Norwegian School of Economics
Bergen, Fall 2017

Underreporting of Income by Self-Employed? A Meta-Analysis

Studies Building on Pissarides & Weber 1989

Erik V. Nilsen and Sunniva Lillestøl

Supervisors: Jarle Møen, Håkon Otneim

MSc in Economics and Business Administration, Finance, and Business Analysis and Performance Management

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This meta-analysis has investigated the relative income underreporting (k) in a sample of 30 international empirical studies, adding up to 342 estimates, which are based on Pissarides & Weber's expenditure-based approach. The effect size, k, is compiled by the exponential function of gamma over beta, and thus do not automatically fit the meta-analysis methodology. Meta-regression analysis shows that publication selection bias is likely present in the literature, implying that researchers and editors systematically select larger estimates to report in their studies. Investigation of the heterogeneity in the literature uncovered diverse characteristics among the studies affecting the estimates. We found that using instrument variables and a proxy for permanent income seem to give systematically lower underreporting of income, though weaker evidence than expected. Published studies tend to report higher estimates than unpublished studies. An economy's tax level is, on the other hand, not important for the estimated underreporting of income in said economy.

When correcting for publication selection bias and applying our assessment of "best practice" in the estimation process, we find that 11.3% of income are left unreported by self-employed from an underreporting factor of k = 1.128.

Preface

This thesis was written as a part of our Master of Science in Economics and Business Administration at the Norwegian School of Economics. The work performed during the fall of 2017 amounts to 30 credits in our two different majors: *Finance* and *Business Analysis and Performance Management*.

We would like to start by thanking Jarle Møen for a challenging, yet immensely educational, task. He introduced us to a new field of study in meta-analysis, a statistical analysis of previously reported research findings on a given phenomenon. During this project, we have learned much about the literature of income underreporting. We have also learned that meta-regression analysis (MRAs) explain much of the excess variation found in empirical economics research, and that any other type of literature review is biased. We want to thank Dr. Tom Stanley for immediate response regarding the insecurities we had about the methodology. Rolf Mirus and Mikael Apel were both so kind to locate a copy of their respective papers from boxes hidden in their homes for decades. For that, we are grateful. The project also challenged us to improve our programming abilities in Stata.

Finally, we would like to give a special thanks to our supervisors, Håkon Otneim and Jarle Møen, for helpful guidance and constructive criticism throughout the process. We are also very grateful to the Norwegian Centre of Taxation at NHH and the Norwegian Tax Administration to be awarded a grant for our master thesis project.

Table of Contents

List of Figures

List of Tables

1. Introduction

Underreporting of income, as part of the shadow economy, has big ramifications in society. Apart from misleading macroeconomic statistics, such behavior reduces the tax base of an economy. A direct consequence of this is reduced tax collections, which in turn directly affects the citizens through funding of public services (Alm, 2012). Reduced tax collections often leads to increased tax rates for compliant tax-paying citizens, which can impose feelings of unjust treatment. A possible consequence is that citizens' trust in the tax system is undermined, reducing overall tax morale.

Furthermore, shadow economy activities, like underreporting of income, distort competition and favor non-compliant businesses over compliant businesses (OECD, 2017a). As non-compliant businesses get more profitable, the failure of honest businesses increases and consequently the shadow economy is able to expand. Another outcome is misallocation of resources when individuals alter their behavior to evade taxes. Examples of this as presented by Alm (2012) are which occupations they choose to enter, how many hours to work, and which investments to undertake.

If decent estimations of the extent and occurrence of such activities are acquired, government authorities may implement tax policies to discourage non-compliance activity designed specifically for the economy in question. Several methods have been used to try to estimate underreporting of income and the shadow economy through the years.

Tanzi (1980/1983) developed the currency demand approach, using demand for cash as an indicator of developments in the shadow economy, by assuming all shadow activities are completed through cash payments. As Schneider & Buehn (2018) point out, Feige's transactions approach follow a similar pattern, using the relationship between total transactions and nominal GDP, and compares this to official GDP. Schneider on the contrary, often bases his estimations on multiple indicators through the model approach (MIMIC) to capture all effects of the shadow economy (Schneider & Buehn, 2018).

We wish to contribute to this research literature by conducting a meta-analysis on the subject of income underreporting. As the number of empirical studies on income underreporting are extensive, we base our meta-analysis solely on studies building on Pissarides and Weber's expenditure-based estimation method. First, this is a widely used and appreciated method in the field. Secondly, a demarcation such as this makes the workload bearable, and it will be easier to compare the studies. Moreover, results can differ significantly between different approaches (Schneider & Buehn, 2018).

Pissarides & Weber (1989) assume that self-employed underreport income, employees report their true incomes, and both groups report food expenditure correctly. They estimate expenditure functions in terms of household characteristics and reported income, and invert the functions to forecast income from reported expenditure. Pissarides & Weber (PW) find that self-employment incomes in Britain have to be multiplied by a factor of 1.55 to arrive at the true incomes. In our meta-analysis, we summarize studies building on this approach. This entails studies from all over the world, using a form of expenditure (food, electricity etc.) to estimate income underreporting by different groups (self-employed, private employees etc.) compared to a reference group (employees, public employees etc.).

“Meta-analysis is now a widely used technique for summarizing evidence for multiple studies” (Sutton et al., 2000, p. 421). As far as we know, this is the first meta-analysis conducted on the topic of income underreporting. In the following, we attempt to provide quantitative answers to the following questions:

i. What is the global average rate of income underreporting by self-employed compared to employees?
ii. Is publication selection bias present in the PW income underreporting literature?
iii. Which sources of heterogeneity within the literature systematically affect the resulting estimates?

We start this project by providing an overview of the analytical and empirical framework developed by Pissarides & Weber (1989) in section 2. Section 3 provides the meta-analysis methodology as proposed by Stanley et al (2012), including the literature search, a presentation of the primary studies and coding procedures. Next follows the first part of the meta-regression methodology used, in the investigation of publication selection bias in section 4. In section 5, we analyze sources of heterogeneity through multiple meta-regression analysis (MRA).

2. Theoretical framework - Expenditure-Based Method

Pissarides and Weber developed and first proposed the expenditure-based method in their 1989 study (shortened to PW) of underreporting by the self-employed in the UK. It uses consumption and income information on households collected in household surveys to estimate the degree of underreporting by the self-employed.

2.1 The Basics

In its simplest form, the expenditure-based method, or PW method, looks at expenditure and income by two different groups, a group deemed the underreporting group – the self-employed, and a reference group which is assumed to report income correctly – employees or wage earners. A few assumptions are made: the two groups both report consumption expenditure, usually food expenditure, correctly, the employees report income correctly and the self-employed underreport income.

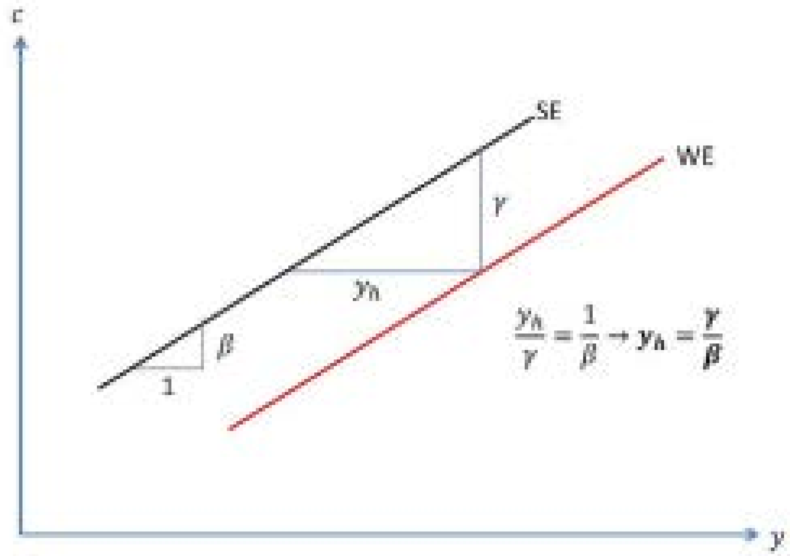
Then, a food expenditure function, a linear Engel curve, is estimated as a function of income, a dummy which is one if the household is self-employed and zero otherwise and control variables which constitute household characteristics for the group which is assumed to report correctly:

$$\ln c_i = z_i \alpha + \beta \ln y_i' + \gamma D_i + \varepsilon_i$$

The marginal propensity to consume, β, and the coefficient vector α on the household characteristics are assumed to be the same for the two groups. Excess food consumption for a given level of income, as indicated by γ, is then an indication of income underreporting, which is estimated by inverting the Engel curve. This is illustrated by the following figure, which is borrowed from Engström & Hagen (2017).

FIGURE 1

Engel-curve: food expenditure as a function of income



Source: Engström & Hagen (2017)

Figure 1 shows the linear Engel curves for self-employed (SE) and employees/wage earners (WE), respectively. β, the marginal propensity to consume, is the slope of the curves, while γ is the difference between the two curves. The amount of income underreported by the self-employed for a given level of consumption is then represented by γ_h. It follows that k, the factor by which self-employed income must be multiplied to obtain their true income, can be calculated as $k = \exp\left(\frac{\gamma}{\beta}\right)$.

This section will first go through Pissarides & Weber's methodology. Then, a few different modifications that different studies have used will be discussed. We will then show what version of the methodology we will use in this meta-analysis, along with the assumptions we will make.

2.2 Thorough review of the PW method

The basis for the PW methodology is to estimate an expenditure function (Engel curve) by using household survey information on consumption of certain goods (food in most studies, including PW), C_i, after-tax reported income, $\mathrm{Y_i}$, and household characteristics, captured by the vector Z_i. The relationship between reported income, $\mathrm{Y_i}$, and actual (current) income, $\mathrm{Y_i^T}$, can be captured by the underreporting factor k_i, which is assumed to be a random variable:

$$Y_i^T = k_i Y_i \tag{1}$$

Employees are assumed to report income truthfully, so for them $k_i = 1$. The self-employed are expected to underreport income, so that $k_i \geq 1$. Both groups of households are assumed to report consumption and Z_i correctly. Then, the following Engel curve is estimated:

$$lnC_i = Z_i\alpha + \beta lnY^P + \varepsilon_i \tag{2}$$

where α is a vector of parameters, β is the marginal propensity to consume, $\ln Y^P$ is the natural logarithm of permanent income and ε_i is a white noise term.

The expenditure function above introduces an important issue in the PW methodology. In line with the permanent income hypothesis, PW argue that consumption is based on permanent income rather than current income and introduce a factor p_i to account for the difference between the two income measures:

$$Y_i^T = p_i Y_i^P \tag{3}$$

p_i is a random variable and PW assume that the mean of p_i is the same for both groups, i.e. $\bar{p}_{SE} = \bar{p}_{EE}$. PW also expect the variance of p_i to be different for the two groups, and it's expected to be bigger for self-employed who might have more volatile income situations.

(1)-(3) imply the following relationship between reported income and true permanent income:

$$k_i Y_i = p_i Y_i^P$$

$$lnk_i + lnY_i = lnp_i + lnY_i^P$$

$$lnY_i^P = lnY_i + lnk_i - lnp_i \tag{4}$$

p_i and k_i are not known, so to make estimation of the amount of underreporting possible, PW argue that the random variables p_i and k_i are log-normally distributed and write them as deviations from their means:

$$lnp_i = \mu_p + u_i \tag{5}$$

$$lnk_i = \mu_k + v_i \tag{6}$$

where u_i and v_i are random variables with zero means and variance σ_u^2 and σ_v^2, respectively.

Combining (4) with (5) and (6) we get:

$$lnY_i^P = lnY_i + (\mu_k + v_i) - (\mu_p + u_i)$$

$$lnY_i^P = lnY_i - (\mu_p - \mu_k) - (u_i - v_i) \quad (7)$$

(7) can then be combined with the Engel curve in (2) to get the following Engel curve:

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

$$lnC_i = Z_i\alpha + \beta lnY_i - \beta(\mu_p - \mu_k) + \eta_i \quad (8)$$

where $\eta_i = \varepsilon_i - \beta(u_i - v_i)$. As the two groups are assumed to have different variances in u_i and v_i, $\sigma^2_{v_{EE}} = 0$ while $\sigma^2_{v_{SE}} > 0$ and $\sigma^2_{u_{SE}} > \sigma^2_{u_{EE}}$, the error term η_i is heteroscedastic. As lnY_i and the error term are correlated, which can be seen from (7), lnY_i is instrumented with a set of X identifying instruments:

$$lnY_i = Z_i\delta_1 + X_i\delta_2 + \xi_i \quad (9)$$

To derive the estimation of the average underreporting factor $\bar{k}$, one can introduce a dummy variable, D_i, that equals 1 for households in the self-employed group and takes a value of 0 otherwise (for households in the employee group):

$$lnC_i = Z_i\alpha + \beta lnY_i + \beta\left(\mu_{k_{EE}} - \mu_{p_{EE}}\right) + \beta D_i\left[\left(\mu_{k_{SE}} - \mu_{p_{SE}}\right) - (\mu_{k_{EE}} - \mu_{p_{EE}})\right] + \eta_i \quad (10)$$

As employees are assumed to report income correctly, $k_{EE} = 1$, which means $\mu_{k_{EE}} = 0$. Equation (10) can then be rewritten as the following:

$$lnC_i = -\beta\mu_{p_{EE}} + Z_i\alpha + \beta lnY_i + \gamma D_i + \eta_i \quad (11)$$

where $\gamma = \beta(\mu_{k_{SE}} - (\mu_{p_{SE}} - \mu_{p_{EE}}))$. γ can be understood as the difference in the constant term between the self-employed and employees, and as such it indicates excess food consumption by the self-employed.

To develop the γ expression further, one can take a closer look at the random variable p_i. By the properties of the log-normal distribution:

$$\bar{p} = \exp(\mu_p + \frac{1}{2}\sigma_u^2)$$

$$ln\bar{p} = \mu_p + \frac{1}{2}\sigma_u^2$$

Then, looking at the two groups of households, remembering that $\bar{p}_{SE} = \bar{p}_{EE}$, yields the following:

$$ln\bar{p}_{SE} = \mu_{p_{SE}} + \frac{1}{2}\sigma^2_{u_{SE}}$$

$$ln\bar{p}_{EE} = \mu_{p_{EE}} + \frac{1}{2}\sigma^2_{u_{EE}}$$

$$\mu_{p_{SE}} - \mu_{p_{EE}} = -\tfrac{1}{2}(\sigma^2_{u_{SE}} - \sigma^2_{u_{EE}}) \leq 0 \qquad (12)$$

Using (12), the expression for γ is then:

$$\gamma = \beta\left(\mu_{k_{SE}} + \tfrac{1}{2}\left(\sigma^2_{u_{SE}} - \sigma^2_{u_{EE}}\right)\right) \qquad (13)$$

Remembering the properties of the log-normal distribution:

$$ln\bar{k}_{SE} = \mu_{k_{SE}} + \frac{1}{2}\sigma^2_{v_{SE}}$$

Combining this with (13), the average underreporting factor, $\bar{k}$, which is the factor by which the average reported income by the self-employed must be multiplied to get their average true income, can be expressed as:

$$\bar{k}_{SE} = exp(\mu_{k_{SE}} + \tfrac{1}{2}\sigma^2_{v_{SE}}) = exp\left(\frac{\gamma}{\beta} + \tfrac{1}{2}\left(\sigma^2_{v_{SE}} - \sigma^2_{u_{SE}} + \sigma^2_{u_{EE}}\right)\right) \qquad (14)$$

To get an interval estimate of the underreporting factor, some inferences about the variance of u_i and v_i can be made from the estimated residual income variances from the income regression or first stage regression as depicted by (9). The error term ξ_i includes unexplained variation in permanent income as well as u_i and v_i, as is clear from (7): $lnY_i = lnY_i^P + \left(\mu_p - \mu_k\right) + (u_i - v_i)$. PW then assume that there is no difference between the two household groups in terms of variance of unexplained variations in permanent income. Then, remembering that $\sigma^2_{v_{EE}} = 0$, the difference in the estimated residual income variance can be expressed as the following:

$$var(\xi_{SE}) - var(\xi_{EE}) = var(u - v)_{SE} - var(u)_{EE} \qquad (15)$$

Expanding yields:

$$\sigma^2_{\xi_{SE}} - \sigma^2_{\xi_{EE}} = \sigma^2_{u_{SE}} + \sigma^2_{v_{SE}} - 2cov(uv)_{SE} - \sigma^2_{u_{EE}} \qquad (16)$$

Then, assuming that the correlation coefficient between u and v is zero, the lower and upper bounds of the underreporting factor can be calculated with one further assumption each. The lower bound is found when $\sigma^2_{v_{SE}}$ takes its lowest value, which is zero, i.e. every self-employed household underreports income by the same proportion, no matter the level of income. The upper bound is found when $\sigma^2_{u_{SE}}$ takes its lowest value, which happens when $\sigma^2_{u_{SE}} = \sigma^2_{u_{EE}}$, i.e. the income variance is the same for the self-employed and employees.

These assumptions yield the interval estimate of $\bar{k}$:

$$\bar{k}_{SE} = exp\left(\frac{\gamma}{\beta} \pm \frac{1}{2}\left(\sigma^2_{\xi_{SE}} - \sigma^2_{\xi_{EE}}\right)\right) \quad (17)$$

It is also possible to allow the correlation coefficient between u and v to be nonzero. However, PW show that a small correlation coefficient does not have a large effect, at least on the UK data.

2.3 Alternative modifications

Several studies have followed the PW approach as outlined above. For example, Shuetze (2002) applied it to Canadian data and Johansson (2005) applied it to Finnish data. However, several studies have later modified the PW methodology by changing the assumptions or using proxies for permanent income to get a point estimate. We draw from Paulus (2015) and give a short overview of some of these alternative specifications.

One modification done to the PW methodology is the use of a proxy for permanent income. There are mainly two ways of doing this. Firstly, one can use a measure of average income, from panel data or register data, instead of current income. Kim, Gibson & Chung (2017) used this alternative specification. This is argued to remove variation in p_i and means (11) is simplified so that $\gamma = \beta\mu_{k_{SE}}$ and (16) is simplified to $\sigma^2_{\xi_{SE}} - \sigma^2_{\xi_{EE}} = \sigma^2_{v_{SE}}$. This leads to the following expression of the average underreporting factor, which is numerically equal to the upper bound in the PW approach:

$$ln\bar{k}_{SE} = \mu_{k_{SE}} + \frac{1}{2}\sigma^2_{v_{SE}} => \bar{k}_{SE} = exp\left(\frac{\gamma}{\beta} + \frac{1}{2}\left(\sigma^2_{\xi_{SE}} - \sigma^2_{\xi_{EE}}\right)\right) \quad (18)$$

Secondly, as Kukk & Staehr (2014) do, one can use a reported measure of regular income as a measure of permanent income. Thus, $Y_i^P = k_i Y_i$ instead of (1) and (3). This leads to the same expression for γ as above and the expression for the average underreporting factor is the same as (18).

Hurst, Li & Pugsley (2014) take a different approach. They assume that the transitory income component, p_i, is the same for the self-employed and employees after controlling for characteristics (instead of $\bar{p}_{SE} = \bar{p}_{EE}$). They also look at the share of true income reported, κ_i, and assume this is a constant instead of a random variable. This changes the model slightly, turning (1) into $\kappa Y_i^T = Y_i$. This means the sign before $\ln k$ in (4) is changed from a plus to a minus, which turns (14) into its equivalent average share of true income reported:

$$\bar{\kappa}_{SE} = exp\left(-\frac{\gamma}{\beta}\right) \tag{19}$$

This approach is based on the combination of the lower and upper bound assumptions in the PW approach: $\sigma^2_{v_{SE}} = 0$ and $\sigma^2_{u_{SE}} = \sigma^2_{u_{EE}}$. Several other studies, including Besim & Jenkins (2005) and Engström & Holmlund (2009), have used the same set of assumptions but used the average underreporting factor k_i instead of κ_i, leading to an underreporting factor

$$\bar{k}_{SE} = exp\left(\frac{\gamma}{\beta}\right) \tag{20}$$

Under these assumptions, there is a simple relationship between the average underreporting factor k_{SE} and the share of true income reported κ_{SE}:

$$k_{SE} = exp\left(\frac{\gamma}{\beta}\right) = \frac{1}{exp(-\frac{\gamma}{\beta})} = \frac{1}{\kappa_{SE}} \tag{21}$$

Lastly, Engström & Hagen (2017) among others, estimate and report the share of true income underreported, $\bar{s}_{SE}$:

$$\bar{s}_{SE} = 1 - \bar{\kappa}_{SE} = 1 - \frac{1}{k} = \frac{k-1}{k} \tag{22}$$

2.4 Facilitation for meta-analysis

To utilize the tools of the meta-analysis framework to the fullest, it is important that we have comparable effect sizes and their standard errors. This is needed to estimate the global average underreporting of income, it is needed for investigations into publication selection bias and it is needed to carry out the meta-regression analyses that seek to explain effect size heterogeneity.

2.4.1 Calculating k

The meta-analysis framework is based on having one effect size, often in the form of a single regression coefficient. One would run regressions with the effect size as the dependent variable and its standard error as the independent variable, with or without several other independent variables. The PW underreporting literature, however, does not fit exactly into this framework. Instead of the effect size simply being a regression coefficient, the effect size in the PW literature, or the underreporting factor k, is made up of two regression coefficients, gamma and beta, and variance terms depending on how k is calculated. Also, sometimes k is reported as an interval estimate, while at other times a point estimate is reported, and that point estimate might be with or without variance terms. The different ways in which k is reported in the various studies means it is difficult to use reported k as a comparable effect size.

It follows that we must have one comparable effect size for each estimate in each study, and we must be able to calculate the standard error of such an effect size. One way of solving this problem is to apply some of the assumptions that were used by many of the later studies that use the expenditure-based estimation method. This means we can use the simplified measure of $\bar{\mathrm{k}}_{\mathrm{SE}} = \exp\left(\frac{\gamma}{\beta}\right)$ or $\bar{\mathrm{s}}_{\mathrm{SE}} = 1 - \bar{\kappa}_{\mathrm{SE}} = \exp\left(-\frac{\gamma}{\beta}\right)$ as effect sizes, such as Hurst et al. (2014), Besim & Jenkins (2005) and Engström & Holmlund (2009) do. Most studies report β and γ and their standard errors, which we can use to calculate the underreporting measure, and the standard error of this measure can be quite easily calculated. Also, some studies report $\bar{\mathrm{k}}_{\mathrm{SE}}$ or $\bar{\mathrm{s}}_{\mathrm{SE}}$ (or $\bar{\kappa}_{\mathrm{SE}}$) directly along with their standard errors.

This approach relies on the two assumptions that make up the lower and upper bound estimates in the PW approach. From the lower bound estimate, we assume a constant k (instead of a random variable), equivalently that all self-employed households underreport the same share of income no matter the absolute level of income, i.e. $\sigma^2_{v_{SE}} = 0$. From the upper bound estimate, we assume that the self-employed and the employees have the same income variance, i.e. $\sigma^2_{u_{SE}} = \sigma^2_{u_{EE}}$. These assumptions might not hold exactly, but Engström & Hagen (2017) state that the simplified measure usually gives a good approximation of the underreporting factor.

The fact that the PW literature does not fit perfectly into the meta-analysis framework means there is little precedence in meta-analysis research in terms of how to deal with a situation in which the effect size is not a single regression coefficient. Based on the remarks above, we choose to use the simplified measure of k as the comparable effect size. When studies report β and γ, we calculate k as (20). When β and γ are not reported, and studies report simplified k estimates and their standard errors, we use these estimates directly. To combine approximated estimates based on β and γ (and their standard errors) with precise reported estimates (and standard errors) is somewhat inconsistent, and may affect the results in some way. However, the latter alternative is only needed for three estimates deriving from Kukk & Staehr (2014). It is unlikely that 3 estimates out of the total of 342 estimates will make much difference.

We discuss the calculation of the standard errors in the next section.

2.4.2 Calculating the standard error of k

We choose to use the Delta-method in order to calculate the standard errors of the average underreporting factors (k). This method is a way of calculating the uncertainty of a function of two or more variables given the uncertainty of those variables. It is important to note that this method of calculation only provide an approximation of the standard error, not precise estimates.

We have the following formula for calculating the approximate standard deviation of a function Y = f(X, Z) (NIST/SEMATECH, 2013, ch.2.5.5):

$$s_y = \sqrt{\left(\frac{\partial Y}{\partial X}\right)^2 s_x^2 + \left(\frac{\partial Y}{\partial Z}\right)^2 s_z^2 + \left(\frac{\partial Y}{\partial X}\right)\left(\frac{\partial Y}{\partial Z}\right) s_{xz}^2} \qquad (23)$$

where s_x is the standard deviation of X, s_z is the standard deviation of Z, $\left(\frac{\partial Y}{\partial X}\right)$ and $\left(\frac{\partial Y}{\partial Z}\right)$ are the partial derivatives of Y with respect to X and Z, respectively, and s_{xz} is the covariance between X and Z.

We have the following equation for the average underreporting factor k:

$$k = \exp\left(\frac{\gamma}{\beta}\right)$$

Applying (23) to k yields the following standard deviation of k:

$$s_k = \sqrt{\left(\frac{\partial k}{\partial \gamma}\right)^2 s_\gamma^2 + \left(\frac{\partial k}{\partial \beta}\right)^2 s_\beta^2 + \left(\frac{\partial k}{\partial \gamma}\right)\left(\frac{\partial k}{\partial \beta}\right) s_{\gamma\beta}^2} \qquad (24)$$

where the partial derivatives of k with respect to γ and β, respectively, are the following:

$$\left(\frac{\partial k}{\partial \gamma}\right) = \frac{1}{\beta} exp\left(\frac{\gamma}{\beta}\right) \qquad (25)$$

$$\left(\frac{\partial k}{\partial \beta}\right) = -\frac{\gamma}{\beta^2} exp\left(\frac{\gamma}{\beta}\right) \qquad (26)$$

How to deal with the covariance is not entirely clear. We tried two different approaches. The first was to assume the covariance is zero. The second was to estimate the covariance between the sample of betas and gammas, assuming it is the same across countries and studies. We chose to apply the first method, with covariance equal to zero. First, even though the correlation between gamma and beta is somewhat high (0.58), when using equation (24), the resulting standard errors are not much different from the ones calculated with the assumption of zero covariance. Second, several of the estimates' standard errors are not possible to calculate when applying the second method, due to a negative variance (and we are therefore unable to take the square root to obtain standard errors). Even though the second method might be more correct theoretically, we believe assuming the covariance is zero makes a fair approximation, as the calculated standard errors are similar (disregarding the estimates with negative variance). Also, it should be mentioned that the covariance term in formula (24) is negative due to the partial derivative in (26) being negative (assuming by our calculations that the correlation is positive). That means our estimate of the standard error of k when using the approximation in equation (24), is upward biased. We account for more uncertainty than might be needed, and that is likely better than estimating a too low standard error.

3. **Data extraction**

We start this section with a definition of meta-analysis from Stanley & Doucouliagos (2012, p. 2): ”Meta-analysis is the statistical analysis of previously published, or reported, research findings on a given hypothesis, empirical effect, phenomenon, or policy intervention. It is a systematic review of all the relevant scientific knowledge on a specific subject…” The empirical effect in question in this meta-analysis is the underreporting of income, as estimated by studies utilizing the expenditure-based method proposed by Pissarides & Weber (1989).

In the spirit of the meta-analysis framework, as proposed by for example Stanley & Doucouliagos (2012), the main objectives of this meta-analysis are calculating a global average degree of income underreporting, investigating the presence of publication selection bias (or not) and investigating the sources of heterogeneity that characterizes the various reported estimates.

In chapter 3, 4 and 5, we will go through the methodology and theory that underpin these quests. We will start this methodology section by going through the literature search and the coding of the primary studies here in chapter 3. In the presentation of the meta-analysis methodology, we draw heavily from Stanley & Doucouliagos (2012).

3.1 Procedure and studies

The literature in question is comprised of studies that estimate underreporting of income (mainly by the self-employed) using the expenditure-based method developed by Pissarides & Weber (1989). We follow the meta-analysis guidelines set forth by Stanley et. al. (2013). The literature search was conducted by two reviewers and consist of a forward citation search, followed by a backward citation search (of the relevant retrieved papers from the forward citation search).

First, we carried out a forward citation search on Google Scholar by searching for all studies that have cited the Pissarides & Weber (1989) study. It is quite reasonable that any reputable paper applying the PW method would cite the original PW paper. This search took place in September 2017 and was completed on September 29, resulting in 382 hits. We then used a set of exclusion criteria to exclude irrelevant studies and to arrive at a preliminary set of primary studies. This process, along with the exclusion criteria, are listed in Table 1 below:

TABLE 1 – Literature search

Preliminary search results	
From Supervisor	3
Total hits from Google Scholar	382
Excluded hits	
Non-English studies	65
Non-empirical studies	78
Remaining studies	242
Incorrect phenomenon - not estimating underreporting of income	114
Remaining studies	128
Not using P&W-method	79
PW studies	49
Duplicates	18
Does not report estimates	2
Missing standard errors	3
Preliminary primary studies	26
Additional search results	
Backward citation search	4
Primary studies	30

We exclude non-English studies because it is important that the reviewer can fully understand the coded studies. It would however be interesting to see if there were a number of non-English empirical studies on the subject. This would require a deep dive into the non-English literature, and to save time and resources we therefore started the literature search by excluding the non-English studies before examining whether the remaining studies were empirical or not. Also, "most empirical economics papers are actually written in English, so that any bias resulting from omitting non-English studies should be of a second order" (Stanley & Doucouliagos, 2012). The exclusion criteria of non-empirical studies, not estimating income underreporting and not using the PW method ensure that we end up with studies that are relevant to this meta-analysis. Furthermore, not all these 46 PW studies can be included in the meta-analysis. As recommended by T.D. Stanley (personal communication, October 3, 2017, see Appendix B), we use the latest version of a study to avoid counting estimates twice and avoid errors (or even typos) that may have been present in earlier versions. This results in some studies being labeled as duplicates and excluded. This also goes for Wangen (2005) which reused Pissarides & Weber's (1989) estimates to develop new estimators. In addition, some studies didn't report sufficient estimates or standard errors and were excluded for that reason.

We then carried out further investigations of the references cited by the preliminary list of primary studies obtained through the forward citation search. This led to the discovery of five relevant studies: Baker (1993), Apel (1994), Cullinan (1997), Mirus & Smith (1997) and Nygård et al (2016). However, we were not able to retrieve the full-text version of the Cullinan paper. Therefore, we exclude this paper from our meta-analysis (also from Table 1 above), while the four other papers are included. Along with the task, three papers were handed to us by one of our supervisors. This includes a master thesis (Skjeggestad & Wæhle, 2015), applying the PW method to Norwegian data, the original Pissarides & Weber (1989) study, and a secluded study by Wangen (2004) (also on Norwegian data).

Table 2 gives an overview of all the 30 primary studies included in this meta-analysis. The table includes the country to which the studies relate and the number of estimates per study. The table also reports the number of observations per study (given by the estimate with the highest number of observations). The three columns with different estimates of k are discussed on pages 25-27. The number of estimates reported by each study vary quite extensively. Four studies report only one estimate (which is comparable and included in this meta-analysis), while three studies report 40 estimates or more. Two of these are Paulus (2015) with 60 estimates and the master thesis by Skjeggestad & Wæhle (2015), which reports 40 estimates. These two studies are also unpublished. It is possible that these studies have a large effect on the regression results, so we conduct robustness checks where we reduce the sample to exclude one or both of these studies.

In addition, it is interesting to note that the study average underreporting factors also vary quite a lot, from a minimum of 1.00 to a maximum of 2.19. This variation in the effect size means we might be able to discern certain patterns in the data.

TABLE 2

Empirical studies estimating underreporting of income using the expenditure-based method

Study	Country	Estimates	Simple average reported k[1]	Simple average calculated k	Weighted average calculated k	Max. Observations
Pissarides & Weber (1989)	Britain	2	1.55	1.49	1.47	1283
Baker (1993)	Britain	24	1.33	1.33	1.30	3092
Apel (1994)	Sweden	3	1.32	1.28	1.24	2230
Mirus & Smith (1997)	Canada	2	1.13	1.14	1.14	4502
Obwona (1999)	Uganda	7	-	1.00	1.01	8559
Schuetze (2002)	Canada	25	1.17	1.17	1.16	8463
Lyssiotou et al (2004)	Britain	1	1.39	1.35	1.35	-
Wangen (2004)	Norway	9	1.01	1.01	1.00	1263
Bernotaité & Piskunova (2005)	Latvia	4	1.20	1.21	1.18	1632
Besim & Jenkins (2005)	North-Cyprus	2	1.14	1.16	1.16	723
Johansson (2005)	Finland	4	1.29	1.39	1.26	2054
Torero et al (2006)	Jamaica	1	1.32	1.37	1.37	1009
Davutyan (2008)	Turkey	4	1.22	1.23	1.23	8550
Engström & Holmlund (2009)	Sweden	12	1.31	1.31	1.27	6004
Kapociute (2013)	Australia	6	1.41	1.38	1.21	2280

[1] See page 26 and 27 for an explanation of the content of this column

TABLE 2 (cont.)

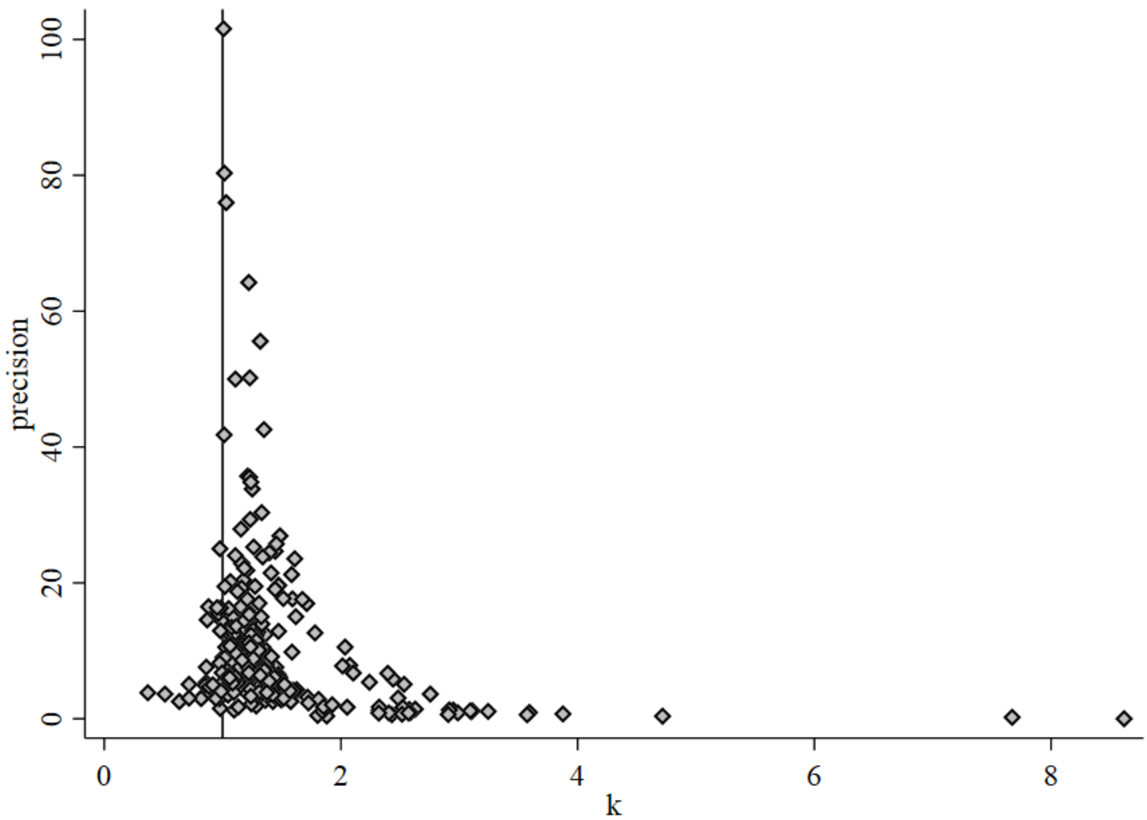
Study	Country	Estimates	Simple average reported k	Simple average calculated k	Weighted average calculated k	Max observations
Martinez-Lopez (2013)	Spain	15	1.30	1.37	1.23	16451
Åstebro & Chen (2014)	US	4	1.46	1.46	1.41	39037
Hurst et al (2014)	US	18	1.36	1.45	1.34	36434
Kukk & Staehr (2014)	Estonia	9	2.26	2.19	1.50	6016
Torosyan & Filer (2014)	Georgia	5	2.06	1.16	1.16	1743
Paulus (2015)	Estonia	60	2.59	1.93	1.16	4754
Skjeggestad & Wæhle (2015)	Norway	40	1.09	1.09	1.12	29097
Ekici & Besim (2016)	North Cyprus	1	1.26	1.26	1.26	861
Nygård et al (2016)	Norway	4	1.22	1.20	1.19	4213
Torregrosa-Hetland (2016)	Spain	6	1.21	1.21	1.22	14442
Anwar et al (2017)	Pakistan	1	2.07	1.37	1.37	12577
Engström & Hagen (2017)	Sweden	43	1.30	1.30	1.25	9165
Kim et al (2017)	Korea and Russia	16	1.37	1.77	1.25	10675
Kukk & Staehr (2017)	Estonia	8	1.70	1.61	1.45	6016
Parvathi & Nguyen (2018)	Laos	6	2.42	2.55	2.53	968
Total		**342**	**1.43**	**1.45**	**1.18**	

The “number of estimates” column is based on our calculation of k from beta and gamma. Not all studies report the same amount of k’s. For instance, the simple average reported k from Paulus (2015) is based on only 12 out of 60 estimates.
Each of the studies’ weighted average is calclulated using equation (27), see section 3.2 below. The total weighted average of 1.18 is also calculated by using equation (27), using all the available estimates.

A summary description can yield a helpful first look at the data collected. Before taking a deep dive into the average measures, we want to get an overview of the distribution of the data. To illustrate this, Stanley & Doucouliagos (2012) propose drawing a funnel graph. Such a funnel plot allows for an illustration of the distribution of the data in terms of the precision of the various effect sizes. The most precise estimates are found at higher levels in the funnel graph, while the least precise ones are located at lower levels. The shape of the graph can give an indication to whether publication selection bias seems to be present. We will address the problem of publication selection soon. This plot makes use of two of the most important variables; it plots the estimates' precision against the underreporting factor k. We produce this funnel plot in Figure 2 below.

FIGURE 2

Funnel Graph of Underreporting factor k



The funnel graph plots the estimates' precision (1/SE) against the (simplified) underreporting factor k. k and its standard error are calculated as explained in sections 2.4.1 and 2.4.2, respectively.

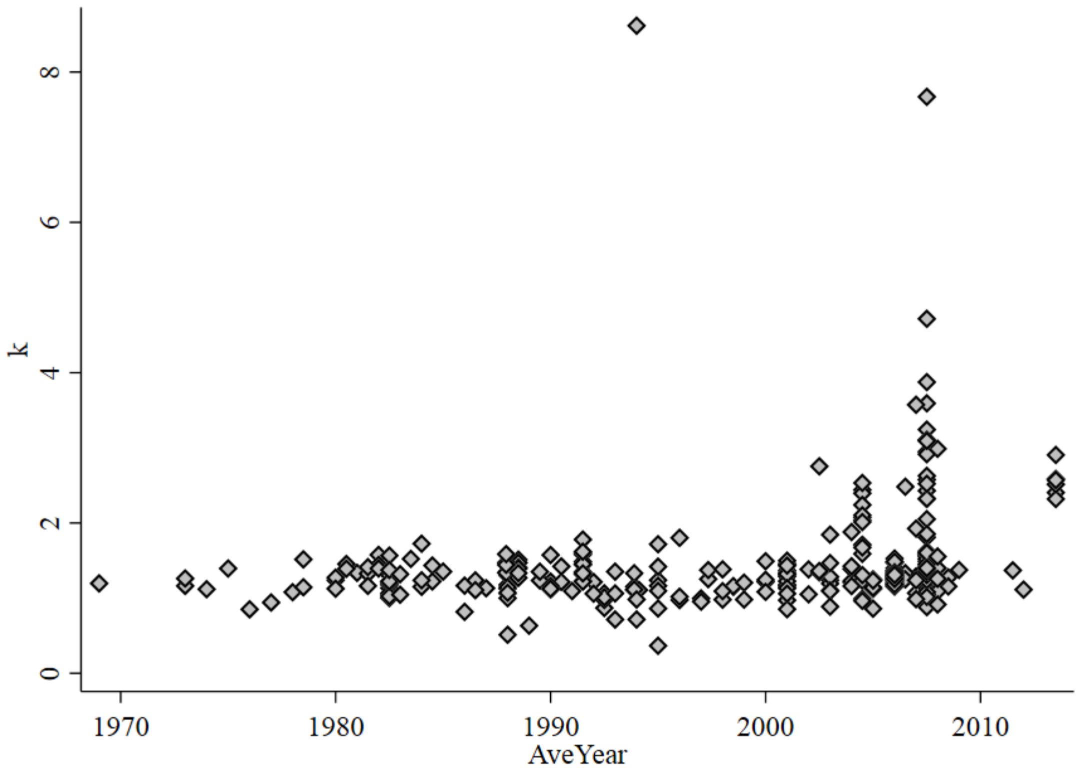
From the plot it is clear that a majority of the estimates report underreporting factors between 1 and 2. A few estimates report very high underreporting factors, while a few of the underreporting factors are below 1, indicating that the suspected underreporting group overreports income. It should also be noted that there is a substantial variance in the precision of the estimates, and the plot makes clear that the most precise estimates are fairly close to an underreporting factor of 1. This will be of high importance in the investigations into publication selection bias.

It should be noted at this point that there is a problem with the way k is distributed and thereby the value of the funnel plot in the investigations into publication selection bias. In order to claim that an asymmetric funnel plot indicates the presence of selection bias, the true distribution of k needs to be symmetrical. However, k is calculated as an exponential function, which means that it can never take on a negative value. This implies an underlying asymmetry in the distribution of k. However, publication selection bias would make the funnel plot even more asymmetric. As discussed in section 2.4.1, we are outside the standard meta-analysis framework, making it difficult for us to make a definite conclusion. To support the analysis on k alone, we complement it with analyses on beta and gamma separately. These are normal regression coefficients and have normal distributions, which means it might be easier to draw conclusions from the funnel plots.

Another helpful plot is a chronological ordering of the effect sizes. Stanley & Doucouliagos (2012) propose the use of such a graph to capture the evolution of the literature. As such, it might show indications of a trend or structural breaks in the effect sizes. We plot the estimate's underreporting factor k against the average year of the data on which the estimate is based. This plot is presented in Figure 3.

FIGURE 3

Chronological plot of underreporting factor k



In this chronological graph we plot k against the variable AveYear to investigate potential time trends in k. k is calculated as in figure 2. AveYear is the average year of the underlying consumption/income data on which the estimates are based.

The chronological plot shows the evolution of the underreporting factor over time. No clear time trend can be discerned from the graph, although larger estimates have been reported based on newer data. Thus, there is the likelihood that underreporting factors may have increased over time. Including a time trend in a meta-analysis is quite common, so we include an average year variable among the possible moderator variables.

The next interesting mission is to calculate the average global underreporting factor. A first attempt at this is made in Table 2, which reports a simple (unweighted) average and a weighted average. Both these average measures are reported for each study and for the total number of estimates at the bottom of the table.

Table 2 reports a simple average underreporting factor of 1.45 for the all the estimates combined. This would mean that one must multiply the reported incomes of the underreporting group, which is mostly self-employed, by 1.45 to reach their true incomes. Equivalently, using equation (22), this average underreporting factor would imply that the underreporting group on average does not report s=1-1/1.45 = 31% of their true incomes.

However, this simple measure is not the statistically optimal measure. According to Stanley & Doucouliagos (2012), a weighted measure, which takes the variance of the estimates into account, is the statistically preferred choice:

$$k_w = \frac{\sum w_i k_i}{\sum w_i} \qquad (27)$$

where k_i is the underreporting factor k of the ith estimate and w_i is the reciprocal of the square of its standard error ($1/SE^2$).

Table 2 reports the weighted average k for all the estimates combined as 1.18, or equivalently that the underreporting group on average fails to report 1-1/1.18 = 15% of their true incomes. As expected based on the funnel plot, this number is substantially lower than the simple average because lower weight is given to the relatively high number of less precise estimates of high k's.

However, these two average numbers (based on all the estimates) may be biased because they fail to account for a phenomenon that is present in many research literatures: publication selection bias. These summary measures should be regarded as pre-corrected measures because they build on the underlying assumption of no publication selection bias and a homogenous dataset. These assumptions are later eased as we use simple MRA models to test for the presence of publication selection and multiple MRA models to investigate the heterogeneity of the literature.

Before addressing the moderator variables, we want to take a look at the way we calculated k. Our measure is a simplified measure which neglects possible variance terms. It is interesting to compare our estimates of k to the ones reported by the studies. As mentioned earlier, some studies report interval estimates, while others report point estimates with or without variance terms. These different estimates of k are summarized in Table 2 in the column named "Simple average k reported". Regrettably, we do not have a full overview of which estimation methods underlie each reported estimate. This entail that the column of reported estimates are ambiguous and not consistent. The underreporting factor k presented per study is either the average reported k by the study, such as the 1.55 number reported by Pissarides & Weber, or an average based on reported intervals or point estimates (either k directly or calculated from kappa). In general, the reported k's and our estimated k's are quite similar. However, there are some bigger differences for some studies. One of these is Torosyan & Filer (2014), which has a reported average of 2.06 and an average estimated k of 1.16. This might be an indication of quite large variance terms creating differences between the reported k and our estimate of k. We did examine this further, and it turns out that this study with data from Georgia operates with a high value for covariance. However, this does not seem to be a problem for most of the studies, where our calculations of k match the reported k's quite well. Thus, this might only be a problem for a few certain studies. Some deviations are detected, without any systematic patterns. Using the simplified measure of k, where we disregard the variance terms, will therefore serve as a good approximation by our assessment. There might be systematic differences between point estimates and interval estimates, as well as between point estimates with or without variance terms. Optimally, if we had enough data on the matter, we would investigate this further. It would however be difficult to standardize the different estimates of k, especially in the cases where effect sizes are intervals. Our chosen effect size is based on beta and gamma, which are easily standardized.

3.2 Moderator variables

Furthermore, the 30 primary studies were coded by two reviewers. We coded several dimensions of the data. However, some of these dimensions, for example dividing the underreporting group into white collar and blue collar workers or into incorporated and unincorporated, are used by very few studies and are of questionable value. That is, a lot of information were coded, but we narrowed it down to 19 moderating variables (also the reference categories). In addition comes the effect sizes, including the underreporting factor k, beta and gamma, and their standard errors (SE). The moderating variables were chosen entirely based on the information available and what seemed to be important factors looking at the literature at hand. Table 3 below lists these variables along with their sample means and standard deviation, as well as the number of estimates to which the variable is relevant.

TABLE 3

Potential Moderator Variables for Meta-Regression Analysis

Moderator variables	Definition	Mean (standard deviation)	No. of estimates
k	Underreporting factor k, calculated as described in section 2.4.1	1.45 (0.73)	342
SE k	standard error of the estimated underreporting factor k	0.47 (4.28)	342
Beta	Propensity to consume, regression coefficient reported by the studies	0.36 (0.22)	339
SE beta	Standard error of beta	0.05 (0.06)	339
Gamma	Excess consumption by underreporting group compared to reference group, regression coefficient reported by the studies	0.11 (0.16)	339
SE gamma	Standard error of gamma	0.04 (0.06)	339
IV	= 1 if estimate is calculated using instrument variables, 0 otherwise	0.66 (0.47)	262
PermProx	= 1 if estimate is calculated using proxy for permanent income, 0 otherwise	0.15 (0.35)	50
Housing	= 1 if housing is used as an instrument variable	0.29 (0.45)	98
Education	= 1 if education is used as an instrument variable	0.49 (0.50)	167
Capital income	= 1 if capital income is used as an instrument variable	0.05 (0.22)	17
TotExp	= 1 if dependent variable in Engel curve is total expenditure, Food is reference category	0.02 (0.15)	8
OtherExp	= 1 if dependent variable in Engel curve is defined otherwise. Food is reference category	0.21 (0.41)	72
Panel	= 1 if estimate relates to panel data, with cross-sectional data as base	0.17 (0.37)	57
Head	= 1 if estimate identifies self-employed (underreporting group) using only head of household	0.39 (0.49)	133
Employees	= 1 if estimate includes employees in underreporting group	0.13 (0.34)	45
Public Employees	= 1 if estimate identifies comparison group as public employees	0.19 (0.39)	64
Share	= 1 if estimate defines self-employed by (only) share of income. Status is reference category	0.26 (0.44)	90
Share&status	= 1 if estimate defines self-employed by share and status. Status is reference category	0.11 (0.31)	38
AveYear	is the average year of the data used, with 2000 as base year	- 1.42 (9.82)	342
Published	= 1 if the estimate comes from a published study	0.54 (0.50)	185
GDP/capita	GDP per capita in current 1000 USD, data from The World Bank (2017b)	23.47 (16.60)	342
TaxSystemQuality	Paying taxes index from Doing Business (2017a), 100 represents highest quality of tax system	85.16 (5.80)	342
Corruption	Corruption Perception Index by Transparency International (2017), scale 0-10 (higher score is less perceived corruption)	7.31 (2.02)	342
TaxLevel	Tax revenue as percentage of GDP, data from The World Bank (2017a) and OECD (2017b)	16.06 (10.16)	342

The moderating variables seek to cover possible sources of heterogeneity in the PW income underreporting literature. The first moderating variables listed (after the effect sizes and standard error variables), deal with which econometric methods are used to deal with the problem of transitory income. The “IV” dummy variable represents the use of instrument variables, while the “PermProx” dummy variable represents the use of a proxy for permanent income, normally a measure of average income. In addition, “Housing”, “Education” and “Capital income” specify which instrument variables are used. Also, Panel refers to the type of data used by the study (or estimate).

Then, a few moderating variables seek to describe how variables are defined. “TotExp” and “OtherExp” specify what kind of consumption the expenditure function is based on for a given estimate. “Head”, “Employees”, “Public employees”, “Share” and “Share&Status” represent how households are defined for a given estimate. For example, if “Share” equals 1, households are divided into the underreporting and reference groups by how big a share their self-employment income consitute of their total income.

Furthermore, the “Published” dummy variable was included to account for journal quality. It has a mean of 0.56, which means that slightly more than half of the estimates come from published studies. It should be noted that 66% of the primary studies are published. However, two studies, Paulus (2015) and Skjeggestad & Wæhle (2015), are unpublished and have 60 and 40 estimates, respectively. This is substantially more than the average number of 11 estimates per study, which pulls the mean of the Published variable downwards. Also, the high number of estimates for these studies means it is prudent to do robustness checks in the meta-regression analyses by excluding these estimates.

Finally, four value-added moderating variables were included. Transparency International’s CPI (2017), short for Corruption Perception Index, was coded to see whether countries that are perceived to have bigger problems with corruption (and thus might have institutions of a lower quality) see more underreporting of income (by the underreporting group compared to the reference group). It should be noted that this might not be a perfect measure of corruption. First, the definition of corruption, and thereby the perception of it, differs across countries. In addition, it is difficult to measure perceptions accurately (T. Søreide, personal communication, january 27, 2017). It is also likely that perceptions of a problem differs from the reality of the problem. We also used index scores from different reports as the estimates are significantly dispersed in time. This might make the results less reliable. Also, the CPI was first developed in 1995, so the estimates that use earlier household data were attributed the first available CPI number.

The tax level of the countries was also coded. This was proxied by tax revenue as percentage of gross domestic product from the World Bank (2017a) for the same year on which each estimate is based. When a tax value was not available for the given year and country, we used data from the OECD when available (adjusting it based on the relationship between World Bank and OECD data for the given country to gain a comparable number) or the closest available World Bank number. The coding of tax level data provides an opportunity to investigate whether a higher tax level is accompanied by higher rates of underreporting (i.e. more underreporting by the underreporting group compared to the reference group). There are many possible measures of taxation that could be used. However, this moderator variable might give a fairly good indication of the general tax level in a country.

A third value-added macrovariable we included is GDP/capita. We used data from the World Bank (2017b) for the same year on which the estimate is based. The data we retrieved was reported in current US$, and to better adapt the output from the coming analyses we transformed the variable to current US$ in thousands. This variable was mainly coded to function as a measure of the development in the country in question. As such, it performs as a substitute for country dummies, and we can reduce the number of necessary variables here from numerous to one single variable. This moderator variable investigates whether there is a systematic connection between relative underreporting of income and how developed a country is in the form of GDP/capita.

Lastly, we wanted to investigate whether the quality of the tax system in a country systematically affects the extent of underreporting of income (by the underreporting group compared to the reference group). For this matter, we use Paying Taxes records from The World Banks Doing Business (Doing Business, 2017a). We coded the economies' distance to frontier, where the frontier represents the best tax system performance oobserved across all economies and years. The index accounts for the administrative burden of paying taxes, in addition to the total tax rate, number of transactions and time necessary to pay all taxes by a medium-sized local company (Doing Business, 2017b). The frontier has a value of 100, which means that the economies closest to 100 in value has tax systems of higher quality compared to other economies. It should be mentioned that the Distance to Frontier ranking is only recorded for two years, 2016 and 2017, with data from every year since 2005. We use the record from 2016 in our dataset, but if available, it would be optimal to use records from the years the specific estimates are based on. As this is not possible, the variable might not give a perfect picture of the tax system quality at the estimation time. We assume it gives sufficient information and can be used as a proxy for tax system quality.

A couple of notes on the Employee and Public Employee variables are in place. These variables were coded because some of the PW studies on underreporting of income looked at groups that were not constrained to estimate the underreporting of income by self-employed compared to employees. The Employees variable picks up whether some groups of employees were included in the underreporting group and the Public Employees variable codes for whether the comparison group is public employees. This means that the underreporting group and the comparison group not only contain self-employed and employees, respectively, which carries onto the global average in chapter 4 and 5, and the investigation into publication selection bias in section 4.2. It should be noted though that the number of estimates this relates to is rather low (the two variables have means of 0.13 and 0.19, respectively). In the multiple MRA analysis in section 5.2, we estimate the corrected effect for self-employed compared to employees by setting Employees and Public Employees equal to zero.

Table 4 on the next page shows a correlation matrix for a sample of the moderator variables and the underreporting factor k. We initially wanted to provide a full correlation matrix, but because we have a big number of moderator variables, that would result in a disordered and chaotic table. Therefore, we chose a sample of moderator variables we regard as particularly interesting and wish to see how correlate especially with k. The table shows a great deal of significant correlations, some of which are high. TaxLevel has a high positive correlation (0.52) with both of the variables GDPpercapita and Corruption. This is expected, as it is reasonable to believe that developed countries with high GDP per capita often have high tax levels, and develped countries often have lower corruption scores.. Furthermore, corruption is significantly negatively correlated with the underreporting factor k. This is also what one would expect, because higher corruption scores mean less perceived corruption which transfer to less underreporting of income. One surprising outcome is however the positive correlation between IV and the underreporting factor k. A possible explanation for this might be the use of poor instruments, which wouldn’t properly solve the transient income problem.

Next, we turn to the issue of publication selection bias, and go through the methodology behind simple meta-regression analysis.

TABLE 4

Correlation matrix

Variables	k	IV	PermProx	TaxLevel	TSQ	GDP	CPI	Head	Emp	P.Emp	P
k	1										
IV	0.108*	1									
PermProx	0.031	-0.263***	1								
TaxLevel	-0.331***	-0.405***	0.075	1							
TaxSystemQuality (TSQ)	-0.037	0.110*	0.094	-0.168**	1						
GDPpercapita (GDP)	-0.168**	-0.392***	0.232***	0.516***	0.138*	1					
Corruption (CPI)	-0.221***	-0.237***	0.114*	0.519***	0.429***	0.583***	1				
Head	0.191***	0.217***	-0.109*	-0.593***	0.222***	-0.142**	-0.283***	1			
Employees (Emp)	-0.018	0.151**	-0.161**	-0.328***	-0.183***	-0.229***	-0.369***	0.257***	1		
PublicEmployees (P.Emp)	0.226***	0.217***	-0.199***	-0.551***	0.327***	-0.174**	-0.225***	0.432***	0.545***	1	
Published (P)	-0.027	-0.016	0.348***	0.089	-0.206***	-0.007	-0.117*	-0.084	-0.284***	-0.476***	1

4. Simple MRA: Publication selection bias

When systematically investigating the literature on a social sciences phenomenon, a significant problem often arises in publication selection bias. "Publication selection is largely the process of choosing research papers, or their results, for statistical significance. As a result, larger, more significant, effects will be overrepresented in the research record" (Stanley & Doucouliagos, 2012, p.51). This behavior might be due to for example the conventional view affecting which papers reviewers accept, that the expected result affects model selection or that significant results are treated more favorably (Card & Krueger, 1995, p.239). In short, researchers and reviewers/editors report only part of the research estimates, which may bias the estimates to one side and, in this instance, distort the global average degree of income underreporting. Researchers must address this issue to avoid possible distorted results.

4.1 Theory

There are several ways to detect publication selection bias in a literature. Two of the most used methods are an informal graphical inspection through a funnel plot and a more formal method with a simple MRA model. The funnel plot in the first method is the same as in chapter 3, plotting the effect size versus its precision (1/SE). The idea is that the estimates should be symmetrically distributed, so that an asymmetrical funnel plot indicates the presence of publication selection bias. Publication selection is less likely to affect the most precise estimates, as these increasingly turn out significant. Thus, these estimates make the top of the funnel graph and should be closest to the true average effect size (Stanley & Doucouliagos, 2012).

The more formal way to test for the presence of publication selection relies on the idea that "researchers who have small samples and low precision, [i.e. high SE], will be forced to search more intensively across model specifications, data and econometric techniques until they find larger [and significant] estimates" (Stanley & Doucouliagos, 2012, p.60). The opposite is true for researchers with higher samples and more precision. Thus, ceteris paribus, publication selection means the estimates are correlated with their standard errors, which leads to the following simple MRA model:

$$effect_i = \beta_0 + \beta_1 SE_i + \varepsilon_i \qquad (28)$$

where $effect_i$ is a reported estimate and SE_i is its standard error. "$\beta_1 SE_i$ models publication selection bias, and estimates of β_0 serve as correction for publication selection bias (as $SE_i \rightarrow 0, E(effect_i) \rightarrow \beta_0$)" (Stanley & Doucouliagos, 2012, p.60). This model can be shown in a slightly transformed funnel plot, with SE is on the horizontal axis and the effect size on the vertical axis.

A problem with model (28) is that we expect different estimates to have different standard errors, making the error term ε_i is heteroskedastic. This problem can be solved by using weighted least squares, with the weights being the inverse of the estimates' variances ($1/SE^2$). An alternative is to divide equation (28) through by SE_i:

$$t_i = \beta_1 + \beta_0(1/SE_i) + v_i \tag{29}$$

where t_i is each estimate's t-statistic ($effect_i/SE_i$) and $v_i = \varepsilon_i/SE_i$, which should make its variance approximately constant (Stanley & Doucouliagos, 2012). The simple WLS-MRA model (29) can then be estimated using ordinary least squares (OLS).

Stanley & Doucouliagos (2012) also include equations (28) and (29) in which the variance is substituted for the standard error. Equation (28) then turns into

$$effect_i = \beta_0 + \beta_1 {SE_i}^2 + \varepsilon_i \tag{30}$$

and equation (29) becomes

$$t_i = \beta_1 SE_i + \beta_0(1/SE_i) + v_i \tag{31}$$

Stanley and Doucouliagos (2014) argue that the 'precision-effect estimate with SE' (PEESE), that is β_0, is the better corrected estimate when a genuine effect is present. They point to simulations showing that PEESE give considerably less bias and is more efficient in most of the cases.

The asymmetry of the funnel plot can be tested when using these simple MRA models. First, the presence of publication selection bias is tested by estimating MRA model (29) and testing the null hypothesis of $\beta_1 = 0$. This is called the funnel-asymmetry test, or FAT for short. If β_1

is significantly different from zero, this is an indication of the presence of publication selection bias.

Next, the presence of a genuine empirical effect when correcting for publication selection bias can be tested by the null hypothesis $\beta_0 = 0$. This is called the precision-effect test (PET). The normal test in meta-analysis is whether β_0 is different from zero. However, in this case, a nonzero effect (i.e. no underreporting) is characterized by k = 1 (and thus $\beta_0 = 1$). If β_0 is significantly different from 1, this indicates a genuine underlying empirical effect beyond the distortion by publication selection.

Finally, if the PET-test indicates that there are a non-zero effect, MRA model (31) can be estimated in order to obtain an estimate of the corrected effect size. The coefficient on precision $(1/SE_i)$, β_0, will be an estimate of this corrected effect size. This estimator is called the precision-effect estimate with standard error (PEESE).

A potential problem with the FAT-PET-PEESE tests is within-study dependence. It is possible that each study is characterized by some idiosyncrasies that cannot be coded explicitly. Thus, there might be dependencies within each study that can have a consequence on the efficiency (not bias), i.e. the standard errors of the estimates in the MRA estimations. These dependencies might also be along other dimensions, such as the dataset source of the different studies.

Stanley & Doucouliagos (2012) propose three ways to mitigate this problem. First, one can use average study effect sizes in the estimations. This is a simple solution, but it reduces the degrees of freedom and the possibility to use within-study variations (in specification etc.) to investigate heterogeneity. Next, they propose using cluster-robust standard errors when estimating the FAT-PET-PEESE MRA models. This yields more conservative standard errors and thus a lower likelihood of finding significant coefficients.

Finally, Stanley and Doucouliagos (2012) propose using unbalanced (multi-level) panel methods. The unbalanced panel version of (28) is

$$effect_{is} = \beta_0 + \beta_1 SE_{is} + v_s + \varepsilon_{is} \quad (32)$$

for estimate i in study s. v_s represents an unobserved study effect. This is assumed either fixed (FEML) or random (REML). The unbalanced panel version of (29) is

$$t_{is} = \beta_1 + \beta_0 (1/SE_{is}) + \mu_s + e_{is} \quad (33)$$

where μ_s represents an unobserved study effect.

A problem with the random-effects model is that it assumes that e_{is} or μ_s are independent of SE or 1/SE. These assumptions are most likely violated as larger standard errors prompt greater experimentation from researchers to find significant results. Thus, the fixed-effects multilevel MRA model might be preferred to the random-effects model. We follow Stanley & Doucouliagos' recommendation and use the fixed effect panel version of the cluster robust WLS-MRA model.

4.2 Results

Publication selection bias is a general problem in economics research. It is easily conceivable that this is a problem in the PW underreporting of income literature as well. Researchers and editors might select higher underreporting factors. We investigate this by first looking at the funnel plot.

The funnel plot in Figure 2 of the estimates' precision versus the underreporting factor shows signs of publication selection bias. The plot is quite asymmetrical with an overweight of estimates on the right side of the plot. In other words, it appears that researchers/editors tend to select and report higher underreporting factors. The direction of this bias is positive, as expected. That means that the simple and weighted average underreporting factors calculated above will be upward biased, yielding too high average underreporting factors.

The funnel plot thereby indicates the presence of publication selection bias. However, we want to investigate this formally as well. To do so, we estimate a few different simple MRA models. These regression results are reported in Table 5 below.

TABLE 5

Simple MRA tests for Publication Selection Bias

	(1) WLS-MRA	(2) FEML	(3) PEESE WLS-MRA
1/SE (β_0) - PET	1.121***	1.066***	1.185***
	(0.0828)	(0.0261)	(0.01000)
SE k			0.0142
			(0.0356)
Constant (β_1) - FAT	1.432*	2.008***	
	(0.657)	(0.271)	
Observations	342	342	342

Standard errors in parentheses

All regressions are estimated using cluster robust standard errors. Column 1 is model (29) estimated using OLS. Column 2 is model (33) estimated using panel methods with fixed effects. Column 3 is model (31) estimated using OLS. The dependent variable is the t-value of k. Estimates are based on the precision (1/SE) of k and/or the standard error of k (SE k).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Before turning to the regression results, it is important to note the use of cluster robust standard errors. A potential problem that needs to be accounted for is potential within-study dependence. 26 of the 30 primary studies report more than one estimate, which means that idiosyncratic factors of the author, consumption survey or otherwise might yield dependencies within each study. To deal with such within-study dependencies, we created a variable that assigns a number to each study and then applied robust standard errors for all the regressions using this variable. Each cluster should contain a study that is independent of the other studies. This is likely not fulfilled for the two Kukk & Staehr studies. These two studies estimate the underreporting factor based on the same consumption survey, so to be sure, we chose to treat the two studies as one for clustering purposes.

Model 1 (WLS-MRA) in Table 5 is our main model. This is model (29) estimated using ordinary least squares (OLS) with cluster robust standard errors. The coefficients β_1 and β_0 allow us to conduct the two tests presented in section 4.1.

The funnel asymmetry test (FAT), which tests for the presence of publication selection bias, looks at the significance and sign of the coefficient β_1. In WLS-MRA, the null hypothesis of $\beta_1 = 0$ is rejected at the 5 percent level. β_1 is positive and significant with a t-value of 2.18 and a p-value of 0.038. This might be evidence of the presence of positive publication selection bias, which supports what the funnel plot suggested.

Next, the precision-effect test (PET) can be used to test whether there is a genuine empirical effect when publication selection bias is accounted for. It is important to remember that we must test whether β_0, and thus k, is different from 1, which is the case of no underreporting. The null hypothesis of $\beta_0 = 1$ is not rejected. β_0 has a t-value of 1.46 and a p-value of 0.15. Thus, PET fails to discover the presence of a genuine empirical effect when the distortion due to publication selection bias is removed.

PET does not indicate a non-zero effect. However, to gain perspective on the numbers, we look at the corrected effect size given by β_0 in column 1. The corrected effect size as estimated by β_0 indicates an underreporting factor of 1.121. Equivalently, this indicates that the underreporting group fails to report 1-1/1.121 = 10.8% of their true incomes. This is substantially less than the 31% left unreported when using the simple average and somewhat less than the 15% left unreported when using the weighted average.

Column 2 shows the results of estimating model (33) with fixed effects (also this with cluster robust standard errors). The use of fixed effects serves to see whether the results are robust across the two different econometric approaches.

The general results in terms of the presence of publication selection bias remain the same with the fixed effect panel model as with the OLS model. FAT indicates the presence of publication selection bias by rejecting the null hypothesis of $\beta_1 = 0$ at the 5 percent level with a t-value of 7.41. However, in contrast to WLS-MRA, PET indicates the presence of a genuine empirical effect beyond selection bias by rejecting the null hypothesis of $\beta_0 = 1$ with a t-value of 2.51 and a p-value of 0.018. The estimated size of the bias and the corrected effect are somewhat different than the OLS model. Publication selection bias is somewhat higher at approximately 2, and the corrected effect is thus as expected a bit lower at 1.066. Using the FEML estimate of k = 1.066, this would imply that the underreporting group on average does not report 1-1/1.066 = 6.2% of their true incomes.

As the fixed effects PET shows evidence of a non-zero effect, it is interesting to use the precision-effect estimate with standard error (PEESE) to gain an estimate of the corrected effect. The estimation result of this model is reported as PEESE WLS-MRA in column 3 in Table 5. The corrected effect size as estimated by β_0 indicates an underreporting factor of 1.185. This estimate is somewhat higher than the estimate of 1.121 given by column 1. The PEESE estimate of k = 1.185 means the underreporting group failed to report 1-1/1.185 = 15.6% of their true incomes. This estimate is very close to the share of unreported income given by the weighted average. In other words, the weighted average might seem like a good estimate of the global underreporting factor.

The overall results of the simple MRA analysis are somewhat robust to the different methods used. We find possible evidence of positive publication selection bias across the different models. There is more disagreement as to the presence of a genuine empirical effect beyond the distortion of publication selection bias. The WLS-MRA results does not reveal such an effect, while the fixed effects model finds evidence of a genuine empirical effect.

4.3 Robustness

As mentioned in section 3.1, our dataset includes two studies that might have a large influence on the results. Paulus (2015) and Skjeggestad & Wæhle (2015) report 60 and 40 estimates, respectively. In addition, they are both not published studies, and the study by Skjeggestad & Wæhle is a master thesis. We want to investigate whether the results in the analysis above still stands in the absence of these studies.

Table 6 reports the results from estimating the models in Table 5 on a reduced sample where the estimates in the Skjeggestad & Wæhle (2015) study have been excluded. Table 7 reports the results from estimating the same regressions on a reduced sample where the estimates from both Skjeggestad & Wæhle (2015) and Paulus (2015) have been excluded.

TABLE 6
Simple MRA tests for Publication Selection Bias: Reduced Sample #1

	(1) WLS-MRA	(2) FEML	(3) PEESE WLS-MRA
1/SE (β_0) - PET	1.115***	1.054***	1.189***
	(0.0873)	(0.0264)	(0.0108)
SE k			0.0146
			(0.0373)
Constant (β_1) - FAT	1.706*	2.368***	
	(0.686)	(0.285)	
Observations	302	302	302

Standard errors in parentheses
All regressions are with cluster robust standard errors. Column 1 is model (29) estimated using OLS. Column 2 is model (33) estimated using panel methods with fixed effects. Column 3 is model (31) estimated using OLS. The dependent variable is the t-value of k. Estimates are based on the precision (1/SE) of k and/or the standard error of k (SE k). The regressions are based on a reduced sample: the master thesis by Skjeggestad & Wæhle (2015) is excluded. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 7
Simple MRA tests for Publication Selection Bias: Reduced Sample #2

	(1) WLS-MRA	(2) FEML	(3) PEESE WLS-MRA
1/SE (β_0) - PET	1.117***	1.068***	1.190***
	(0.0919)	(0.0277)	(0.0120)
SE k			0.00759
			(0.0405)
Constant (β_1) - FAT	1.816*	2.408***	
	(0.836)	(0.335)	
Observations	242	242	242

Standard errors in parentheses
All regressions are with cluster robust standard errors. Column 1 is model (29) estimated using OLS. Column 2 is model (33) estimated using panel methods with fixed effects. Column 3 is model (31) estimated using OLS. The dependent variable is the t-value of k. Estimates are based on the precision (1/SE) of k and/or the standard error of k (SE k).
The regressions are based on a reduced sample: Skjeggestad & Wæhle (2015) and Paulus (2015) are excluded.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results from estimating the models on the reduced samples in Table 6 and 7 show very similar results to those found based on the full sample in Table 5. The funnel asymmetry test yields possible evidence of the presence of publication selection bias. The null hypothesis of $\beta_1 = 0$ is rejected at the 5-percent level in both the WLS-MRA and FEML regressions and for both samples. As in section 4.2, the precision effect test fails to uncover a genuine empirical effect beyond the distortion of selection bias with the WLS-MRA regressions. The null hypothesis of $\beta_0 = 1$ is not rejected at the 5 percent level with t-values of 1.32 and 1.27 for sample #1 and sample #2, respectively. Also as in section 4.2, the fixed effects results in column 2 in Table 6 and Table 7 find evidence of a genuine empirical effect with t-values of 2.05 and 2.45 for sample #1 and sample #2, respectively.

Furthermore, the PEESE estimates of k = 1.189 and k = 1.190 for sample #1 and sample #2, respectively, are very similar to the PEESE estimate of k = 1.185 from the full-sample estimation in section 4.2. The similarity between the results with and without the Skjeggestad & Wæhle (2015) and Paulus (2015) studies shows that the results are robust to including or excluding these studies. Thus, it seems that these two studies did not have an undue influence on the results.

The analysis in section 4.2 above was carried out on the calculated underreporting factor k. Although exp(gamma/beta) treats the different estimates in the same way, it does make some simplifying assumptions and disregard the residual income variances. In addition, the funnel plot and the formal FAT and PET tests require the true distribution of k to be symmetrical in order to see an asymmetrical funnel plot or a significant $\beta 1$ as evidence of publication selection bias. As discussed in section 3.1, there are reasons to believe that the true distribution of k is not symmetrical.

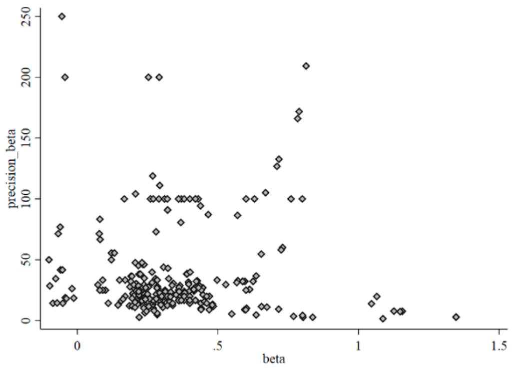
Thus, the results from the funnel plot and simple MRA tests on k should be taken with caution. Even though the FAT tests consistently showed signs of publication selection bias, it is possible that this shouldn't be read as evidence of this bias. However, some further robustness checks are possible to undertake to investigate whether the results from the analyses on k are supported. By carrying out the same analysis on beta and gamma, respectively, we can investigate whether the likely problem for k arises with beta or gamma. Beta and gamma are normal regression coefficients with normal distributions, so it is easier to conclude on selection bias from funnel plots or simple MRAs.

Beta is an estimate of the propensity to consume and is given by the coefficient in front of income in the expenditure function that the primary studies estimate in their application of the PW method. Beta is an important part of our estimate of the underreporting factors: k = exp(gamma/beta). Thus, it will have an important effect on the resulting underreporting factors and it is interesting to investigate whether beta can cause problems of publication selection bias for k.

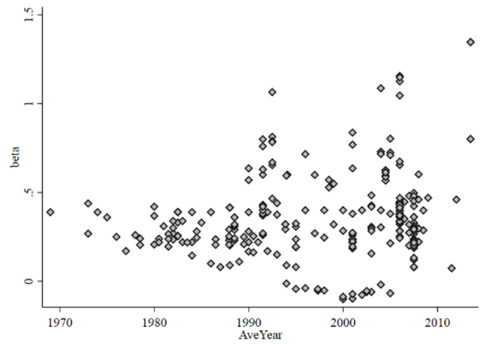
To do this, we produce the same plots for beta that we did for the underreporting factor k. The left plot in Figure 4 shows a funnel plot for beta where beta's precision ($1/SE_{beta}$) is plotted against beta. The right plot in Figure 4 shows a chronological ordering of the betas, where beta is plotted against the average year of the underlying household data on which each estimate is based.

FIGURE 4

4a. Funnel plot of beta



4b. Chronological plot of beta



The funnel plot shows no clear sign of asymmetry. A majority of the estimates is in the region of 0.1 - 0.5, which is quite reasonable, and the rest is spread out rather evenly. Thus, the funnel plot shows no signs of problems with publication selection bias. The chronological plot shows a small indication that betas based on newer data are a bit higher than those based on older data. However, this is not clear.

We also do the more formal tests of publication selection bias. Table 8 below shows the results of the simple MRA analysis conducted on beta. The main model (WLS-MRA) yields a β_1 of -0.8, which is not statistically significant. The fixed effects panel method estimation in column 2 agrees with the main model and yields insignificant β_1. Although researchers and editors don't select beta directly, the beta estimates are important in calculating the size of k. The insignificant β_1's indicate that a potential publication selection bias does not arise with beta.

It can be noted that WLS-MRA gives a significant β_0. The PEESE estimate then yields a propensity to consume of 0.382. However, FEML does not give a significant β_0. It should also be noted that the number of observations are 339, three less than the 342 observations in the regressions based on k. This is because three estimates reported k and its standard error directly without reporting beta and gamma.

TABLE 8

Simple MRA tests for Publication Selection Bias: beta

	(1) WLS-MRA	(2) FEML	(3) PEESE WLS-MRA
1/SEbeta (β_0) - PET	0.375*	0.233	0.382***
	(0.136)	(0.157)	(0.0227)
SEbeta			9.067
			(12.58)
Constant (β_1) - FAT	-0.833	4.200	-1.576
	(3.336)	(5.542)	(1.453)
Observations	339	339	339

Standard errors in parentheses

All regressions are with cluster robust standard errors. Column 1 is model (29) estimated using OLS. Column 2 is model (33) estimated using panel methods with fixed effects. Column 3 is model (31) estimated using OLS.

The dependent variable is the t-value of beta. Estimates are based on the precision (1/SE) of beta and/or the standard error of beta (SE beta).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

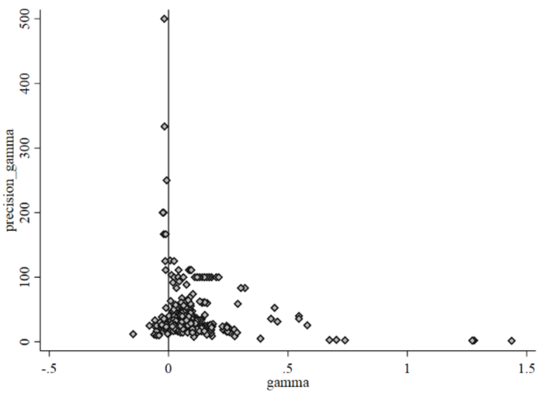
The other part of our estimate of k in exp(gamma/beta) is gamma, which is an estimate of excess consumption at a given level of income. As such it is the coefficient in the consumption function that picks up the differences between the underreporting group and the reference group in terms of consumption and thus income. Thus, it is interesting to investigate whether gamma picks up the same problems with publication selection bias that the analysis on k indicated.

We produce the same plots for gamma in Figure 5. The left plot in Figure 5 shows the funnel plot, which plots gamma's precision ($1/SE_{gamma}$) against gamma. The right plot in Figure 5 shows a chronological ordering of the data. It plots gamma against the average year of the underlying household data on which each estimate is based.

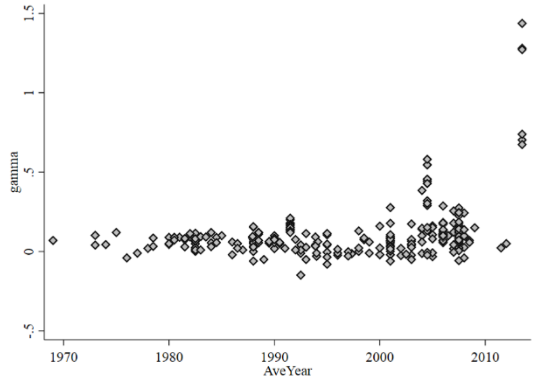
The chronological plot shows no clear signs of a time trend in the gamma estimates. However, there is a slight uptick in gamma estimates based on the newest data. The newer data also shows indications of a slightly bigger variation than earlier data. The funnel plot very much resembles the funnel plot of the underreporting factor k. A majority of the gamma estimates are gathered slightly above zero, and there are signs of asymmetry as the gamma estimates seem biased to the right side of the plot. Thus, the funnel plot indicates the presence of publication selection bias.

FIGURE 5

5a. Funnel plot of gamma



5b. Chronological plot of gamma



We investigate the presence of publication selection bias more formally for gamma as well by estimating the various simple MRA models we also used on k and beta. The regression results of the simple MRAs are shown in Table 9. The results show clear signs of the presence of selection bias across the models. Both the WLS-MRA model in column 1 and the fixed effects model in column 2 yield β_1's that are signficant at the five percent level. β_1 varies from a low of 3.140 and a p-value of 0.013 in the WLS-MRA model to a high of 3.777 and a p-value of approximately zero in the FEML model. This indicates positive selection bias, which is in line with the funnel plot.

TABLE 9
Simple MRA tests for Publication Selection Bias: gamma

	WLS_MRA	FEML	PEESE WLS_MRA
1/SEgamma (β_0) - PET	0.00805	-0.00624	0.00416
	(0.24)	(-0.53)	(0.70)
SEgamma			-9.239*
			(-2.12)
Constant (β_1) - FAT	3.140*	3.777***	3.690***
	(2.67)	(7.18)	(8.41)
Observations	339	339	339

t statistics in parentheses
All regressions are with cluster robust standard errors. Column 1 is model (29) estimated using OLS. Column 2 is model (33) estimated using panel methods with fixed effects. Column 3 is model (31) estimated using OLS.
The dependent variable is the t-value of gamma. Estimates are based on the precision (1/SE) of gamma and/or the standard error of gamma (SE gamma).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The selection bias results in Table 9 corroborates the results we found in the analysis on underreporting factor k. Although researchers and editors don't select gamma directly, the gamma estimates are what indicates the difference between the underreporting group and the reference group (in terms of excess consumption). As such, the positively significant β_1's in Table 9 serve at least as indications of publication selection bias in the literature.

It is also interesting to look at β_0 in Table 9. In the results from the analysis on k in tables 5-7, the WLS-MRAs did not produce significant β_0's, while the fixed effects estimations did. In Table 9 however, β_0 is not found to be significant in any of the models. Thus, this further questions the presence of underreporting because when the distortion from possible publication selection bias is corrected for, no excess consumption remains.

5. Multiple MRA: Heterogeneity

The most important objective of this meta-analysis is to investigate the heterogeneity that characterizes the literature on income underreporting.

5.1 Theory

Although excess heterogeneity, i.e. variation in excess of what random sampling error would imply alone, is present in most economics research, Stanley and Doucouliagos (2012) propose using Cochran's Q-test to test for the presence of excess heterogeneity. Q can be calculated as the sum of squared errors from estimating the simple MRA (29) with no intercept, i.e. regressing the t-value on precision (1/SE). Cochran's Q is chi-squared distributed and the degrees of freedom for this test equals the number of estimates subtracted by one. The Q-test will produce a large test statistic and will be rejected if there is much variation in excess of the random variation captured by the standard error. Based on the Q-statistic, we can easily compute a measure that quantifies the effect of heterogeneity, $I^2 = \frac{Q-df}{Q} \times 100\%$ where df = degrees of freedom. This is a measure of the degree of inconsistency in the studies' results, more precisely the percentage of variation across studies that is caused by heterogeneity rather than chance. Higgins, Thompson, Deeks & Altman (2003) argue that this is the better quantification if one should compare several meta-analyses because it does not depend on the number of studies in the meta-analysis. They also suggest that heterogeneity is low if the measure is between 25-50%, moderate if it is between 50-75% and high if over 75 %.

In order to explain the heterogeneity in the results, we can expand the simple MRA models to multivariate or multiple MRA models. This entails that we use the coded moderator variables as explanatory variables in a regression to account explicitly for the systematic variation in the estimate dataset. Stanley and Doucouliagos (2012) propose the following multiple MRA model:

$$effect_i = \beta_0 + \sum \beta_k Z_{ki} + \beta_1 SE_i + \sum \delta_j SE_i K_{ji} + \varepsilon_i \qquad (34)$$

In this multiple MRA model, there are two sets of moderator variables. Z_k is a set of k different moderator variables that are assumed to represent sources of heterogeneity, while K_j is a set of j moderator variables that "affect the likelihood of selecting an empirical estimate" (Stanley and Doucouliagos, 2012, p.84). Thus, the terms $\beta_1 SE_i$ and $\sum \delta_j SE_i K_{ji}$ represent publication selection bias, while the term $\sum \beta_k Z_{ki}$ represents the heterogeneity. Potential Z-K variables was presented in Table 3 in section 3.1.

The problem of heteroscedasticity from the previous section on publication selection bias is still a problem here. To mitigate this problem, MRA model (34) can be divided through by SE_i to get the following WLS-MRA model, which can be estimated with OLS:

$$t_i = \beta_1 + \sum \delta_j K_{ji} + \frac{\beta_0}{SE_i} + \frac{\sum \beta_k Z_{ki}}{SE_i} + u_i \qquad (35)$$

where $u_i = \varepsilon_i / SE_i$ and has a variance of approximately 1.

As with the simple FAT-PET-PEESE MRA models, there is also the problem of dependence in the multiple MRA case when there are several estimates from each study. The solution is the same as in the simple MRA case, either using cluster-robust standard errors or using unbalanced/multilevel panel methods. We can expand the simple MRA panel model (33) in section 3.3 to the multiple case:

$$t_{is} = \beta_1 + \sum \delta_j K_{jis} + \frac{\beta_0}{SE_{is}} + \frac{\sum \beta_k Z_{kis}}{SE_{is}} + v_s + u_{is} \qquad (36)$$

for estimate i in study s. v_s represents an unobserved study effect. This can be either fixed (FEML) or random (REML). The assumption that the study-level effect variable is uncorrelated with SE or 1/SE is likely violated, which makes fixed-effects the recommended method (Stanley and Doucouliagos, 2012).

When estimating the multiple MRA models, the number of possible models often exceed the number of observations. To reduce the problems of data mining and multicollinearity, Stanley and Doucouliagos (2012) recommend using the general-to-specific (G-to-S) approach, or backward selection, as a structured method. This approach will also ensure that one reaches a specific model in an ordered way. One possibility is therefore to start with an all-inclusive model with all the moderator variables and then remove the least significant variables one by one until only significant variables remain in the estimation.

Finally, the corrected effect, or in our case the corrected rate of average global underreporting, can be calculated by substituting values for the moderator variables into MRA model (34) and using the estimated coefficients from estimating model (35) to calculate the corrected effect. First, SE is set to zero to remove publication selection bias. The chosen values for the other variables are based on professional judgement, either following “best practice”, setting all to the sample means or using the reference group one wishes to look at.

To conclude the methodology section, it is important to note that the simple MRA models usually provide a good approximation of the results. However, to investigate the robustness of the results, it is essential to use more complex multiple MRA models and the econometric solutions to within-study dependencies. We look for consistent results in terms of the presence of publication selection bias. In terms of explaining heterogeneity, moderator variables that are significant across WLS-MRA model (35) without and with cluster-robust standard errors and the FEML MRA give an indication that these are in fact sources of systematic variance in the literature.

5.2 Results

Estimating the multiple MRA models presented above in section 5.1 means we can utilize the coded moderator variables to investigate sources of heterogeneity, look into whether the problem of publication selection bias remains in the multivariate case and estimate a corrected underreporting factor based on different Z-variable settings.

As Stanley & Doucouliagos (2012) point out, they have yet to see a literature (on a specific subject) where differences in the procedure, model specifications and estimation methods did not systematically affect the resulting estimates. However, we wish to formally investigate the presence and extent of heterogeneity. As mentioned in the theory section above (5.1), an appropriate tool for this undertaking is Cochran’s Q-test[2]. We test for excess heterogeneity with Cochran’s Q-test. Applying the Q-test to this data provides a Q-value of 2717.13 with a p-value of approximately zero. This strongly implies that there is excess heterogeneity present in the data (beyond what the random error term would imply alone). That given Q-value provides a heterogeneity measure of 87% (I^2), which indicates the proportion of total variation across studies that is due to heterogeneity rather than chance. The regression coefficients presented earlier contain more variation than they are supposed to and are thereby possibly biased. It is therefore important to take the effect of excess heterogeneity into account so that our overall

[2] This standard meta-analysis approach of testing for heterogeneity has recently been criticized in several studies. No later than last year, Hoaglin (2016) questioned this approach in his article on misunderstandings about Q and Cochran’s Q-test in meta-analysis. Questioning the Q-test naturally entails questioning I^2 as well.

results are not biased. Having tested for and found evidence of excess heterogeneity in the data[3], we move forward with the multiple MRA analysis.

We utilize a multivariate meta-regression model to analyze the sources of heterogeneity. The moderator variables (see Table 3) included in the model were, as explained previously, chosen entirely based on the information available, and what seemed as important factors looking at the literature at hand. Return to section 3.2 for a more thorough walkthrough of the moderating variables listed in Table 3.

Before looking at all the variables in a General-to-Specific approach, we wish to supplement the correlation matrix in Table 4 and gain a perspective on the different moderator variables. We do this by doing group-wise analyses on different groups of variables. This might allow us to identify particularly interesting factors of heterogeneity.

5.2.1 Group-wise Analysis

We do a group-wise analysis by applying model (35) to three different dimensions of the data: variables representing a method to account for transitory income, value added macro variables, and a group we call definition variables. These were chosen because they make a natural separation of our coded moderator variables into groups. In addition, to account for transitory income, and how, are important differences in method across this empirical literature. We also wish to examine if we can explain some variation in a wider context, based on the macro variables we have chosen to collect. Based on meta-regressions on these groups, we will move forward with a General-to-Specific approach on a set of variables. We will then use the specific model to calculate a corrected effect.

Table 10 below shows regressions on variables representing econometric techniques to account for the endogeneity problem due to transitory income fluctuations. Studies account for this problem either with instrument variables or with a proxy for permanent income. The reference category is therefore studies that do not account for the endogeneity problem at all.

[3] We also tested for heterogeneity in beta and gamma, and found evidence for both.

TABLE 10
Multiple MRAs: Method variables

	(1) WLS-MRA: k	(2) FEML: k	(3) WLS-MRA: gamma	(4) WLS-MRA: beta
IV/SE	0.104*	0.00817	-0.00186	-0.0478
	(0.0477)	(0.0188)	(0.0442)	(0.123)
PermProx/SE	0.180**	-0.0302	0.0287	-0.323
	(0.0503)	(0.0276)	(0.0702)	(0.167)
1/SE (β_0)	1.074***	1.062***	0.00446	0.452***
	(0.0711)	(0.0314)	(0.0565)	(0.117)
Constant (β_1)	0.976**	2.048***	3.160*	-0.928
	(0.305)	(0.240)	(1.210)	(2.523)
Observations	342	342	339	339

Standard errors in parentheses
All four regressions are calculated with cluster robust SEs. Column (1) is model (35) estimated using OLS on the underreporting factor k (dependent variable is t = exp(γ/β)/SE k). Column (2) is model (36) estimated using panel methods with fixed effects on k (dependent variable is t = exp(γ/β)/SE k). Estimates in column (1) and (2) are based on the variables divided by SE of k. Column (3) and (4) are model (35) estimated using OLS on gamma and beta separately. Estimates in column (3) are based on the variables divided by SE of gamma. Estimates in column (4) are based on the variables divided by SE of beta.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We see that both the included Z-variables, IV and PermProx, are significant for “k” in the WLS regression at the 5% level. They are however not significant in the fixed-effects model, nor separately for gamma and beta. A few comments on the coefficients are in order. The surprising case of a positive sign for IV’s correlation with “k” emerges here as well. The use of IV as an econometric technique to account for the problem caused by transitory income fluctuations, as opposed to not addressing the issue (using OLS), gives systematically higher estimates of relative underreporting of income. This same result, only higher coefficient, also goes for PermProx. We would expect that the use of instrument variables or a proxy for permanent income reduces the underreporting factor. This is due to possible attenuation bias which pulls beta downwards and thus k upwards because of the transitory income fluctuations. There is measurement noise when using current income instead of permanent income. In this case, the unexpected result might be explained by omitted variable bias. We know from Table 4 that several variables significantly correlate with IV and PermProx and exclusion of these may affect the size and direction of the resulting coefficients. The insignificance in the FEML model also means that too much weight should not be put on the WLS-MRA results. When analyzing these variables’ importance, we are mostly interested in the results from the fixed-effects regression. The resulting coefficients show within-study variations, and an insignificant coefficient means that the specific method (IV or PermProx) has no significant impact on the estimates of k. Significant coefficients in WLS can be caused by other differences between the studies that are not accounted for. It therefore does not seem that IV and PermProx are important for estimation of k, surprisingly.

Due to significance of IV in WLS, we wanted to examine further the specific instruments used. We therefore excluded IV from the previous regression and included the instruments housing, education and capital income instead. The result is presented in Table 11 below.

TABLE 11
Multiple MRAs: Method variables with specific instruments

	(1)	(2)	(3)	(4)
	WLS-MRA: k	FEML: k	WLS-MRA: gamma	WLS-MRA: beta
Education/SE	0.115*	0.00327	0.0697*	0.0957
	(0.0478)	(0.0225)	(0.0305)	(0.0832)
Housing/SE	-0.0349	0.0260	-0.0625*	-0.173
	(0.0673)	(0.0194)	(0.0272)	(0.142)
Capitalincome/SE	-0.0605	-0.0725***	0.0344	0.154
	(0.0764)	(0.0180)	(0.0177)	(0.144)
PermProx/SE	0.154**	-0.0325	-0.000336	-0.314
	(0.0518)	(0.0262)	(0.0539)	(0.156)
1/SE (β_0)	1.085***	1.060***	0.0430	0.436***
	(0.0710)	(0.0300)	(0.0316)	(0.114)
Constant (β_1)	1.135**	2.056***	1.188	-1.414
	(0.389)	(0.227)	(0.626)	(2.857)
Observations	342	342	339	339

Standard errors in parentheses
All four regressions are calculated with cluster robust SEs. Column (1) is model (35) estimated using OLS on the underreporting factor k (dependent variable is t = exp(γ/β)/SE k). Column (2) is model (36) estimated using panel methods with fixed effects on k (dependent variable is t = exp(γ/β)/SE k). Estimates in column (1) and (2) are based on the variables divided by SE of k. Columns (3) and (4) are model (35) estimated using OLS on gamma and beta separately. Estimates in column (3) are based on the variables divided by SE of gamma. Estimates in column (4) are based on the variables divided by SE of beta.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

PermProx is still significant in WLS-MRA on “k”. Education is significant for WLS-MRA on “k” and gamma at the 5% level, housing is only significant for gamma at the 5% level, and Capitalincome is only significant in the fixed effect model for “k”. The three instruments’ coefficients give quite varying significance and directions, but including only one of them (for example Education) does not make much sense. In the following, we will therefore disregard the specific instruments, and use IV instead.

The second group of variables we investigate are the value-added macro variables Corruption, Tax Level, GDP per capita and Tax System Quality. The results of regressions on “k”, gamma and beta are given in Table 12 below.

TABLE 12
Multiple MRAs: Macro variables

	(1) WLS-MRA: k	(2) WLS-MRA: gamma	(3) WLS-MRA: beta
CorruptionScore/SE	0.0326	0.0272***	0.0709*
	(0.0248)	(0.00490)	(0.0305)
TaxLevel/SE	-0.00875	-0.00605*	-0.0107*
	(0.00473)	(0.00256)	(0.00436)
GDPpercapita/SE	0.00129	-0.000133	-0.00241
	(0.00205)	(0.000856)	(0.00228)
TaxSystemQuality/SE	-0.000348	-0.00278	-0.0502***
	(0.00866)	(0.00170)	(0.0102)
1/SE (β_0)	1.075	0.195	4.208***
	(0.608)	(0.129)	(0.756)
Constant (β_1)	0.964**	1.231	3.316*
	(0.324)	(0.844)	(1.392)
Observations	342	339	339

Standard errors in parentheses
All four regressions are calculated with cluster robust SEs. Column (1) is model (35) estimated using OLS on the underreporting factor k (dependent variable is t = exp(γ/β)/SE k). Estimates in column (1) are based on the variables divided by SE of k. Columns (2) and (3) are model (35) estimated using OLS on gamma and beta separately. Estimates in column (2) are based on the variables divided by SE of gamma. Estimates in column (3) are based on the variables divided by SE of beta.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Surprisingly, none of the macro variables are significant for “k” in this regression, only two of them are significant for gamma, and three are significant for beta. These variables also have several significant correlations in Table 4, and may potentially have interesting/important implications if they prove significant in the end. We therefore wish to include them in the general model.

At last, we investigate the significance of different ways variables are defined across the 30 primary studies. In this meta-regression, we included Head, Employees, Public Employees, Share, Share&Status, TotExp and OtherExp. See the results below in Table 13.

TABLE 13
Multiple MRAs: Variable definitions

	(1) WLS-MRA k	(2) FEML k	(3) WLS-MRA gamma	(4) WLS-MRA beta
Head/SE	0.237***	0.0125	0.0420	-0.117
	(0.0444)	(0.0388)	(0.0259)	(0.111)
Employees/SE	-0.0871*	-0.0679	0.0183	0.357*
	(0.0421)	(0.0392)	(0.0547)	(0.142)
PublicEmployees/SE	-0.127*	-0.129**	-0.0986	-0.370**
	(0.0568)	(0.0462)	(0.0569)	(0.117)
Share/SE	0.0916	0.0599	0.0623	0.0902
	(0.0883)	(0.0339)	(0.0530)	(0.107)
Share&Status/SE	0.104**	-0.0372	0.0980	0.131
	(0.0315)	(0.0507)	(0.0544)	(0.107)
TotExp/SE	-0.0167	0.00833	0.115**	0.425***
	(0.0150)	(0.0348)	(0.0374)	(0.108)
OtherExp/SE	-0.000811	0.0343*	0.0669*	0.168
	(0.0328)	(0.0156)	(0.0309)	(0.0973)
1/SE (β_0)	1.033***	1.068***	-0.0405	0.346*
	(0.0252)	(0.0279)	(0.0242)	(0.167)
Constant (β_1)	1.257**	2.011***	3.208**	-0.382
	(0.363)	(0.170)	(1.045)	(2.208)
Observations	342	342	339	339

Standard errors in parentheses
All four regressions are calculated with cluster robust SEs. Column (1) is model (35) estimated using OLS on the underreporting factor k (dependent variable is t = exp(γ/β)/SE k). Column (2) is model (36) estimated using panel methods with fixed effects on k (dependent variable is t = exp(γ/β)/SE k). Estimates in column (1) and (2) are based on the variables divided by SE of k. Columns (3) and (4) are model (35) estimated using OLS on gamma and beta separately. Estimates in column (3) are based on the variables divided by SE of gamma. Estimates in column (4) are based on the variables divided by SE of beta.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In this table, the variable Public Employees stands out with significance on all WLS-MRA regressions at the 5% level, and with consistent signs across all models. This is an unexpected result. It is likely that public employees have weaker opportunities to underreport income than employees in the private sector. Thus, we would expect to see larger underreporting factors when the underreporting group is compared against public employees relative to being compared against both public and private employees. It should be noted that nearly all the estimates that use only public employees as the reference group are from Paulus (2015). Thus, the result is based on limited data and should be taken with caution. The FEML result is perhaps particularly affected by this.

Head and Employees are significant for "k" in the WLS-MRA. So is Share&Status, but Share turns out insignificant across all models. TotExp are significant for both gamma and beta, and OtherExp is significant for the fixed effects model for "k", as well as for WLS on gamma.

5.2.2 The General-to-Specific Approach

Having obtained an overview of the different moderator variables, we now combine the variables in a multiple MRA model. We choose to follow Stanley & Doucouliagos' (2012) recommendation by including all moderator variables in the general model. These variables then make up the Z- and possible K-variables in the general model. The Z variables represent research dimensions that explain the reported heterogeneity among results, while the K variables seek to explain publication selection bias. The General-to-Specific approach is then used to simplify the general model, ending up with a specific model. We follow Stanley & Doucouliagos' example in their book about meta-analysis methodology (Stanley & Doucouliagos, 2012) and conduct the general-to-specific approach on WLS-MRA model (35) without cluster robust standard errors. After the specific model is obtained based on the WLS model without cluster robust standard errors, the specific model is reported as cluster robust WLS and FEML as well.

The variables that make up the general model include IV and PermProx from the Method MRA analysis in Table 11. These were both significant at the 5-percent level in the WLS-MRA estimations on k in column 1. The macro variables are interesting to include in the G-to-S model. Thus, we choose to include all four macro variables despite varying significance. From the variable definition variables, we choose to include head and both the employee variables, as these are all significant in the WLS-MRA with k as the dependent variable. The two dummy variables representing the consumption category, TotExp and OtherExp, are also included. In addition to variables from the group-wise analyses, we also include AveYear to control for a potential time trend. The Published variable is also included. This will seek to explain differences between published and unpublished studies, and as such it is a measure of quality. Finally, we included Panel as a Z-variable.

We also include a set of K-variables to account for publication selection bias in the multivariate case. To begin with, all the included Z-variables are also used as K-variables. However, the macro variables as well as the AveYear and Published variables are either study-invariant variables or little likely to affect selection decisions. We also consider Share, Share&Status, TotExp and Otherexp sa unlikely to affect selection decisions. We thus end up with IV, PermProx, Head, Employees, PublicEmployees and Panel as the K-variables.

These Z and K-variables are then used along with the t-value of the underreporting factor k and k's precision (1/SE) in the estimation of WLS-MRA model (35). To see the full general model, go to table A1[4] in appendix A. The least significant variable was removed one by one until only significant variables remained. The resulting specific model is reported as column 1 in Table 14. Going from the general to the specific model, TaxLevel, Corruption, Share, ShareStatus, TotExp and OtherExp were removed as Z-variables due to insignificance, and employees was excluded as a K-variable of the same reason.

We need to take within-study dependence into account here like we did in the simple MRA context. We do this in the same way that we did before, by using cluster robust standard errors as well as panel methods. The results of applying cluster robust standard errors to the WLS-MRA model is reported in column 2 in Table 14. Column 3 of the same table shows the results of estimating the fixed effects panel model.

[4] The regressions of the general model are placed in the appendix because it is only used as a step to obtain the specific model.

TABLE 14
Multiple MRAs - specific model

	(1) WLS-MRA	(2) WLS-MRA cluster robust	(3) FEML cluster robust
IV/SE	-0.123***	-0.123	0.0288
	(0.0278)	(0.0735)	(0.0311)
PermProx/SE	-0.191***	-0.191	-0.0787***
	(0.0553)	(0.110)	(0.0187)
TaxLevel/SE	-0.00268*	-0.00268	
	(0.00121)	(0.00324)	
TaxSystemQuality/SE	0.0204***	0.0204**	
	(0.00248)	(0.00695)	
Head/SE	0.161***	0.161**	0.0189
	(0.0262)	(0.0482)	(0.0373)
Employees/SE	0.197***	0.197*	-0.0582
	(0.0444)	(0.0946)	(0.0309)
PublicEmployees/SE	-0.579***	-0.579**	-0.141***
	(0.0685)	(0.159)	(0.0357)
Published/SE	0.0804***	0.0804*	
	(0.0234)	(0.0374)	
1/SE (β_0)	-0.517**	-0.517	1.064***
	(0.189)	(0.508)	(0.0341)
IV	0.976**	0.976*	-0.468
	(0.337)	(0.420)	(0.513)
PermProx	3.357***	3.357	0.776*
	(0.644)	(1.852)	(0.298)
Head	-0.822*	-0.822	-0.542
	(0.354)	(0.431)	(0.587)
Public Employees	2.365***	2.365***	0.615***
	(0.466)	(0.452)	(0.146)
Constant (β_1)	0.0579	0.0579	2.425***
	(0.304)	(0.327)	(0.472)
Observations	342	342	342

Standard errors in parentheses
Columns 1 and 2 are model (35) estimated using OLS. Column 3 is model (36) estimated using panel methods with fixed effects, excluding study-invariant variables. Columns 2 and 3 are with cluster robust standard errors. The dependent variable is the t-value of k. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of estimating the specific model in Table 14 provide a basis for investigating the sources of heterogeneity. We will start this discussion by looking at the most interesting variables. We will then take a more general, overlooking view and test for publication selection bias and whether there are genuine, systematic factors in the data. Finally, we will estimate a corrected underreporting factor by removing the distortion due to publication selection bias and using either the mean values of the Z variables or what we think might be considered as "best practice".

The first two Z variables in the specific model are IV and PermProx. These variables represent the econometric technique employed by an estimate to deal with the endogeneity problem due to transitory income fluctuations. The use of instrument variables or a proxy for permanent income is compared against the use of no such technique. The results in column 1 in Table 14 indicate that the use of instrument variables or a proxy for permanent income reduces the underreporting factor. However, when taking within-study dependence into account, IV is insignificant in both the WLS-MRA and the fixed effect model. Thus, the result for IV seems to be rather weak. For PermProx, the results are slightly stronger, with a negative and significant coefficient in the fixed effect model. This means that the use of a proxy for permanent income gives systematically lower estimates of underreporting, which contrasts with the positive correlation coefficients in Table 4 and the positive regression coefficients in Table 10. This result is more in line with expectations. As Engström & Hagen (2017) note, the use of permanent income based on average income measures of several years of register income data can reduce the attenuation bias that transitory income fluctuations entail. Attenuation bias is said to lead to lower betas and thus higher underreporting factors, and as such, the negative coefficient on PermProx corroborates the results from Engström & Hagen (2017) and Hurst et al (2014).

The only macro variables that are left in the specific model are TaxLevel and TaxSystemQuality. TaxLevel is negative and significant at the 5-percent level with a coefficient of -0.00268 in column 1. This indicates that a country with a higher tax level systematically sees less underreporting of income (by the underreporting group relative to the reference group). However, the absolute value of the coefficient is very low, so it does not seem to be practically significant. Also, more importantly, the significance disappears as we take within-study dependence into account in column 2. This is in line with Bovi's (2002) findings that the size of the underground economy is affected to a lesser degree by taxation. TaxSystemQuality has a positive and significant coefficient in both column 1 and column 2. This indicates that a higher measure of tax system quality is linked with higher rates of underreporting. This result is quite unexpected and in contrast with the negative simple correlation coefficient in Table 4.

One of the variable definition variables that are left in the specific model is Head. Its coefficient is positive and significant in the WLS-MRA regression with cluster robust standard errors. This is in line with the simple correlation coefficient in Table 4 and the results in the group-wise analysis in Table 13. The positive coefficient would imply that the estimates that define households by looking at only the head of household usually report higher underreporting factors. It should be noted that the significance is lost with the fixed effects model.

In addition, the Published variable is positive and significant in the WLS-MRA with cluster robust standard errors in column 2 in Table 14. The positive coefficient of 0.0804 indicates that published studies generally report higher rates of underreporting.

In the following, we will conduct simple restrictions tests to investigate the presence of publication selection bias and genuine, systematic patterns in the multivariate context. In the simple MRA analyses, we found indications that publication selection bias might be present in the PW income underreporting literature. In the multivariate context, publication selection bias is captured by the intercept and the K-variables. By testing the joint hypothesis that all K-variables' coefficients are zero with the F-test, we examine whether selection bias is present after taking different sources of heterogeneity into account. Using column 1 in Table 14, the joint hypothesis is rejected ($F(5, 328) = 20.55$, $p \approx 0$), which is a strong indication that selection bias is present. The result still stands when looking at the cluster robust WLS-MRA model in column 2 and the fixed effects model in column 3. The resulting F-values are 10.35 and 44.57, respectively, with p-values approximately equal to zero. It is also interesting to see whether there is evidence of genuine, systematic patterns in the literature after accounting for the effect of publication selection bias. An F-test of whether all Z variables, in the specific model, are zero can be used to test this. The result of this test applied to the column 1 results in Table 14 is $F(8, 328) = 32.56$, $p \approx 0$. This result remains when looking at the within-study dependence models. The F-values of applying the test of joint significance to columns 2 and 3 in Table 14 are 41.85 and 38.54, respectively, with $p \approx 0$. This is evidence that there are in fact genuine systematic patterns among reported income underreporting research findings.

Next, we want to calculate the corrected effect. We use the coefficients from column 1 in Table 14 and substitute values for the coefficients and for the Z into model (34). First, we remove the effect of publication selection bias by setting $SE = 0$. That means that the effect of the K-variables is set to zero. Next, we substitute reasonable values for the Z-variables in two ways. First, to reduce the likelihood and effect of poor judgement in selecting values for the Z-variables, we choose to substitute the sample means of the Z-variables into the model (34). This leads to a corrected estimate of the underreporting factor of k = 1.09. This is slightly lower than the corrected effect of k = 1.121 from the simple MRA (WLS-MRA) and would imply that the underreporting group failed to report 1 - 1/1.09 = 8.3% of their true incomes.

Second, we attempt to substitute what we think constitute a good study in this research field for some relevant Z-variables and substitute sample means for the rest. We decide to set the variable PermProx to one, indicating the use of a proxy for permanent income in the consumption function. This means instrument variable methods were not used, so IV is set to zero. We assume that published studies are of a higher quality than unpublished studies, so we set Published to one. In addition, the main underreporting group of interest is self-employed, so Employees is set to zero, and the main reference group of interest is employees, so PublicEmployees is also set to zero. The rest of the Z-variables are set to their sample means. This gives an underreporting factor of 1.128, which is slightly higher compared to using sample means for all Z-variables. An underreporting factor of k = 1.128 would imply that self-employed households on average do not report 1- 1/1.128 = 11.3% of their true incomes compared to employees.

5.3 Robustness

As we did in the simple-MRA section, we also want to investigate whether the results in the multiple MRA above still stand in the absence of the master thesis by Skjeggestad & Wæhle (2015) and the Paulus (2015) study.

Table 15 reports column (2) and (3) from table 14 for two reduced samples. Column (1) and (2) show the MRA results for a sample of 29 studies where Skjeggestad & Wæhle (2015) is excluded. This exclusion results in 302 observations, as the master thesis has 40 estimates. Columns (3) and (4) are the meta-regressions of a sample of 28 studies, because now Paulus (2015) is excluded in addition to Skjeggestad & Wæhle (2015). This sample consists of 242 observations, because further 60 observations are excluded due to Paulus.

TABLE 15

Multiple MRAs - specific model reduced sample

	(1) WLS-MRA sample1	(2) FEML sample1	(3) WLS-MRA sample2	(4) FEML sample2
IV/SE	-0.128	0.0298	-0.125	0.0590
	(0.0779)	(0.0377)	(0.0802)	(0.0386)
PermProx/SE	-0.174	-0.0692***	-0.211	-0.0766**
	(0.109)	(0.0180)	(0.114)	(0.0229)
TaxLevel/SE	-0.00339		-0.00641	
	(0.00328)		(0.00317)	
TaxSystemQuality/SE	0.0190*		0.0207*	
	(0.00720)		(0.00746)	
Head/SE	0.192***	0.0302	0.174***	0.0115
	(0.0486)	(0.0558)	(0.0463)	(0.0574)
Employees/SE	0.184	-0.0588	0.219	0.0279
	(0.110)	(0.0304)	(0.120)	(0.0234)
PublicEmployees/SE	-0.553**	-0.137**	-0.150	0.112
	(0.161)	(0.0373)	(0.150)	(0.0751)
Published/SE	0.0840		0.0607	
	(0.0559)		(0.0687)	
1/SE (β_0)	-0.421	1.048***	-0.495	1.044***
	(0.529)	(0.0235)	(0.563)	(0.0194)
IV	0.717	-0.503	0.586	-1.360*
	(0.542)	(0.562)	(0.706)	(0.541)
PermProx	3.024	0.646	3.455	0.457
	(1.911)	(0.347)	(1.798)	(0.351)
Head	-1.145**	-0.676	-0.978	-0.462
	(0.351)	(0.632)	(0.484)	(0.789)
Public Employees	2.242***	0.603***	-0.172	-0.365
	(0.476)	(0.159)	(0.902)	(0.251)
Constant (β_1)	0.669	2.900***	0.733	3.351***
	(0.446)	(0.496)	(0.611)	(0.488)
Observations	302	302	242	242

Standard errors in parentheses

All regressions are estimated using cluster robust standard errors. The variables are calculated using SE of k. Column (1) and (2) are based on 29 studies, the primary studies excluded Skjeggestad & Wæhle (2015). Column (3) and (4) are based on 28 studies, the primary studies excluded Skjeggesatd & Wæhle (2015) and Paulus (2015). Column (1) and (3) are model (35) estimated using OLS. Column (2) and (4) are model (36) estimated using panel methods with fixed effects, excluding study-invariant variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15 shows quite similar results as column (2) and (3) in Table 14 for the Z-variables. IV and TaxLevel are still not significant, but PermProx, TaxSystemQuality and Head give similar significant results as in Table 14. Head actually gives somewhat higher significant coefficients in Table 15. PublicEmployees are still significant for sample 1 with slightly lower coefficients (in absolute value), but shows no sign of significance in sample 2. This variable's coefficient actually changes direction from negative in WLS to positive in FEML for sample 2. Most of the estimates using public employees as reference group come from Paulus (2015). The interpretation of PublicEmployees' coefficients should not be emphasized because the significance is lost when we exclude Paulus (2015). In addition, Employees and Published are no longer significant with reduced samples.

This robustness check mainly supports our conclusions in section 5.2. IV and TaxLevel do not seem important. The variables PermProx, TaxSystemQuality and Head systematically affect the reported estimates of relative income underreporting. As Published and Employees no longer turn out significant, the inferences made in section 5.2 about these two variables have weaker reliability. Overall, we think the patterns found in section 5.2 are consistent. When we test for genuine systematic patterns with an F-test as in the previous section, we find evidence supporting this for all four models[5]. It should be noted that publication selection bias is also present here after taking the different sources of heterogeneity into account. This is examined with an F-test by testing the joint hypothesis that all K-variables' coefficients are zero.[6]

As we continuously have investigated the separate cases of gamma and beta throughout the thesis, we wish to do a robustness check with these as well. Table 16 reports column (2) and (3) from Table 14 (WLS-MRA and FEML, all with cluster robust standard errors) on both gamma and beta.

[5] The null hypothesis is rejected in all four cases, with F-statistics of 84.29, 35.75, 91.06 and 15.15.
[6] The null hypothesis is rejected in all four cases, with F-statistics of 10.79, 96.95, 3.26 and 77.62.

TABLE 16: specific model on gamma and beta

	(1)	(2)	(3)	(4)
	WLS-MRA	FEML	WLS-MRA	FEML
IV/SE	-0.0590	0.0194	0.0107	-0.0845
	(0.0373)	(0.0292)	(0.136)	(0.121)
PermProx/SE	-0.0786	0.0266	-0.470***	-0.423***
	(0.0449)	(0.0358)	(0.103)	(0.104)
TaxLevel/SE	-0.00657*		-0.00574	
	(0.00286)		(0.00342)	
TaxSystemQuality/SE	0.00184		-0.0274**	
	(0.00141)		(0.00866)	
Head/SE	-0.0243	-0.0541***	0.0169	-0.151
	(0.0330)	(0.0144)	(0.101)	(0.111)
Employees/SE	0.0467	-0.0522	0.0168	0.129
	(0.0494)	(0.0279)	(0.0954)	(0.115)
PublicEmployees/SE	-0.267*	0.0175	-0.228	-0.193*
	(0.111)	(0.0564)	(0.119)	(0.0719)
Published/SE	-0.0279		-0.0154	
	(0.0285)		(0.0577)	
1/SE (β_0)	0.0334	0.0293	2.759***	0.537**
	(0.125)	(0.0200)	(0.669)	(0.165)
IV	2.348	-0.633	1.153	13.14*
	(1.307)	(1.536)	(3.624)	(5.940)
PermProx	7.355***	-2.462	20.23***	17.78***
	(1.953)	(2.097)	(4.143)	(4.721)
Head	1.559	0.344	-2.032	2.190
	(1.288)	(1.299)	(2.884)	(4.084)
Public Employees	3.216	-0.0212	5.438	4.938*
	(2.389)	(1.443)	(2.838)	(2.101)
Constant (β_1)	0.652	3.457*	-0.944	-12.67
	(1.168)	(1.444)	(2.692)	(7.735)
Observations	339	339	339	339

Standard errors in parentheses. All columns are calculated using cluster robust standards errors. Column (1) is model (35) on gamma estimated using OLS. Column (2) is model (36) on gamma estimated using panel methods with fixed effects, excluding study-invariant variables. Both column (1) and (2) are estimations based on variables calculated with SE of gamma. Column (3) is model (35) on beta estimated using OLS. Column (4) is model (36) on beta estimated using panel methods with fixed effects, excluding study-invariant variables. Both column (3) and (4) are estimations based on variables calculated with SE of beta.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Some of the patterns found earlier are also existent in the Table 16. IV, Employees and Published are significant for neither gamma nor beta. PermProx and TaxSystemQuality are only significant for beta, while Head and TaxLevel show signs of significance for gamma. Public Employees gives significant results for both.

The same F-tests as previously used are employed to examine publication selection bias and genuine systematic patterns for beta and gamma. We find evidence of both.[7] Regarding publication selection bias, we have to remember that gamma and beta are selected in pairs and not seperately. It is therefore not certain that this test applies in the same way for gamma and beta. We do however take the overwhelming indications of publication selection bias as support of our conclusions regarding the underreporting factor k.

In the process of analyzing, we have experienced that meta-regression analysis can be highly sensitive. The general-to-specific approach gave different specific models depending on which variables were included in the general model in the first place, naturally. We also saw from Table 15 that the sample of studies (sample 1 and 2 as opposed to the full sample) included in the analysis also has an impact, and we wanted to investigate this further. The study authored by Paulus (2015) was excluded in sample 2 due to the number of estimates (60) it provided and thereby its possible influence. The master thesis by Skjeggestad & Wæhle (2015) was excluded in sample 1 and 2 partly due to its number of estimates (40) and partly due to the quality of the study. There is another study, in this case published, we question the quality of. This regards the Obwonas (1999) study from Uganda, mainly because we had do make some assumptions in the coding process due to lack of information. Therefore, we run the specific model on a sample of 29 studies, only this time by excluding Obwona (1999). The results are shown in Table A2 in the Appendix A.

In this case, the results are quite different compared to earlier, except for a consistent significance of PermProx. Now suddenly, using IV as a method to account for the endogeneity problem caused by transitory income fluctuations proves significant at the 5% level. With a negative coefficient, this gives systematically lower estimates of income underreporting (by the underreporting group compared to the reference group). This coincides with what we would expect from the beginning. Another analysis where we exclude Obwona (1999) also speaks in

[7] The null hypothesis of no publication selection bias is rejected across all models with F-statistics of 10.52, 30.22, 7.26 and 19.88. The null hypothesis of no genuine systematic patterns is also rejected across all models with F-statistics of 3.58, 33.40, 29.82 and 41.74.

favor of the importance of IV and PermProx. We repeated the regressions in Table 10 (this time with sample 3), resulting in significant coefficients in the fixed-effects model of -0.04 and -0.06, respectively.

Another variable that turn out significant is Published. We would expect that published studies tend to report higher estimates of income underreporting. This analysis supports that hypothesis with a significantly positive coefficient of 0.105 at the 5% level. The empirical study by Obwona estimating underreporting of income by self-employed in Uganda, is a published study, and finds no evidence of relative underreporting[8]. It is therefore reasonable that Published becomes (at least more) positively significant when such a study is excluded from the sample.

It should also be mentioned that Head and Employees are no longer significant, another piece of evidence that suggests the results are highly sensitive and conclusions should not be made lightly.

[8] See reported k in Table 2 in section 3.1

6. Conclusion

This meta-analysis has uncovered evidence that publication selection bias is likely a problem in the PW underreporting literature. The funnel-asymmetry test was rejected across all MRA specifications and samples, which indicates a certain robustness to the result. Both the simple MRA tests and the multiple MRA tests corroborated what the asymmetrical-looking funnel plots suggested, and the results from the analysis on k alone was supported by the analysis on beta and gamma separately. However, the problems with symmetry of the true distribution of k means this result should be taken with caution.

The likely presence of publication selection bias is not unexpected in economics research. However, when present, it can introduce significant biases to simple average measures and needs to be taken into account and corrected for.

The multiple MRA analyses uncovered genuine, systematic patterns across all specifications and methods, including those that considered within-study dependence. This is also not an unexpected result in economics research. Furthermore, the multiple MRAs found several variables that seem to be sources of heterogeneity in the PW underreporting literature. One variable which was significant across different MRA models and samples was PermProx. It seems that using a proxy (typically an average income measure) for permanent income in the expenditure function estimations systematically leads to lower underreporting factors. This means that the use of register based income data has a significant effect on underreporting research. MRA results did however suggest that IV is not important when estimating income underreporting, despite what was expected. Further robustness analyses questioned this conclusion and indicated a possible (weak) negative systematic effect on k.

One of the important aspects of this meta-analysis is the estimation of a global average underreporting factor. This estimation is likely affected by publication selection bias. While the simple average measure of the underreporting factor yielded underreporting of 31%, the simple MRA implied that 16% of true incomes were left unreported. However, it is important to utilize the heterogeneity that the moderator variables capture in this calculation. A “best practice” PW study might use register based income data and might be one that is published. Using these inputs, the multiple MRA results imply that self-employed underreport 11% of their true incomes compared to employees.

We should remind you that these numbers are based on our calculations of the underreporting factors using the Hurst et. al (2014) method. Along the way, we learned that a more thorough meta-analysis investigating several dimensions of estimations of k in the PW income underreporting literature is desirable. Due to time and resource restrictions, we were not able to accomplish that at this time, but is it a suggestion to further work on the topic.

Appendix A

TABLE A1: Multiple MRAs - General model

	(1) WLS	(2) Cluster robust	(3) FEML
IV/SE	-0.116***	-0.116	0.0288
	(0.0290)	(0.0682)	(0.0313)
PermProx/SE	-0.181**	-0.181	-0.0787***
	(0.0594)	(0.121)	(0.0187)
Corruption/SE	0.00555	0.00555	
	(0.0102)	(0.0167)	
TaxLevel/SE	-0.00277	-0.00277	
	(0.00183)	(0.00466)	
GDPpercapita/SE	-0.000253	-0.000253	
	(0.00000141)	(0.00000132)	
TaxSystemQuality/SE	0.0187***	0.0187**	
	(0.00389)	(0.00631)	
Head/SE	0.155***	0.155*	0.0189
	(0.0319)	(0.0646)	(0.0374)
Employees/SE	0.239***	0.239	-0.0596
	(0.0581)	(0.133)	(0.0653)
PublicEmployees/SE	-0.529***	-0.529**	-0.141***
	(0.0761)	(0.148)	(0.0376)
Share/SE	-0.0170	-0.0170	
	(0.0311)	(0.0654)	
ShareStatus/SE	-0.0345	-0.0345	
	(0.0357)	(0.0673)	
TotExp/SE	-0.0243	-0.0243	
	(0.0185)	(0.0186)	
OtherExp/SE	-0.0217	-0.0217	
	(0.0221)	(0.0219)	
Published/SE	0.0956***	0.0956	
	(0.0255)	(0.0522)	
AveYear/SE	-0.00132	-0.00132	
	(0.00169)	(0.00304)	
1/SE (β_0)	-0.422	-0.422	1.064***
	(0.278)	(0.429)	(0.0342)
IV	0.985**	0.985**	-0.467
	(0.360)	(0.356)	(0.526)
PermProx	3.486***	3.486	0.776*
	(0.658)	(1.942)	(0.300)
Head	-0.791*	-0.791	-0.542
	(0.402)	(0.563)	(0.589)
Employees	-0.608	-0.608	0.0123
	(0.503)	(0.581)	(0.402)
Public Employees	2.457***	2.457***	0.611**
	(0.492)	(0.475)	(0.174)
Constant (β_1)	0.0794	0.0794	2.425***
	(0.312)	(0.365)	(0.473)
Observations	342	342	342

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE A2
Multiple MRAs - specific model with sample 3

	(1) WLS-MRA	(2) FEML
IV/SE	-0.161*	-0.0454*
	(0.0643)	(0.0197)
PermProx/SE	-0.213*	-0.123***
	(0.0988)	(0.0159)
TaxLevel/SE	-0.00894*	
	(0.00382)	
TaxSystemQuality/SE	-0.00762	
	(0.00735)	
Head/SE	0.00186	-0.0207
	(0.0697)	(0.0365)
Employees/SE	-0.0500	-0.0484
	(0.0827)	(0.0312)
PublicEmployees/SE	-0.212	-0.173***
	(0.117)	(0.0361)
Published/SE	0.105*	
	(0.0425)	
1/SE (β_0)	2.052**	1.185***
	(0.659)	(0.0272)
IV	1.183*	0.134
	(0.466)	(0.339)
PermProx	3.676*	1.080***
	(1.765)	(0.267)
Head	0.101	-0.307
	(0.581)	(0.519)
Public Employees	1.361*	0.546***
	(0.501)	(0.124)
Constant (β_0)	-0.312	1.605***
	(0.289)	(0.314)
Observations	335	335

Both models are calculated using cluster robust standard errors. Sample 3 equals the full sample excluding Obwona (1999), that is 29 studies. Column (1) is model (35) estimated using OLS. Column (2) is model (36) estimated using panel methods with fixed effects, excluding all study-invariant variables. The estimations are based on variables calculated using SE of k.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix B

Stanley, Tom <Stanley@hendrix.edu>
ti 03.10, 15.25

ReportingGuidelines.pdf
432 KB

Last ned

Dear Erik and Sunniva:

Well, good luck with your thesis. They are always daunting, at least to begin with.

I am attaching our MAER-Net guidelines. They give a very nice yet short sketch about what we expect 'state-of-the-art' meta-analyses in economics to do.

Below are my answers to your questions.

Best

Tom

T.D. Stanley
Julia Mobley Professor of Economics, Emeritus
Hendrix College

...

To: Stanley, Tom <Stanley@hendrix.edu>
Subject: Meta-analysis questions

Dear Dr. Stanley,

We are two students from the Norwegian School of Economics and are currently writing our master thesis. The thesis is a meta-analysis of findings in the international literature on underreporting of income by self-employed.

It has been very helpful reading your book "Meta-Regression Analysis in Economics and Business" and we are convinced that you're the person to ask for help in this field.
We hope that you are able to make some time to answer the following questions.

First, in the second chapter of your book regarding unpublished studies you state that *"One danger with the use of unpublished studies is that there is a risk that the estimates in the published version might change"*. Our question is, when we have a working paper which is later published and the estimates and/or specifications/samples are a bit different, should we only include the published study? We would code only the published paper version, or whatever is the last version. 2 reasons. You do not want to count the same estimate multiple times, and perhaps some errors (even typos) were discovered along the way.

Secondly, we are wondering whether every reported estimate in a study has to be included - for example, can we include only the main specification and disregard alternative specifications /robustness checks?
We include them all and code for the differences. These multiple estimates are important in explaining the heterogeneity in methods, approaches, specification, etc. But then, you also need to account for the dependence with a study—FE panel; cluster robust. . .

Appendix C

DOFILE

Clear
capture log close
set more off

cd "\\Lire\Stud$\s136020\Masteroppgave\STATA\Final dataset\Multiple MRA\Final
log using Meta-analysis_log, text replace
import excel "M:\Masteroppgave\Final dataset\Final dataset.xlsx", sheet("Estimates") firstrow
ssc install estout

replace Studie = 14 if Studie == 25

//Simple MRA – Publication selection bias
gen t = k/SEk
gen precision = 1/SEk
gen AveYear= AveYearbaseyear2000+2000

//Funnel plot, k
scatter precision k, xline(1) msize(small) msymbol(diamond) plotregion(style(none)) ///
graphregion(fcolor(white)) mfcolor(gs12) mlcolor(gs0) mlwidth(medium) ///
saving(graph_funnel, replace)

// Chronological ordering of effect sizes, k
scatter k AveYear, msize(medsmall) msymbol(diamond) plotregion(style(none)) ///
graphregion(fcolor(white)) mfcolor(gs12) mlcolor(gs0) mlwidth(medium) ///

//Correlation matrix
estpost correlate k IV PermProx Taxsystemleveltaxrevenue ///
Payingtaxeskvalitetpåskatte GDPpercapita Corruptionindexscore Head ///
Employees PublicEmployees Published, matrix listwise
est store c1
esttab * using Correlation_matrix_k.rtf, unstack not noobs compresssaving(graph_chrono, replace)
eststo clear

// Simple MRAs for FAT-PET-PEESE-testing
eststo clear
eststo: quietly regress t precision, robust cluster(Studie) //WLS-MRA cluster robust
test precision=1 //PET
xtset Studie
eststo: quietly xtreg t precision, fe vce(cluster Studie) //FEML
test precision=1 //PET
eststo: quietly regress t SEk precision, noconst //PEESE
eststo: quietly regress t precision if Studie != 26, robust cluster(Studie) //WLS-MRA cluster robust
wo/masteroppgave
test precision=1 //PET
eststo: quietly regress t precision if Studie != 26 & Studie != 19, robust cluster(Studie) //WLS-MRA
cluster robust wo/masteroppgave&Paulus
esttab using PublicationSelection_simpleMRAs3.rtf, label ///
title(Selection publication bias simple MRAs) ///
mtitles("WLS-MRA" "FEML" "PEESE WLS-MRA" "WLS-MRA reduced sample1" ///

"WLS-MRA reduced sample2") addnote("Remember: all regressions are with cluster ///
robust standard errors") se replace
eststo clear

//Robustness checks - publication selection bias

/*Without Masteroppgave (studie 26)*/
eststo clear
eststo: quietly regress t precision if Studie != 26, robust cluster(Studie) //WLS-MRA cluster robust
test precision=1 //PET
xtset Studie
eststo: quietly xtreg t precision if Studie != 26, fe vce(cluster Studie)
test precision=1 //PET
eststo: quietly regress t SEk precision if Studie != 26, noconst //PEESE
test precision=1 //PET
esttab using Robustness_simpleMRAs1.rtf, label ///
title(Selection publication bias simple MRAs Reduced sample1) ///
mtitles("WLS-MRA" "FEML" "PEESE WLS-MRA") ///
addnote("Remember: all regressions are with cluster robust standard errors") se replace
eststo clear

/*Without Materoppgave & Paulus (studie 19 og 26)*/
eststo clear
eststo: quietly regress t precision if Studie != 26 & Studie != 19, robust cluster(Studie) //WLS-MRA cluster robust wo/masteroppgave&Paulus
test precision=1 //PET
xtset Studie
eststo: quietly xtreg t precision if Studie != 26 & Studie != 19, fe vce(cluster Studie)
test precision=1 //PET
eststo: quietly regress t SEk precision if Studie != 26 & Studie != 19, noconst //PEESE
test precision=1 //PET
esttab using Robustness_simpleMRAs2.rtf, label ///
title(Selection publication bias simple MRAs Reduced sample2) ///
mtitles("WLS-MRA" "FEML" "PEESE WLS-MRA") ///
addnote("Remember: all regressions are with cluster robust standard errors") se replace
eststo clear

/*Publication selection with beta and gamma*/

gen t_beta = Betaestimat/ Betastandardavvik
gen precision_beta = 1/Betastandardavvik
gen t_gamma = Gammaestimat/Gammastandardavvik
gen precision_gamma = 1/Gammastandardavvik

scatter precision_beta Betaestimat, ytitle(precision_beta) xtitle(beta) msize(small) ///
msymbol(diamond) plotregion(style(none)) graphregion(fcolor(white)) mfcolor(gs12) mlcolor(gs0) ///
mlwidth(medium) saving(graph_funnel_beta, replace)

scatter Betaestimat AveYear, ytitle(beta) xtitle(AveYear) msize(medsmall) msymbol(diamond) ///
plotregion(style(none)) graphregion(fcolor(white)) mfcolor(gs12) mlcolor(gs0) mlwidth(medium) ///
saving(graph_chrono_beta, replace)

```
eststo clear
eststo: reg t_beta precision_beta, robust cluster(Studie)
xtset Studie
eststo: xtreg t_beta precision_beta, fe vce(cluster Studie)
eststo: reg t_beta Betastandardavvik precision_beta
eststo: reg t_beta precision_beta if Studie != 26, robust cluster(Studie)
esttab using PublicationSelection_MRAs_beta.rtf, label ///
title(Selection publication bias simple MRAs - beta) ///
mtitles("WLS-MRA" "FEML" "PEESE WLS-MRA" "WLS-MRA reduced sample") ///
addnote("Say that all use cluster robust SEs.") se replace
eststo clear

scatter precision_gamma Gammaestimat, ytitle(precision_gamma) xtitle(gamma) xline(0) ///
msize(small) msymbol(diamond) plotregion(style(none)) graphregion(fcolor(white)) mfcolor(gs12) ///
mlcolor(gs0) mlwidth(medium) saving(graph_funnel_gamma, replace)

scatter Gammaestimat AveYear, ytitle(gamma) xtitle(AveYear) msize(medsmall) ///
msymbol(diamond) plotregion(style(none)) graphregion(fcolor(white)) mfcolor(gs12) ///
mlcolor(gs0) mlwidth(medium) saving(graph_chrono_gamma, replace)

eststo clear
eststo: reg t_gamma precision_gamma, robust cluster(Studie)
xtset Studie
eststo: xtreg t_gamma precision_gamma, fe vce(robust)
eststo: reg t_gamma Gammastandardavvik precision_gamma
eststo: reg t_gamma precision_gamma if Studie != 26, robust cluster(Studie)
esttab using PublicationSelection_MRAs_gamma.rtf, label ///
title(Selection publication bias simple MRAs - gamma) ///
nonumbers mtitles("WLS_MRA" "FEML" "PEESE WLS_MRA" "WLS-MRA reduced sample") ///
addnote("Can add note here") replace
eststo clear
```

// end of Publication Selection Section

// Multiple MRA - Heterogeneity

// Cochran's Q-test - is there evidence of excess heterogeneity

```
reg t precision, noconst
reg t_beta precision_beta, noconst
reg t_gamma precision_gamma, noconst
```

// Method MRAs

```
gen IVSE = IV/SEk
gen PermProxSE = PermProx/Sek
gen IVSE_gamma = IV/Gammastandardavvik
gen PermProxSE_gamma = PermProx/Gammastandardavvik
gen IVSE_beta = IV/Betastandardavvik
gen PermProxSE_beta = PermProx/Betastandardavvik

xtset Studie
eststo clear
eststo: qui reg t IVSE PermProxSE precision, robust cluster(Studie)
eststo: qui xtreg t IVSE PermProxSE precision, fe vce(cluster Studie)
```

```
eststo: qui reg t_gamma IVSE_gamma PermProxSE_gamma precision_gamma, robust cluster(Studie)
eststo: qui reg t_beta IVSE_beta PermProxSE_beta precision_beta, robust cluster(Studie)

esttab using MethodMRAs.rtf, label ///
title(Method MRAs, all with cluster robust SE) ///
mtitles("k" "k FEML" "gamma" "beta") ///
addnote("Can add note here") se replace
eststo clear
```

// Method MRAs – specific instruments

```
gen EducationSE = Education/SEk
gen HousingSE = Housing/SEk
gen CapitalincomeSE = Capitalincome/SEk
gen EducationSE_gamma = Education/Gammastandardavvik
gen HousingSE_gamma = Housing/Gammastandardavvik
gen CapitalincomeSE_gamma = Capitalincome/Gammastandardavvik
gen EducationSE_beta = Education/Betastandardavvik
gen HousingSE_beta = Housing/Betastandardavvik
gen CapitalincomeSE_beta = Capitalincome/Betastandardavvik

xtset Studie
eststo clear
eststo: qui reg t EducationSE HousingSE CapitalincomeSE PermProxSE ///
precision, robust cluster(Studie)
eststo: qui xtreg t EducationSE HousingSE CapitalincomeSE PermProxSE ///
precision, fe vce(cluster Studie)
eststo: qui reg t_gamma EducationSE_gamma HousingSE_gamma CapitalincomeSE_gamma ///
PermProxSE_gamma precision_gamma, robust cluster(Studie)
eststo: qui reg t_beta EducationSE_beta HousingSE_beta CapitalincomeSE_beta ///
PermProxSE_beta precision_beta, robust cluster(Studie)

esttab using MethodMRAs_instruments.rtf, label ///
title(Method MRAs - instruments) ///
mtitles("k" "k FEML" "gamma" "beta") ///
addnote("Can add note here") se replace
eststo clear
```

//End Method MRAs

// Macro MRAs

```
gen CorruptionindexscoreSE = Corruptionindexscore/SEk
gen TaxsystemleveltaxrevenueSE = Taxsystemleveltaxrevenue/SEk
gen GDPpercapitaSE = GDPpercapita/SEk
gen TaxSystemQualitySE = Payingtaxeskvalitetpåskatte/SEk

gen CorruptionindexscoreSE_gamma = Corruptionindexscore/Gammastandardavvik
gen TaxsystemleveltaxrevenueSE_gamma = Taxsystemleveltaxrevenue/Gammastandardavvik
gen GDPpercapitaSE_gamma = GDPpercapita/Gammastandardavvik
gen TaxSystemQualitySE_gamma = Payingtaxeskvalitetpåskatte/Gammastandardavvik

gen CorruptionindexscoreSE_beta = Corruptionindexscore/Betastandardavvik
gen TaxsystemleveltaxrevenueSE_beta = Taxsystemleveltaxrevenue/Betastandardavvik
gen GDPpercapitaSE_beta = GDPpercapita/Betastandardavvik
gen TaxSystemQualitySE_beta = Payingtaxeskvalitetpåskatte/Betastandardavvik
```

```
eststo clear
eststo: qui reg t CorruptionindexscoreSE TaxsystemleveltaxrevenueSE GDPpercapitaSE ///
TaxSystemQualitySE precision, robust cluster(Studie)
eststo: qui reg Gamma_t CorruptionindexscoreSE_gamma TaxsystemleveltaxrevenueSE_gamma ///
GDPpercapitaSE_gamma TaxSystemQualitySE_gamma precision_gamma, robust cluster(Studie)
eststo: qui reg Beta_t CorruptionindexscoreSE_beta TaxsystemleveltaxrevenueSE_beta ///
GDPpercapitaSE_beta TaxSystemQualitySE_beta precision_beta, robust cluster(Studie)

esttab using MacroMRAs.rtf, label ///
title(Macro MRAs, all with cluster robust SE) ///
nonumbers mtitles("k" "gamma" "beta") ///
addnote("Can add note here") se replace
eststo clear
```

//End macro MRAs

//Definition of variables MRAs

```
gen HeadSE = Head/SEk
gen EmployeesSE = Employees/SEk
gen PublicEmployeesSE = PublicEmployees/SEk
gen ShareSE = Share/SEk
gen ShareStatusSE = ShareStatus/SEk
gen TotExpSE = TotExp/SEk
gen OtherExpSE = OtherExp/SEk

gen HeadSE_gamma = Head/Gammastandardavvik
gen EmployeesSE_gamma = Employees/Gammastandardavvik
gen PublicEmployeesSE_gamma = PublicEmployees/Gammastandardavvik
gen ShareSE_gamma = Share/Gammastandardavvik
gen ShareStatusSE_gamma = ShareStatus/Gammastandardavvik
gen TotExpSE_gamma = TotExp/Gammastandardavvik
gen OtherExpSE_gamma = OtherExp/Gammastandardavvik

gen HeadSE_beta = Head/Betastandardavvik
gen EmployeesSE_beta = Employees/Betastandardavvik
gen PublicEmployeesSE_beta = PublicEmployees/Betastandardavvik
gen ShareSE_beta = Share/Betastandardavvik
gen ShareStatusSE_beta = ShareStatus/Betastandardavvik
gen TotExpSE_beta = TotExp/Betastandardavvik
gen OtherExpSE_beta = OtherExp/Betastandardavvik

xtset Studie
eststo clear
eststo: qui reg t HeadSE EmployeesSE PublicEmployeesSE ShareSE ShareStatusSE TotExpSE ///
OtherExpSE precision, robust cluster(Studie)
eststo: qui xtreg t HeadSE EmployeesSE PublicEmployeesSE ShareSE ShareStatusSE TotExpSE ///
OtherExpSE precision, fe vce(cluster Studie)
eststo: qui reg t_gamma HeadSE_gamma EmployeesSE_gamma PublicEmployeesSE_gamma ///
ShareSE_gamma ShareStatusSE_gamma TotExpSE_gamma OtherExpSE_gamma ///
precision_gamma, robust cluster(Studie)
eststo: qui reg t_beta HeadSE_beta EmployeesSE_beta PublicEmployeesSE_beta ///
ShareSE_beta ShareStatusSE_beta TotExpSE_beta OtherExpSE_beta ///
precision_beta, robust cluster(Studie)
```

```
esttab using DefinitionMRAs.rtf, label ///
title(Method MRAs, all with cluster robust SE) ///
mtitles("k" "k FEML" "gamma" "beta") ///
addnote("Say: all with cluster robust SEs") se replace
eststo clear
```

//General-to-Specific

```
gen PublishedSE = Published/SEk
gen AveYearSE = AveYearbaseyear2000/SEk
gen PanelSE = Panel/SEk

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
GDPpercapitaSE TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ///
ShareSE ShareStatusSE TotExpSE OtherExpSE PanelSE PublishedSE AveYearSE ///
precision IV PermProx Head Employees PublicEmployees Panel

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ShareSE ///
ShareStatusSE TotExpSE OtherExpSE PanelSE PublishedSE AveYearSE precision ///
IV PermProx Head Employees PublicEmployees Panel

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ShareSE ///
ShareStatusSE TotExpSE OtherExpSE PanelSE PublishedSE AveYearSE precision ///
IV PermProx Head Employees PublicEmployees

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ShareSE ///
ShareStatusSE TotExpSE OtherExpSE PublishedSE AveYearSE precision IV ///
PermProx Head Employees PublicEmployees

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ShareStatusSE ///
TotExpSE OtherExpSE PublishedSE AveYearSE precision IV PermProx Head ///
Employees PublicEmployees

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE ShareStatusSE TotExpSE OtherExpSE ///
PublishedSE AveYearSE precision IV PermProx Head Employees PublicEmployees

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE TotExpSE OtherExpSE PublishedSE ///
AveYearSE precision IV PermProx Head Employees PublicEmployees

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE TotExpSE PublishedSE AveYearSE ///
precision IV PermProx Head Employees PublicEmployees

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE PublishedSE AveYearSE precision ///
IV PermProx Head Employees PublicEmployees
```

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx ///
Head Employees PublicEmployees

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx ///
Head PublicEmployees

//Specific model

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx ///
Head PublicEmployees
eststo reg1
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns, without precision (the intercept)

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE ///
HeadSE EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx ///
Head PublicEmployees, robust cluster(Studie)
eststo reg2
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns, without precision (the intercept)

//remove variables that are assumed to not vary within studies

xtset Studie
xtreg t IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE precision ///
IV PermProx Head PublicEmployees, fe vce(cluster Studie)
eststo reg3
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE // test of the presence of genuine, systematic patterns, without precision (the intercept)

esttab reg1 reg2 reg3 using MultipleMRA_specific_model.rtf, label ///
title(Multiple MRAs - specific model) se replace
eststo clear

//Robustness - specific model on reduced sample 1 og 2

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE EmployeesSE ///
PublicEmployeesSE PublishedSE precision IV PermProx Head ///
PublicEmployees if Studie !=26, robust cluster(Studie)
eststo reg1
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE EmployeesSE ///
PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE EmployeesSE ///
PublicEmployeesSE PublishedSE precision IV PermProx Head ///

```
PublicEmployees if Studie !=26 & Studie !=19, robust cluster(Studie)
eststo reg3
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE EmployeesSE ///
PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns among
underreporting research, without precision (the intercept)


/*remove variables that are assumed to not vary within studies*/

xtset Studie
xtreg t IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE precision IV PermProx Head ///
PublicEmployees if Studie !=26, fe vce(cluster Studie)
eststo reg2
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE // test of the presence of genuine,
systematic patterns among underreporting research, without precision (the intercept)

xtreg t IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE precision IV PermProx Head ///
PublicEmployees if Studie !=26 & Studie !=19, fe vce(cluster Studie)
eststo reg4
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE // test of the presence of genuine,
systematic patterns among underreporting research, without precision (the intercept)

esttab  reg1 reg2 reg3 reg4 using MultipleMRA_specific_model_reduced_sample.rtf, label ///
title(Multiple MRAs - specific model reduced sample) se replace
eststo clear
```

//Robustness - gamma og beta

```
reg t_gamma IVSE_gamma PermProxSE_gamma TaxsystemleveltaxrevenueSE_gamma ///
TaxSystemQualitySE_gamma HeadSE_gamma EmployeesSE_gamma PublicEmployeesSE_gamma ///
PublishedSE_gamma precision_gamma IV PermProx Head PublicEmployees, robust cluster(Studie)
eststo reg1
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE_gamma PermProxSE_gamma TaxsystemleveltaxrevenueSE_gamma ///
TaxSystemQualitySE_gamma HeadSE_gamma EmployeesSE_gamma PublicEmployeesSE_gamma ///
PublishedSE_gamma // test of the presence of genuine, systematic patterns among underreporting
research, without precision (the intercept)

reg t_beta IVSE_beta PermProxSE_beta TaxsystemleveltaxrevenueSE_beta ///
TaxSystemQualitySE_beta HeadSE_beta EmployeesSE_beta PublicEmployeesSE_beta ///
PublishedSE_beta precision_beta IV PermProx Head PublicEmployees, robust cluster(Studie)
eststo reg3
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE_beta PermProxSE_beta TaxsystemleveltaxrevenueSE_beta TaxSystemQualitySE_beta ///
HeadSE_beta EmployeesSE_beta PublicEmployeesSE_beta PublishedSE_beta // test of the presence
of genuine, systematic patterns among underreporting research, without precision (the intercept)

xtset Studie
xtreg t_gamma IVSE_gamma PermProxSE_gamma HeadSE_gamma EmployeesSE_gamma ///
PublicEmployeesSE_gamma precision_gamma IV PermProx Head ///
PublicEmployees, fe vce(cluster Studie)
```

eststo reg2
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE_gamma PermProxSE_gamma HeadSE_gamma EmployeesSE_gamma ///
PublicEmployeesSE_gamma // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

xtreg t_beta IVSE_beta PermProxSE_beta HeadSE_beta EmployeesSE_beta ///
PublicEmployeesSE_beta precision_beta IV PermProx Head PublicEmployees, fe vce(cluster Studie)
eststo reg4
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE_beta PermProxSE_beta HeadSE_beta EmployeesSE_beta PublicEmployeesSE_beta // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

esttab reg1 reg2 reg3 reg4 using Specific_model_FEML_WLS_gamma_beta_robusthet.rtf, label ///
title(Multiple MRAs - specific model_gamma_beta_robusthet) se replace
eststo clear

//Robusthet - specific model without Obwona

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx Head ///
PublicEmployees if Studie !=3
eststo reg1
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

reg t IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE precision IV PermProx Head ///
PublicEmployees if Studie !=3, robust cluster(Studie)
eststo reg2
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE EmployeesSE ///
PublicEmployeesSE PublishedSE // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

/*remove variables that are assumed to not vary within studies*/

xtset Studie
xtreg t IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE precision IV ///
PermProx Head PublicEmployees if Studie !=3, fe vce(cluster Studie)
eststo reg3
test IV PermProx Head PublicEmployees _cons //test of publication selection bias
test IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE // test of the presence of genuine, systematic patterns among underreporting research, without precision (the intercept)

esttab reg1 reg2 reg3 using MultipleMRA_specific_model_wo_Obwona2.rtf, label ///
title(Multiple MRAs - specific model wo Obwona2) se replace
eststo clear

//Robustness method (table 10) - without Obwona
xtset Studie

```
eststo clear
eststo: qui reg t IVSE PermProxSE precision if Studie !=3, robust cluster(Studie)
eststo: qui xtreg t IVSE PermProxSE precision if Studie !=3, fe vce(cluster Studie)
eststo: qui reg t_gamma IVSE_gamma PermProxSE_gamma ///
precision_gamma if Studie !=3, robust cluster(Studie)
eststo: qui reg t_beta IVSE_beta PermProxSE_beta precision_beta if Studie !=3, robust cluster(Studie)

esttab using MethodMRAs_obwona.rtf, label ///
title(Method MRAs obwona, all with cluster robust SE) ///
mtitles("k" "k FEML" "gamma" "beta") ///
addnote("Say: all with cluster robust SEs") se replace
eststo clear
```

//General model for appendix

```
eststo clear
reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
GDPpercapitaSE TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ///
ShareSE ShareStatusSE TotExpSE OtherExpSE PublishedSE AveYearSE precision ///
IV PermProx Head Employees PublicEmployees
eststo reg1

reg t IVSE PermProxSE CorruptionindexscoreSE TaxsystemleveltaxrevenueSE ///
GDPpercapitaSE TaxSystemQualitySE HeadSE EmployeesSE PublicEmployeesSE ///
ShareSE ShareStatusSE TotExpSE OtherExpSE PublishedSE AveYearSE precision ///
IV PermProx Head Employees PublicEmployees, robust cluster(Studie)
eststo reg2

//remove variables that are assumed to not vary within studies

xtset Studie
xtreg t IVSE PermProxSE HeadSE EmployeesSE PublicEmployeesSE precision IV ///
PermProx Head Employees PublicEmployees, fe vce(cluster Studie)
eststo reg3

esttab  reg1 reg2 reg3 using MultipleMRA_general_model.rtf, label ///
title(Multiple MRAs - general model) se replace
eststo clear
```

//End of heterogeneity section

//Calculating corrected effect

```
foreach var of varlist IV - Payingtaxeskvalitetpåskatte {

    egen `var'_mean = mean(`var')

}

reg t precision IVSE PermProxSE TaxsystemleveltaxrevenueSE TaxSystemQualitySE HeadSE ///
EmployeesSE PublicEmployeesSE PublishedSE IV PermProx Head PublicEmployees

display _b[precision] + IV_mean*_b[IVSE] + ///
PermProx_mean*_b[PermProxSE] +
Taxsystemleveltaxrevenue_mean*_b[TaxsystemleveltaxrevenueSE] + ///
Payingtaxeskvalitetpåskatte_mean*_b[TaxSystemQualitySE] + ///
```

```
Head_mean*_b[HeadSE] + Employees_mean*_b[EmployeesSE] + ///
PublicEmployees_mean*_b[PublicEmployeesSE] + Published_mean*_b[PublishedSE]


                    // =1.0906
```

References

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance, 19*(1), 54-77. Doi: 10.1007/s10797-011-9171-2

Anwar, S., Akbar, R., Akbar, M. W., & Azhar, A. (2017). Measuring the Size of Under Ground Economy in Pakistan: A Microeconomic Approach. *Journal of Applied Environmental and Biological Sciences*, 7(8), 84-93. Retrieved from https://www.researchgate.net/profile/Sofia_Anwar/publication/319039876_Measuring_the_Size_of_Under_Ground_Economy_in_Pakistan_A_Microeconomic_Approach/links/598c530ca6fdcc58acb89f6e/Measuring-the-Size-of-Under-Ground-Economy-in-Pakistan-A-Microeconomic-Approach.pdf

Apel, M. (1994). *An expenditure-based estimate of tax evasion in Sweden* (PHD). Uppsala-Working Paper Series.

Baker, P. (1993). Taxpayer compliance of the self-employed: estimates from household spending data. *IFS Working Paper Series, 14*. Doi: 10.1920/wp.ifs.1993.9314

Bernotaite, R., & Piskunova, A. (2005). An expenditure-based estimate of Latvia's shadow economy. *SSE Riga Working Papers*, (75). Retrieved 5.september 2017 from http://www.sseriga.edu/en/research/student-research/page:12/

Besim, M., & Jenkins, G. P. (2005). Tax compliance: when do employees behave like the self-employed?. *Applied Economics, 37*(10), 1201-1208. Doi: 10.1080/00036840500109407

Bovi, M. (2002). *The nature of the underground economy: some evidence from OECD countries*. Roma: ISAE, Institute for Studies and Economic Analyses. Retrieved 20.november 2017 from https://s3.amazonaws.com/academia.edu.documents/41544519/The_nature_of_the_underground_economy_So20160125-12285-1tuogge.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1511806363&Signature=EMxYLlXFlnHFjmjtb3Q3hHrKaTc%3D&response-content-disposition=inline%3B%20filename%3DThe_nature_of_the_underground_economy_So.pdf

Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: a meta-analysis. *The American Economic Review, 85*(2), 238-243. Retrieved 17.october 2017 from http://www.jstor.org/stable/2117925

Davutyan, N. (2008*)*. Estimating the size of Turkey's informal sector: an expenditure-based approach. *Journal of Economic Policy Reform, 11*(4), 261-271. Doi: 10.1080/17487870802598393

Doing Business. (2017a). Paying Taxes – Distance to Frontier. Retrieved 29. November 2017 from http://www.doingbusiness.org/data/exploretopics/paying-taxes/frontier

Doing Business. (2017b). Paying Taxes Methodology. Retrieved 29.November 2017 from http://www.doingbusiness.org/methodology/paying-taxes

Ekici, T., & Besim, M. (2016). A Measure of the Shadow Economy in a Small Economy: Evidence from Household-Level Expenditure Patterns. *Review of Income and Wealth, 62*(1), 145-160. Doi: 10.1111/roiw.12138

Engström, P., & Hagen, J. (2017). Income underreporting among the self-employed: a permanent income approach. *European Economic Review, 92*, 92-109. Doi: 10.1016/j.euroecorev.2016.12.001

Engström, P., & Holmlund, B. (2009). Tax evasion and self-employment in a high-tax country: evidence from Sweden. *Applied Economics, 41*(19), 2419-2430. Doi: 10.1080/00036840701765452

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557. Doi: 10.1136/bmj.327.7414.557

Hoaglin, D. C. (2016). Misunderstandings about Q and 'Cochran's Q test'in meta-analysis. *Statistics in medicine, 35*(4), 485-495. Doi: 10.1002/sim.6632

Hurst, E., Li, G., & Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics, 96*(1), 19-33. Doi: 10.1162/REST_a_00363

Johansson, E. (2005). An estimate of self-employment income underreporting in Finland. *Nordic Journal of Political Economy, 31*(1), 99-109. Retrieved 5.september 2017 from https://www.researchgate.net/profile/Edvard_Johansson2/publication/23954607_An_Estimate_of_Self-Employment_Income_Underreporting_in_Finland/links/004635224df4dc8b08000000.pdf

Kapociute, M. (2013). *Do Entrepreneurs Under-report Their Income?* (Master's thesis). Erasmus University, Rotterdam.

Kim, B., Gibson, J., & Chung, C. (2017). Using Panel Data to Estimate Income Under-Reporting by the Self-Employed. *The Manchester School, 85*(1), 41-64. Doi: 10.1111/manc.12135

Kukk, M., & Staehr, K. (2014). Income underreporting by households with business income: evidence from Estonia. *Post-Communist Economies, 26*(2), 257-276. Doi: 10.1080/14631377.2014.904110

Kukk, M., & Staehr, K. (2017). Identification of Households Prone to Income Underreporting: Employment Status or Reported Business Income?. *Public Finance Review, 45*(5), 599-627. Doi: 10.1177/1091142115616182

Lyssiotou, P., Pashardes, P., & Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal, 114*(497), 622-640. Doi: 10.1111/j.1468-0297.2004.00234.x

Martinez-Lopez, D. (2013). The underreporting of income by self-employed workers in Spain. *SERIEs*, *4*(4), 353-371. Doi: 10.1007/s13209-012-0093-8

Mirus, R., & Smith, R. S. (1997). *Self-employment, tax evasion, and the underground economy: Micro-based estimates for Canada*. Working paper series, International Tax Program, Harvard Law School.

NIST/SEMATECH (2013, 30. October) *e-Handbook of Statistical Methods*. Retrieved 25. October 2017 from http://www.itl.nist.gov/div898/handbook/toolaids/pff/mpc.pdf

Nygard, O. E., Slemrod, J. B., & Thoresen, T. O. (2016). *Distributional Implications of Joint Tax Evasion*. CESifo Working Paper Series No. 5915. Retrieved from https://ssrn.com/abstract=2798601

Obwona, M. (1999). *Estimating unreported income of the self-employed and tax evasion in Uganda: An expenditure-based approach* (No. 9). Economic Policy Research Center.

OECD. (2017a). *Shining Light on the Shadow Economy: Opportunities and Threats.* (OECD report). Retrieved 16. October 2017 from http://www.oecd.org/tax/crime/shining-light-on-the-shadow-economy-opportunities-and-threats.pdf

OECD. (2017b). Tax revenue (indicator). Retrieved 10. November 2017 from https://data.oecd.org/tax/tax-revenue.htm

Parvathi, P., & Nguyen, T. T. (In press, 2018). Is Environmental Income Reporting Evasive in Household Surveys? Evidence From Rural Poor in Laos. *Ecological Economics*, 143, 218-226.

Paulus, A. (2015). Income underreporting based on income expenditure gaps: Survey vs tax records. *ISER Working Paper Series,* University of Essex. Retrieved 5. September 2017 from https://www.econstor.eu/bitstream/10419/126467/1/828388474.pdf

Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics, 39*(1), 17-32. Doi:10.1016/0047-2727(89)90052-2

Schneider, F., & Buehn, A. (In press, 2018). Shadow Economy: Estimation Methods, Problems, Results and Open Questions. *Open Economics. 1*(1), 1-29. Doi: 10.1515/openec-2017-0001

Schuetze, H. J. (2002). Profiles of tax non-compliance among the self-employed in Canada: 1969 to 1992. *Canadian Public Policy/Analyse de Politiques, 28*(2), 219-238. Doi: 10.2307/3552326

Skjeggestad, S. O., & Wæhle, S. (2015). *Underrapportering av inntekt blant selvstendig næringsdrivende-Et utgiftsbasert estimat.* (Master's thesis). The Norwegian School of Economics, Bergen.

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. Oxford: Routledge.

Stanley, T. D., & Doucouliagos, H. (2014). Meta‐regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60-78. Doi: 10.1002/jrsm.1095

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., ... & Rost, K. (2013). Meta‐analysis of economics research reporting guidelines. *Journal of Economic Surveys, 27*(2), 390-394. Doi: 10.1111/joes.12008

Stanley, T. D., Doucouliagos, C., & Jarrell, S. B. (2008). Meta-regression analysis as the socio-economics of economics research. *The Journal of Socio-Economics, 37*(1), 276-292. Doi: 10.1016/j.socec.2006.12.030

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). *Modelling publication bias in meta-analysis: a review*. Statistical methods in medical research, 9(5), 421-445. Doi: 10.1177/096228020000900503

Tanzi, V. (1983). The underground economy in the United States: annual estimates, 1930-80. Staff Papers, 30(2), 283-305. Doi: 10.2307/3867001

The World Bank. (2017a). Tax Revenue (% of GDP). Retrieved 10. November 2017 from https://data.worldbank.org/indicator/GC.TAX.TOTL.GD.ZS?locations=GB-SE-CA-CY-TR-KR-AU-US

The World Bank. (2017b). GDP per capita (current US$). Retrieved 29. November 2017 from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Torero, M., Robles, M., Hernández, M., De la Roca, J., Webber, M., & Thomas, D. (2006). *The Informal Sector in Jamaica* (Inter-American Development Bank). Retrieved 5.september 2017 from https://publications.iadb.org/bitstream/handle/11319/4326/The%20Informal%20Sector%20in%20Jamaica.pdf?sequence=1&isAllowed=y

Torosyan, K., & Filer, R. K. (2014). Tax reform in Georgia and the size of the shadow economy. *Economics of Transition*, 22(1), 179-210. Doi: 10.1111/ecot.12034

Torregrosa-Hetland, S. (2016). Sticky income inequality in the Spanish transition (1973-1990). *Revista de Historia Economica-Journal of Iberian and Latin American Economic History*, 34(1), 39-80. Doi: doi:10.1017/S0212610915000208

Transparency International. (2017). Corruption Perception Index. Retrieved 10. November 2017 from https://www.transparency.org/research/cpi

Wangen, K. R. (2004). *How extensive is the self-employed's tax evasion?* (Forskning om skatteøkonomi:71). Oslo: Statistisk Sentralbyrå.

Åstebro, T., & Chen, J. (2014). *The entrepreneurial earnings puzzle: Mismeasurement or real?*. Journal of Business Venturing, 29(1), 88-105. Doi: 10.1016/j.jbusvent.2013.04.003