0.16 | 433 | 0.18 |
| 4 | 924 | 0.30 | 668 | 0.22 |
| 5 | 1441 | 0.38 | 1019 | 0.27 |
| 6 | 3207 | 0.68 | 1513 | 0.32 |
| 7 | 5218 | 0.92 | 2210 | 0.39 |
| 8 | 5295 | 0.78 | 3195 | 0.47 |
| 9 | 5560 | 0.69 | 4309 | 0.53 |
| 10 | 10947 | 0.87 | 6826 | 0.54 |
| All | 3319 | 0.67 | 2063 | 0.41 |
| | | | | |
| Gini coeff. | 0.9168 | | 0.9203 | |

Notes: Households ranked by collusive tax-evasion-adjusted post-tax income in 2012. All values refers to decile mean. Engel elasticity and price assumption: $\phi = 1$, $p_r = p_h$. Income shares as share of collusive-tax-evasion-adjusted post-tax income.

sellers and buyers from participating in the hidden market. Thus, for suppliers, non-reported income that forms the basis for the income distributions reported in Table 6 is taxed, and the distribution of the tax burden depending on whether non-reported income is taxed or not ($SG$) is shown in the first column of Table 7. For the consumers, as discussed in Section 4.1, the reduced tax burden (the consumer gain, $CG$) is equal to the indirect taxation of the hidden consumption (say the VAT), given the assumption that the pre-tax prices of the hidden and the regular market are equal, $p_r = p_h$, and the reduced tax burden due to buying in the hidden market is equal to the difference in after-tax prices ($p_r(1 + \pi)$ vs $p_h$). As explained in Section 2.2, our approach implies that the gains in terms of reduced prices are exactly equal to the reduced tax burden for the consumers (see footnote 8).

Table 7 shows that the average size of $CG$ is smaller than the average of $SG$, which follows from the difference between taxation of income (supply side) and taxation of consumption (demand side). As expected, given the results of Table 4 and Table 5, tax evasion on the demand side also generates a more unequal "true" income distribution; see the last column of the table, where consumer gain is described in terms of income share. Table 7 shows that the reductions in tax burdens are fairly modest, in both absolute and relative terms, among both suppliers and consumers in deciles 1–3. At the top of the income distribution, by contrast, the gains for both sides are large.

The joint supply and demand side effect of tax evasion on income distribution is summarised in Table 8, which shows the distribution of tax-evasion-adjusted income, accounting for evasion at both sides of the market, and compared to an income distribution based on a conventional definition. Thus, the post-tax income of Table 6 has been added to the price gain of consumers in Table 7, and turned into an overall tax-evasion-corrected measure of disposable income. We see that the Gini coefficient for this income concept is 0.366, which is higher than the Gini coefficient income inequality for conventional income of 0.364. Thus, our method suggests that the "hidden-economy-controlled" income inequality is higher than the official one.

Finally, in Table 9 we show how the distribution of tax revenue is affected by tax evasion, accounting for effects through both direct and indirect taxation. As shown, the compliance rate decreases with income, implying that the tax schedule in reality is less progressive than shown by official figures. The mean compliance rate for the whole population is 97.96%, which correponds to a revenue loss of approximately NOK 12 billion, in total. The

*Table 8: Conventional Income Distribution Compared to Income Distribution Controlled for Tax Evasion of Suppliers and Consumers. Household Post-Tax Income, 2012*

| Decile | Reported post-tax income | Post-tax income adjusted by collusive tax evasion | Reported as share of adjusted post-tax income (%) |
|---|---|---|---|
| 1 | 95,934 | 96,379 | 99.54 |
| 2 | 192,600 | 193,308 | 99.63 |
| 3 | 246,991 | 248,565 | 99.37 |
| 4 | 311,508 | 314,809 | 98.95 |
| 5 | 382,814 | 387,821 | 98.71 |
| 6 | 464,709 | 474,647 | 97.91 |
| 7 | 557,900 | 573,581 | 97.27 |
| 8 | 664,454 | 680,985 | 97.57 |
| 9 | 799,271 | 816,879 | 97.84 |
| 10 | 1,229,612 | 1,261,989 | 97.43 |
| All | 494,579 | 504,896 | 97.96 |
| | | | |
| Gini coeff. | 0.3642 | 0.3662 | |

Notes: Households ranked by collusive tax-evasion-adjusted post-tax income in 2012. All values refer to decile mean. Engel elasticity and price assumption: $\phi = 1$, $p_r = p_h$.

VAT loss is calculated to be about NOK 4 billion,[37] whereas the tax revenue from the personal income tax would have been approximately NOK 8 billion higher if the self-employed (and wage-earning craftsmen) reported all their (true) income.

## *4.3 Sensitivity Analysis: Gain Splitting and the Slope of the Engel Curve*

Our quantitative descriptions of the distributional effects of joint tax evasion depend on some fairly strong assumptions. To test how robust our results are with respect to the main assumptions of the framework, we consider

[37]We recognize that the loss in indirect tax revenue may be somewhat overstated, as there may be some remittance of VAT in earlier stages of the production process.

*Table 9: Distribution of the Tax Compliance Rate, 2012*

| Decile | Actual tax revenue | Full compliance tax revenue | Compliance rate (%) |
|---|---|---|---|
| 1 | 43,114 | 43,321 | 99.52 |
| 2 | 71,152 | 71,536 | 99.46 |
| 3 | 111,507 | 112,316 | 99.28 |
| 4 | 147, 564 | 148, 832 | 99.15 |
| 5 | 172,250 | 174,362 | 98.79 |
| 6 | 210,169 | 214,997 | 97.75 |
| 7 | 261,514 | 268,655 | 97.34 |
| 8 | 324,079 | 332,172 | 97.56 |
| 9 | 420,220 | 429,196 | 97.91 |
| 10 | 776,726 | 792,611 | 98.00 |
| All | 253,518 | 258,493 | 98.08 |

Notes: Households ranked by collusive tax-evasion-adjusted post-tax income in 2012. All values refer to decile mean. Engel elasticity and price assumption: $\phi = 1$, $p_r = p_h$. Revenue effects from indirect and direct taxation are derived from the tax-benefit model LOTTE and national accounts data (effective indirect tax rates). The marginal propensity to consume is set to 0.7 for all income groups.

alternative scenarios. As discussed in Section 4.1, we would like to see if the main results survive under different assumptions concerning the split of the economic gain between sellers and buyers and with respect to our assumption about the slope of the hidden economy Engel curve ($\varphi$). Note that the economic gain split is regulated by how the non-tax price of the hidden market, $p_h$, relates to the before-tax price of the regular market, $p_r$. As supplier gain is defined by the (hypothetical) income tax ($x_r^* \tau$) on non-reported income, if the price in the hidden market is low and falls short of the price in the regular market (before tax) ($p_h < p_r$), $x_r^*$ decreases, and more of the gain is transferred to the buyers. In principle, the buyers may obtain the whole advantage (or, vice versa, the whole gain may go to the suppliers).

As in the descriptions of distributional effects seen so far, we shall describe the effects both in terms of supplier and consumer gains, $SG$ and $CG$, and in terms of the effects on income distribution, in Table 10 and Table 11, respectively. As an alternative to $\varphi = 1$, we set $\varphi = 0$, i.e., assuming that the amount of hidden purchases is independent of income. However, note that the empirical approach still allows the probability of purchasing in the hidden market depending on income, as in Table 4. Further, in addition to the approach used so far, which assumes that consumer prices in the hidden market are 20% lower than in the regular market, corresponding to $p_h = p_r$, results for 10 and 30% lower prices are presented in Table 10, whileTable 11 also shows results for 0% and 40% lower prices in the hidden market. The figures reflect that when the gain of the supplier in the hidden market falls short of the pre-VAT price in the regular market (for example, for a 30% price reduction in Table 10), the total amount of income he receives is less than the pre-tax income of the regular sector, and the supplier side gain is reduced accordingly.

The change to $\varphi = 0$ implies that the distributional gains of consumers move to lower parts of the income distribution. This can be seen by comparing $CG$ of the middle alternative in Table 10 with the results reported in Table 7. When this is combined with a large price reduction in the hidden market (30%), consumer gains increase and, in particular, gains are larger in the lower part of the distribution.

However, Table 11 shows that, for all alternatives, income inequality is higher than the income inequality of conventional income, the latter reported in Table 6. The change to $\varphi = 0$ reduces the overall income inequality from 0.366 (see Table 8) to 0.365 (for $p_h = p_r$, or a 20% lower price). The income

*Table 10: Economic Gain as Share of Tax-Evasion-Adjusted Income under Different Consumer Price Reduction Assumptions.* $\varphi = 0$

| | Price reduction: 10% | | Price reduction: 20% | | Price reduction: 30% | |
|---|---|---|---|---|---|---|
| Decile | *SG* (%) | *CG* (%) | *SG* (%) | *CG* (%) | *SG* (%) | *CG* (%) |
| 1 | 0.13 | 0.63 | 0.10 | 1.30 | 0.06 | 1.99 |
| 2 | 0.09 | 0.23 | 0.07 | 0.48 | 0.05 | 0.69 |
| 3 | 0.20 | 0.21 | 0.16 | 0.46 | 0.11 | 0.82 |
| 4 | 0.37 | 0.20 | 0.30 | 0.44 | 0.20 | 0.73 |
| 5 | 0.46 | 0.20 | 0.38 | 0.45 | 0.26 | 0.77 |
| 6 | 0.83 | 0.20 | 0.68 | 0.44 | 0.51 | 0.75 |
| 7 | 1.09 | 0.20 | 0.92 | 0.44 | 0.69 | 0.75 |
| 8 | 0.94 | 0.20 | 0.78 | 0.45 | 0.58 | 0.77 |
| 9 | 0.82 | 0.18 | 0.69 | 0.42 | 0.52 | 0.75 |
| 10 | 1.02 | 0.12 | 0.87 | 0.29 | 0.69 | 0.51 |
| All | 0.79 | 0.18 | 0.67 | 0.42 | 0.50 | 0.71 |

Notes: Households sorted by collusive tax-evasion-adjusted post-tax income in 2012. All values refer to decile mean.

inequality is fairly insensitive to the incidence assumption, except that we see that inequality rises to 0.366 when there is no price reduction (i.e., the producers retain the whole benefit).

Thus, all of the estimates of evasion-controlled income inequality are somewhat higher than the income inequality arrived at using standard income concepts. The clear pro-rich pattern of the suppliers is a principal explanation. However, this conclusion cannot automatically be transferred to the case of tax savings distribution, where the gains split between suppliers and consumers becomes particularly decisive. If the consumers obtain most to the gain (30% price reduction), and the expenditure function is constant with respect to income ($\varphi = 0$), the sizeable gains of the consumers at the low end of the income distribution may dominate; see Table 10. However, for most alternatives of the tax savings perspective, we see distributional effects that point to tax evasion benefitting the rich more than the poor.

*Table 11: Reported Post-Tax Income as Share of Tax Evasion Adjusted Post-Tax Income under Different Consumer Price Reduction Assumptions.* $\varphi = 0$

| | Lower price in the hidden market (%) | | | | |
|---|---|---|---|---|---|
| Decile | 0 | 10 | 20 | 30 | 40 |
| 1 | 99.71 | 99.23 | 98.74 | 98.23 | 97.74 |
| 2 | 99.79 | 99.56 | 99.32 | 99.12 | 99.01 |
| 3 | 99.54 | 99.34 | 99.09 | 98.74 | 98.28 |
| 4 | 99.16 | 98.97 | 98.74 | 98.45 | 98.10 |
| 5 | 98.97 | 98.78 | 98.53 | 98.22 | 97.81 |
| 6 | 98.22 | 98.03 | 97.80 | 97.50 | 97.11 |
| 7 | 97.64 | 97.45 | 97.22 | 96.92 | 96.53 |
| 8 | 98.03 | 97.84 | 97.59 | 97.28 | 96.87 |
| 9 | 98.36 | 98.19 | 97.95 | 97.63 | 97.18 |
| 10 | 97.96 | 97.84 | 97.68 | 97.47 | 97.15 |
| All | 98.36 | 98.18 | 97.96 | 97.67 | 97.29 |
| Adj. post-tax income, Gini | 0.3656 | 0.3654 | 0.3652 | 0.3649 | 0.3647 |

Notes: Households sorted by collusive tax-evasion-adjusted post-tax income in 2012. All values refer to decile mean.

# 5 Summary

The distributional effect of tax evasion depends on the income profile of both buyers and sellers, as well as how the gain resulting from evasion is shared between the two sides of the market. In this paper we offer a quantitative assessment of this question for Norway by evaluating new data from both sides of the market. To learn about the suppliers of informal goods and services, we use the expenditure method. To evaluate the consumer side, we utilise sample survey information. Finally, we incorporate these two sources of information into a comprehensive framework that imposes consistency between supply and demand and which allows us to perform robustness checks concerning the shifting parameter that we do not observe. This type of analysis has not been attempted before for any country.

We find that accounting for the hidden economy in this comprehensive way provides an estimate of the income inequality in Norway that is higher than the official one, with the income profile distribution of informal suppliers contributing the larger share of the change in measured inequality. Thus, the usual omission of the demand side effects is not crucial, according to the present study. The compliance rate of suppliers decreases with income, implying that effective income tax progressivity is less than that indicated by official figures.

We recognise that these conclusions are provisional, as both the data and the methods of inference from the data are imperfect, and we do not offer these conclusions as the final word on this issue. We do argue, however, that our methodology shows that this kind of comprehensive approach holds promise of shedding light on the distributional effects of the informal economy.

## *Affiliation*

Nygård: Research Department, Statistics Norway, PO box 8131, Dep. NO-0033 Oslo, Norway (e-mail: oen@ssb.no); Slemrod: Stephen M. Ross School of Business and Department of Economics, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109 (e-mail: jslemrod@umich.edu); Thoresen: Research Department, Statistics Norway, PO box 8131, Dep. NO-0033 Oslo, Norway, and Oslo Fiscal Studies, Department of Economics, University of Oslo (e-mail: tot@ssb.no).

# References

Aasness, J., Dagsvik, J.K. and Thoresen, T.O. (2007). 'The Norwegian tax-benefit model system LOTTE', in (A. Gupta and A. Harding, eds.), *Modelling Our Future: Population Ageing, Health and Aged Care, International Symposia in Economic Theory and Econometrics*, pp. 513–518, Amsterdam: Elsevier.

Abraham, M., Lorek, K., Richter, F. and Wrede, M. (2016). 'Collusive tax evasion and social norms', *International Tax and Public Finance*, vol. 24(2), pp. 179–197.

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

Alm, J. (2012). 'Measuring, axplaining, and controlling tax evasion: lessons from theory, experiments, and field studies', *International Tax and Public Finance*, vol. 19(1), pp. 54–77.

Alm, J. and Finlay, K. (2013). 'Who benefits from tax evasion', *Economic Analysis and Policy*, vol. 43(2), pp. 139–154.

Alm, J. and Sennoga, E. (2010). 'Mobility, competition, and the distributional effects of tax evasion', *National Tax Journal*, vol. 63(4), pp. 1055–1084.

Andreoni, J., Erard, B. and Feinstein, J. (1998). 'Tax compliance', *Journal of Economic Literature*, vol. 36(2), pp. 818–60.

Alstadsæter, A., Johannesen, N. and Zucman, G. (2017). 'Tax evasion and inequality', NBER Working Paper 23772, Cambridge (MA).

Banks, J., Blundell, R. and Lewbel, A. (1997). 'Quadratic Engel curves and consumer demand', *Review of Economics and Statistics*, vol. 79(4), pp. 527–539.

Benedek, D. and Lelkes, O. (2011). 'The distributional implications of income under-reporting in Hungary', *Fiscal Studies*, vol. 32(4), pp. 539–560.

Bishop, J.A., Formby, J.P. and Lambert, P. (2000). 'Redistribution through the income tax: the vertical and horizontal effects of noncompliance and tax evasion', *Public Finance Review*, vol. 28(4), pp. 335–350.

Boadway, R., Marceau, N. and Mongrain, S. (2002). 'Joint tax evasion', *Canadian Journal of Economics*, vol. 35(3), pp. 417–35.

Chang, J.-j. and Lai, C.-c. (2004). 'Collaborative tax evasion and social norms: why deterrence does not work', *Oxford Economic Papers*, vol. 56(2), pp. 344–368.

Christian, C.W. (1994). 'Voluntary compliance with the individual income tax: results from the 1988 TCMP study', IRS Research Bulletin, 1993/1994, Publication 1500, 35–42.

Cremer, H. and Gahvari, F. (1993). 'Tax evasion and optimal commodity taxation', *Journal of Public Economics*, vol. 50(2), pp. 261–275.

De Økonomiske Råd (2011). 'Dansk økonomi, efterår 2011', Copenhagen. (in Danish, with English summary)

Dhami S. and al-Nowaihi, A. (2007). 'Why do people pay taxes? Prospect theory versus expected utility theory', *Journal of Economic behaviour & Organization*, vol. 64(1), pp. 171–192.

Dunbar, G.R. and Fu, C. (2015). 'Sheltered income: Estimating income under-reporting in Canada, 1998 and 2004', Bank of Canada Working Paper No. 2015-22, Ottawa.

Eide, E., von Simson, K. and Strøm, S. (2011). 'Rank-dependent utility, tax evasion and labor supply', *FinanzArchiv/Public Finance Analysis*, vol. 67(3), pp. 261–281.

Engström, P. and Holmlund, B. (2009). 'Tax evasion and self-employment in a high-tax country: Evidence from Sweden', *Applied Economics*, vol. 41(19), pp. 2419–2430.

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

Feldman, N.E. and Slemrod, J. (2007). 'Estimating tax noncompliance from unaudited tax returns', *The Economic Journal*, vol. 117(518), pp. 327–352.

Fiorio, C.V. and D'Amuri, F. (2005). 'Workers' tax evasion in Italy', *Giornale degli Economisti e Annali di Economia,* vol. 64(2-3), pp. 247–270.

Fortin, B., Lacroix, G. and Pinard, D. (2010). 'Evaluation of the underground economy in Quebec: a microeconomic approach', *International Economic Journal*, vol. 24(4), pp. 463–479.

Groves, R.M. (2006). ‘Nonresponse rates and nonresponse bias in household surveys’, *Public Opinion Quarterly*, vol. 70(5), pp. 646–675.

Holmøy, A. and Lillegård, M. (2014). ‘Forbruksundersøkelsen 2012, dokumentasjonsrapport’, Notater 2014/17, Statistics Norway. (in Norwegian)

Hurst, E., Li, G. and Pugsley, B. (2014). ‘Are household surveys like tax forms: Evidence from income underreporting of the self-employed’, *Review of Economics and Statistics*, vol. 96(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.

Johns, A. and Slemrod, J. (2010). ‘The distribution of income tax noncompliance’, *National Tax Journal*, vol. 63(3), pp. 397–418.

Kesselman, J.R. (1989). ‘Income tax evasion: an intersectoral analysis’, *Journal of Public Economics*, vol. 38(2), 137–182.

Kim, B., Gibson, J. and Chung, C. (2017). ‘Using panel data to estimate income under-reporting by the self-employed’, *The Manchester School*, vol. 85(1), pp. 41–64.

Kleven, H.J., Knudsen, M., Kreiner, C.T., Pedersen, S.L. and Saez, E. (2011). ‘Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark’, *Econometrica*, vol. 79(3), 651–692.

Kleven, H.J., Kreiner, C.T. and Saez, E. (2016). ‘Why can modern governments tax so much? An agency model of firms as fiscal intermediaries’, *Economica*, vol. 83(330), 219–246.

KRISINO (2011). ‘Kriminalitets- og sikkerhetsundersøkelsen i Norge’, Næringslivets sikkerhetsråd (in Norwegian).

Lyssiotou, P., Pashardes, P. and Stengos, T. (2004). ‘Estimates of the black economy based on consumer demand approaches’, *The Economic Journal*, vol. 114(497), pp. 622–640.

Martinez-Lopez, D. (2013). ‘The underreporting of income by self-employed workers in Spain’, SERIEs, vol. 4(4), pp. 353–371.

Matsaganis, M. and Flevotomou, M. (2010). ‘Distributional implications of tax evasion in Greece’, GreeSE Paper No 31, Hellenic Observatory Papers on Greece and Southeast Europe, LSE, London.

Ognedal, T. (2016). ‘Morality in the market’, *Journal of Economic Behaviour & Organization*, vol. 129, pp. 100–115.

Opinion (2006). Privates kjøp av svart arbeid, Opinion AS. (in Norwegian)

Opinion (2014). Undersøkelse om svart arbeid, Opinion AS. (in Norwegian)

Opinion (2016). Undersøkelse om svart arbeid, Opinion AS. (in Norwegian)

Pashardes, P. and Polycarpoub, A. (2008). ‘Income tax evasion, inequality and poverty’, *Cyprus Economic Policy Review*, vol. 2(2), pp. 37–49.

Paulus, A. (2015). ‘Income underreporting based on income-expenditure gaps: Survey vs tax records’, ISER Working Paper Series, No. 2015-15, University of Essex, Colchester.

Persson, M. and Wissén, P. (1984). 'Redistributional aspects of tax evasion', *The Scandinavian Journal of Economics*, vol. 86(2), pp. 131–149.

Pissarides, C.A. and Weber, G. (1989). 'An expenditure-based estimate of Britain's black economy', *Journal of Public Economics*, vol. 39(1), pp. 17–32.

Schneider, F. and Enste, D.H. (2000). 'Shadow economies: Size, causes, and consequences', *Journal of Economic Literature*, vol. 38(1), pp. 77–114.

Schuetze, H. (2002). 'Profiles of tax non-compliance among the self-employed in Canada: 1969–1992', *Canadian Public Policy*, vol. 28(2), pp. 219–238.

Slemrod, J. (2007). 'Cheating ourselves: The economics of tax evasion', *Journal of Economic Perspectives*, vol. 21(1), pp. 25–48.

Slemrod, J. and Yitzhaki, S. (2002). 'Tax avoidance, evasion, and administration', in (A.J. Auerbach and M.S. Feldstein, eds.), *Handbook of Public Economics*, Volume 3, 1423–70, Amsterdam: Elsevier.

Slemrod, J. and Weber, C. (2012). 'Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy', *International Tax and Public Finance*, vol. 19(1), pp. 25–53.

Statistics Norway (2017). 'Income and wealth statistics for households', http:www.ssb.no/en/inntekt-og-forbruk/statistikker/ifhus.

Tedds, L.M. (2010). 'Estimating the income reporting function for the self-employed', *Empirical Economics*, 38(3), pp. 669–687.

TNS Gallup (2009). 'Kjøp av svart arbeid i privatmarkedet', Skattedirektoratet. (in Norwegian)

Torgler, B. (2007). *Tax Compliance and Tax Morale*, Cheltenham: Edward Elgar.

Wangen, K.R. (2005). 'An expenditure estimation of Britain's black economy revisited', Discussion Papers, No. 414, Statistics Norway.

Williams, C.C. (2008). *The Hidden Enterprise Culture: Entrepreneurship in the Underground Economy*, Cheltenham: Edward Elgar.

Yaniv, G. (1988). 'Withholding and non-withheld tax evasion', *Journal of Public Economics*, vol. 35(2), 183–204.

Yaniv, G. (1992). 'Collaborated employee-employer tax evasion', *Public Finance*, 47(2), 312–21.

Yitzhaki, S. (1974). 'A note on 'Income tax evasion: A theoretical analysis'', *Journal of Public Economics*, vol. 3(2), pp. 201–202.

# A Sample Defined by Matches to Permanent Income

As described in Section 3.2, we match the expenditure data to income data through the use of common background variables, since matching through national identification numbers cannot be done. The fact that we find matches for approximately 70% of the observations raises the question of to what extent the restricted sample is representative. Therefore, in Table A1, we show how the full sample and matched sample compare with respect to some key characteristics. As there is close correspondence between the characteristics of the two datasets, there is no reason to assume that the sample used in the estimation is biased as a result of of the matching procedure.

*Table A1: Descriptive Statistics. Estimation Sample vs. Sample Prior to Matching. Pooled Consumer Expenditure Data 2003–2009, 2012*

| | Sample prior to matching | | Estimation sample | |
|---|---|---|---|---|
| | Mean | Std.dev. | Mean | Std.dev. |
| Log net income | 13.24 | 0.45 | 13.27 | 0.43 |
| Log food expenditures | 10.82 | 0.65 | 10.89 | 0.65 |
| Number of adults | 1.90 | 0.56 | 1.87 | 0.55 |
| Number of children | 1.26 | 1.17 | 1.32 | 1.17 |
| Age, head of household | 44.15 | 11.23 | 45.20 | 10.85 |
| Size of house | 137.86 | 61.39 | 137.92 | 58.82 |
| Education | 0.41 | 0.49 | 0.42 | 0.49 |
| Number of cars | 1.36 | 0.75 | 1.37 | 0.74 |
| Observations | 8968 | | 6179 | |

# B Supplementary Tests

Using a method similar to that of Pissarides and Weber (1989), we test for non-linearities in the effect of income on expenditure by introducing the square of the log of income as an additional regressor. In Table B1 we report the estimation results for an OLS regression (permanent income as the income variable) when the square of log (permanent) income is also included among the regressors. As seen from the table, this yields non-significant coefficients of the income variables. Thus, these results do not provide support for a rejection of the assumed log-linear relationship. We have also tested for alternative relationships between expenditure and income by interacting

*Table B1: Estimation Results when Square of Log Income is Included in the Specification. Pooled Consumer Expenditure Data 2003–2009, 2012*

| | Coef. | SE | t-value |
|---|---|---|---|
| Slope, $\beta$ | 0.994 | 0.785 | 1.27 |
| Square of log income | -0.025 | 0.030 | -0.84 |
| Self-employed, $\delta_{SE}$ | 0.044* | 0.025 | 1.73 |
| Salary worker, $\delta_{SW}$ | 0.011 | 0.019 | 0.61 |
| Number of adults | 0.275*** | 0.017 | 16.08 |
| Number of children | 0.193*** | 0.007 | 26.65 |
| Age, head of household | 0.020*** | 0.005 | 4.13 |
| Age squared | -0.0002*** | 0.0001 | -2.88 |
| Size of house | 0.0004*** | 0.0001 | 2.86 |
| Education | 0.048*** | 0.014 | 3.41 |

$p < 0.1$* $p < 0.05$** $p < 0.01$***

income with a dummy variabel indicating high income. This variable does not prove to have a significant coefficient either.

Further, Table B2 reports estimates obtained by running regressions separately for each year. The table illustrates that there is substantial varation in parameter estimates, including estimates of $\bar{k}$, across the small samples. However, formal tests of parameter estimate differences across subsamples suggest that only the parameter $\beta$ for 2008 is significantly different from the $\beta$'s of the other years, and among the dummy shift estimates, only the estimate for salary workers ($\delta_{SW}$) in 2009 is significantly different from the others.

Correspondingly, and as expected, the statistical power of many of the estimates of $k$ for subgroups is low, reflecting that the number of observations is relatively small. Power test results are reported in Table B3, given that the level of significance is set to 0.1. We find power values above 0.8 only for the $k$ estimate when restricting to the self-employed with low income and

*Table B2: Separate Estimations for Each Year, 2003–2009 and 2012*

| Year | $\beta$ | $\delta_{SE}$ | $\delta_{SW}$ | $\overline{k}_{SE}$ | $\overline{k}_{SW}$ | Obs. |
|---|---|---|---|---|---|---|
| 2003 | 0.401 (4.38)*** | 0.137 (3.61)* | 0.019 (0.33) | 1.44 (2.28) | 1.04 (0.07) | 492 |
| 2004 | 0.347 (4.38)*** | 0.194 (3.61)*** | 0.067 (1.28) | 1.79 (6.23)** | 1.20 (1.23) | 501 |
| 2005 | 0.333 (4.35)*** | 0.003 (0.04) | -0.036 (-0.54) | 1.03 (0.02) | 0.88 (0.39) | 512 |
| 2006 | 0.324 (3.54)*** | -0.016 (- 0.17) | -0.046 (-0.68) | 0.97 (0.65) | 0.97 (0.65) | 465 |
| 2007 | 0.363 (4.11)*** | 0.013 (0.13) | 0.060 (1.05) | 1.06 (0.04) | 1.17 (0.80) | 461 |
| 2008 | 0.618 (4.71)*** | 0.025 (0.29) | 0.067 (1.03) | 1.06 (0.18) | 1.10 (0.80) | 557 |
| 2009 | 0.261 (3.33)*** | 0.099 (1.26) | -0.117 (-1.83)* | 1.49 (1.09) | 0.63 (4.29)** | 531 |
| 2012 | 0.337 (7.98)*** | 0.005 (0.11) | 0.032 (1.05) | 1.03 (0.07) | 1.09 (0.82) | 2660 |

$p < 0.1^{*}$ $p < 0.05^{**}$ $p < 0.01^{***}$

Notes: Estimation of $k$ based on using information about variance of errors in the case when we use the approximation to permanent income, see Section 3.2. In parentheses are the t-statistics and chi-square statistics for the regression coefficients and the $k$ values, respectively.

high age.

In Table B4 we present results of a further investigation into differences in the slope parameter, $\beta$, across the three groups used in the identification of underreporting. The expenditure approach postulates an equal slope assumption, and Table B4 shows that our data to a large degree comply with this.

*Table B3: Power of Tests for Statistical Significance of k, Significance Level 0.1*

| | OLS, annual | IV, annual | OLS, permanent | IV, permanent |
|---|---|---|---|---|
| Implied $\overline{k}_{SE}$ | 0.443 | 0.612 | 0.581 | 0.706 |
| Implied $\overline{k}_{SW}$ | 0.142 | 0.170 | 0.137 | 0.152 |
| | | | | |
| Implied $\overline{k}_{SE,HI,HA}$ | 0.129 | 0.103 | 0.109 | 0.100 |
| Implied $\overline{k}_{SE,LI,HA}$ | 0.783 | 0.893 | 0.894 | 0.932 |
| Implied $\overline{k}_{SE,HI,LA}$ | 0.187 | 0.140 | 0.180 | 0.109 |
| Implied $\overline{k}_{SE,LI,LA}$ | 0.320 | 0.124 | 0.175 | 0.107 |
| | | | | |
| Implied $\overline{k}_{SW,HI,HA}$ | 0.100 | 0.107 | 0.100 | 0.102 |
| Implied $\overline{k}_{SW,LI,HA}$ | 0.135 | 0.223 | 0.185 | 0.256 |
| Implied $\overline{k}_{SW,HI,LA}$ | 0.198 | 0.129 | 0.140 | 0.109 |
| Implied $\overline{k}_{SW,LI,LA}$ | 0.109 | 0.102 | 0.100 | 0.107 |

*Table B4: Slope Parameter Estimates for the Self-Employed (SE), Salary Workers (SW), and Non-Evading Wage Earners. Pooled Consumer Expenditure Data 2003–2009 and 2012. t-Statistics in Parentheses*

| | Self-employed | Salary workers | Non-evaders |
|---|---|---|---|
| Slope, $\beta$ | 0.356 (5.37)$^{***}$ | 0.382 (4.61)$^{***}$ | 0.337 (11.30)$^{***}$ |
| No. of obs. | 414 | 787 | 4978 |

$p < 0.1^{*}$ $p < 0.05^{**}$ $p < 0.01^{***}$

# C Amounts Evaded and Income on the Demand Side

Note that we do not use information from this data source about amounts evaded, only whether the individual has been involved in tax evasion or not. Instead we proceed by specifying demand functions and calibrate demand parameters according to what we perceive as the best available knowledge at the moment, which includes results found in De Økonomiske Råd (2011), and employ information from the supply side when we calibrate total amounts evaded.

However, Table C1 reports the amounts evaded as seen in TNS Gallup (2009) and Opinion (2006; 2014),[38] although at very wide intervals, which is the main reason why these data are less useful. A clear majority of those who buy hidden services report having purchased for less than NOK 50,000. Amounts evaded also appear to increase with gross income. Among evaders, the share spending more than NOK 10,000 increases with the income level (excluding the income group with income above NOK 1.5 million).

[38] The 2016 survey does not include this type of information.

*Table C1: Number of Observations Sorted by Annual Gross Household Income and Hidden Market Expenditures. Pooled Information, 2006, 2009 and 2014 Survey*

| | Expenditure (NOK) | | | | | |
|---|---|---|---|---|---|---|
| Income (NOK) | 0 | 0–10K | 10K–50K | 50K–100K | >100K | All |
| 0–200,000 | 211 | 11 | 1 | 0 | 1 | 224 |
| 200K–500K | 1509 | 129 | 46 | 7 | 1 | 1692 |
| 500K–800K | 1272 | 142 | 49 | 6 | 5 | 1474 |
| 800K-1.5M | 1014 | 111 | 60 | 10 | 2 | 1197 |
| >1.5M | 257 | 21 | 11 | 1 | 1 | 291 |
| All | 4263 | 414 | 167 | 24 | 10 | 4878 |

Notes: Expenditures measured over 2 years.
1€ = 7.47 Norwegian kroner (NOK), and 1$ = 5.82 NOK in 2012.