Assessing quality of income data in a survey vs an administrative source

Dmitri Romanov[[1]](#footnote-1)\* and Yury Gubman[[2]](#footnote-2)\*\*

Abstract

Research assesses the quality of labour income reported in a social survey against indivuduals’ records in the administrative income tax file. Income data are known to be misreported in both sources, for different reasons, giving rise to substantial descrepancies at micro and macro levels of analysis. Income reportage in surveys is also plagued by high rates of item non-response. We focus on three novel issues: analyzing the distribution of discrepancies; investigating factors associated with positive and negative discrepancies, separately for the wage-earners and the self-employed individuals; utilizing the survey software Blaise Auditrail feature for pinpointing the effect of gros- and net-income items design on the quality of response. We find a clear regression to the mean: overstatement of labor income in the survey is more common among low-income respondents whereas understatement is more widespread among high-income respondents. We conclude that income mis-reportage by a wage-earner is determined by individual’s employment characteristics and fringe benefits. Misreporing by the self-employed, however, remains merely unexplained, supposedly due to the conceptual abiguity of income definition and measurement in the survey. The Auditrail data appear to make an invaluable contribution to the analysis of a respondent’s cognitive behavior, monitoring and improvement of data quality.

1. Introduction

Income is a crucial variable in profiling the socioeconomic status of an individual, a household, or a geographical region in statistical publications. One may obtain income data at the individual level from two sources: surveys and data from administrative sources such as tax authorities or social security authorities. Survey data are prone to errors of sampling, measurement, and non-response, and are costly to gather. Individual-level administrative data are susceptible to biased reportage due to the natural relationship between reporting income and paying taxes or receiving benefits. While such sources do provide “census”-type coverage of a specific type of income, they seldom provide information about income that is not taxable. Income data from both types of sources, administrative and survey, also differ in definitions, periods of reference, and populations covered (Yitzhaki, 2007).

Israel's Social Survey gathers information regarding labor income (from wage/salary and self-employment) by a question with ten predetermined income bands. Use of a banded question is supposed to raise the rate of item response in view of the known sensitivity of questions relating to income (Tourangeau and Smith, 1994). The band method is widely used for the reportage of income data in non-economic surveys. It is also employed in economic surveys that suffer from high non-response rates (Hurd et al., 2003). Reconstruction strategies by respondents who lack clear or documented knowledge about their income level often make even “accurate” reporting approximate and rounded (Moore et al., 2001), making banded questioning all the more justified (Czajka and Denmead, 2008).

According to the edifying review by Moore et al. (2001) "response bias estimates for wage/salary income are generally small and without a consistent sign" (p. 356). However, no systematic analysis has been conducted on the issue of reporting income from business and self-employed activity, which is more difficult in terms of conceptualization, definition, questioning, and reporting than wage/salary income (Martin et al., 1996).

Another issue that usually eludes the attention of researchers is the distribution of the income measurement errors. If, for example, errors in income reportage in a survey correlate negatively with “true” income (Bound and Krueger, 1991; Romanov and Furman, 2006) so that errors among the wealthy in one direction are offset by errors among the poor in the opposite direction, the average error may be negligible but the estimates of income inequality would be biased.

Accordingly, the current study has two main goals. First, it aims to compare banded labor income reported in a social survey with information from an administrative file, matched at the individual level, by distinguishing between negative and positive discrepancies and analyzing separately employees’ wage/salary income and business income of the self-employed. The second goal is to analyze factors that may explain the disparities between the two sources in individuals’ reported income. Here reference to three types of factors will be made: those related to question design and the cognitive process of response; those related to differences between employees and the self-employed in the nature and definition of labor income; and employment-related factors that make it hard to correctly reconstruct and report the level of labor income in the survey, even in a banded question.

The rest of the article is organized as follows: Part 2 presents the data. Part 3 shows the results of the estimation and Part 4 concludes.

2. Research Data and Variables

2.1 Database and Definition of Variables

The study is based on data from the 2008 Social Survey, performed each year in Israel by the Central Bureau of Statistics (see for details Social Survey methodology report, Central Bureau of Statistics, 2008 http://www.cbs.gov.il/www/skarim/social\_surv/metoda\_e.pdf).

In 2008, a systematic random sample of 8,899 individuals who belong to the survey population was extracted from the Population Registry; all were residents of Israel aged 20+. Among them, 7,327 responded (response rate of 82% of eligible persons). The survey is conducted by means of a computer-assisted personal interview in the sampled person’s home.

The income question refers to the respondent’s gross income from all sources (hereinafter: “gross income”). The question is administered only to individuals who reported working as employees or self-employed in the month preceding the survey.

The question for employees is: “Last month, what was your gross income, before deductions, from all places where you worked?” Self-employed respondents and those who manage their own companies are asked: “Last month, what was your gross income, before deductions, from all places where you worked, including wages and income from a business?” Of the 4,493 survey participants who were asked the question, 8.26% did not respond (refuse or do not know), lowering the number of respondents to the gross income item to 4,122. To answer the gross income question, the respondent is asked by the interviewer to avail him/herself of a card on which ten bands, specified in Table 1, are listed. The item non-response rate to the gross income question in Israel’s Social Survey is low by the standards of main U.S. government surveys, in which rates reach 25% or more (Moore et al., 2001; Czajka and Denmead, 2008; Groves et al., 2001).

The gross income question is immediately followed by one about net income: “What was your net income after deductions such as income tax, social security and mandatory health insurance contributions?” The question was added to the survey in 2007, in response to findings of Romanov and Furman (2006), as a tool with which respondents could control the distinction between gross income and net income by themselves. When respondents are asked about their net income after having answered the gross income question, if they mistakenly cited their net income in response to the first question, they are expected to realize their mistake and go back and correct the mistaken gross income variable.

In 2008, the survey participants were asked about their history in the labor market, the components of their wages, their fringe benefits, and their attitudes toward and expectations about their present and future status as employees/self-employed. We enrich the analysis with this information that general surveys usually lack.

The Social Survey is performed by means of Blaise, which, in a special application known as Audit Trail (AT), allows for the creation of a log file that records every motion the interviewer does during the interview. This log captures the following variables among others: time of entrance to fill in a field (a question), time of exit from field, value of variable upon entering field, and value entered by interviewer. If the respondent corrects a previous response, as many lines are recorded in the relevant field in the AT file as the number of times the interviewer repeated the specific question and corrected the value of the field. Basing ourselves on the AT file, we calculated three variables for each respondent who answered the gross income question: Did the respondent go back to the question? (0 if s/he answered once and 1 if s/he went back to the question and corrected h/her previous response); time (in seconds) that it took the respondent to answer the gross income question; and the size of the correction to gross income when the question was repeated, calculated as the difference between the first answer and the last one.

The survey records were matched with the administrative income tax data by a unique national ID number. The income tax data for employees itemize all jobs that the individual held during the tax year, annual wage/salary and the months in which s/he held them in each job. The average monthly income was calculated for all employee jobs by dividing the annual income by the number of months in the job. To match the Social Survey’s question of gross income from all jobs, employees’ income in the month preceding the enumeration month from the income tax file was calculated as the total average gross income from all jobs that the individual held in the pre-enumeration month. Since the self-employed do not report to the tax authorities working months, we calculated the gross monthly income of the self-employed by dividing the reported annual income by 12.

Individuals for whom records were not found in the income tax file or for whom the AT file contained no valid records (usually for technical reasons) were deleted from the database. The final database comprised records of 408 self-employed (12%) and 3,009 employees (88%). Table 1 shows the gross income values and the distribution of the responses obtained in the survey. All analyses in this study were performed on the basis of the final calibrated weights that CBS produced for the 2008 Social Survey.

Figure 1 contrasts the income reported in the survey with the same individuals’ income in the administrative file, grouped into the same bands as are specified in the survey. Figure 1 indicates that low-income respondents (bands 1, 2, and 3) over-reported their labor income in the survey relative to the administrative source and that those of relatively high income (bands 5 and above) under-reported it. This phenomenon—over-reporting of low income in the survey and under-reporting of high income—is known in statistical terms as regression to the mean.

Table 2 presents indices of the distribution of the specified variables on the basis of data in the aforementioned AT file. We see that 22% of the respondents went back to the question at least once and that most corrections were made in an upward direction, by about half a band on the scale of responses in the survey. By conjecture, they made this correction after realizing, when asked the net income question, that when responding to the previous question they had stated a net income value instead of a gross income value. Then, they returned to the gross income question and offered a higher value. By implication, adding a net income item to the questionnaire may improve the accuracy of reportage of gross income. It also allows those who do not know their gross income to respond. Notably, self-employed participants took, on average, much longer to respond than employees did (27.7 seconds as against 21.8, respectively) with a higher standard deviation (22.3 vs. 17.8).

2.2 Defining the measurement error

Most studies concerning measurement error treat income data from the administrative source as “true”. This approach assumes that, ideally, survey respondents would report their income exactly as it appears in the tax authorities’ files. However, Moore et al. (2001) disapprove of this approach, emphasizing:

Data from independent sources are almost never completely comparable to the survey data—due to sampling frame differences, timing differences, definitional differences, etc.—and the adjustments necessary to make them comparable are often inadequate. The flawed adjustments, and the fact that the independent estimates themselves are subject to various errors and omissions, add uncertainty to any comparison of survey and benchmark estimates (p. 333).

In contrast to the common practice, Abowd and Stinson (2011) assume as a point of departure that administrative data are not a “gold standard” due to errors in employers’ reportage to the tax authorities and the existence of black-market employment and income. Furthermore, any comparison of administrative data and survey reportage at the individual level entails record linkage, which in itself may bring about errors.

Accordingly, we treated discrepancies discovered in a comparison between the survey reportage and administrative-source data not as a “reporting error” by the individual but as a “measurement error” that is affected by the full range of factors.

Let *F* represent the distribution of income from the administrative source (as a continuous variable) across the population of survey respondents. For each band specified in the survey, an estimator of the expected income from the administrative file that would have been reported in the same band had it not been for measurement errors, according to distribution *F*, will be:

(1) 

—the predicted value of administrative income of individual *i* from all jobs, calculated on the basis of estimated distribution  in income band  as defined in the survey.

—the number of individuals in group k.

To find an estimator for distribution *F*, we fitted by maximum likelihood several accepted parametric distributions to the continuous administrative data of the survey respondents’ gross income. We examined log-normal and Gamma distributions (Banerjee et al., 2006), elliptic, asymmetric, skew-normal, and skew-t distributions (Azzalini and Capitanio, 2003). The best fit, i.e., the one that minimizes the deviations of the actual data from the value predicted on the basis of a theoretical parametric distribution, was obtained for the skew-t distribution. Figure 2 presents the distribution of the survey respondents’ administrative income and the estimated distribution line.

Having estimated distribution function *F* under the assumption , we may calculate the measurement error of individual *i* (in%):

(2) 

where I is an indicator that is assigned the value of 1 if the “true” income of individual *i*, as shown in the administrative file, does not belong to the band in which the individual reported h/her income in the survey; otherwise, it is 0.

The measurement error calculated by Formula (2) is positive when the income recorded in the administrative file is smaller than the expectation of the response in the survey (possibly over-reporting in the survey) and negative in the opposite case (possibly under-reporting in the survey).

Table 3 presents main indicators of the distribution of measurement error by respondent’s labor status. It may be seen that, on average, employees are more inclined than the self-employed to negative measurement errors (44.6% vs. 24%, respectively). The share of respondents who have no measurement error is higher among employees (40.4% vs. 33.8%, respectively) and the variance of measurement errors is wider among the self-employed (S.D. 97.7 vs. 67.3). By implication, there are meaningful differences between employees and the self-employed in the distribution of measurement errors.

3. Analysis of Factors for Measurement Error

3.1 Econometric Model

To identify the contribution of the factors that affect the measurement errors and estimate the intensity of their effect, we estimate the following regression (in matrix notation):

(3) 

e—measurement error;

*X*—set of individual’s background variables;

*L*—set of variables representing the individual’s employment and wage terms and variables of h/her subjective assessment of job mobility;

*AT*—set of Audit Trail variables, constructed in the manner described above;

—vector of residuals.

The estimation method was the weighted OLS and the final weights of the 2008 Social Survey were used.

* 1. Estimation Results

The factors associated with measurement error were analyzed separately for employees and for the self-employed due to a material difference in the regularity of these populations’ income and a conceptual difficulty in defining it among the self-employed. Also, a possibility of heterogeneous relations between the explained variable and variables representing the cognitive process of responding entails separate analysis of positive and negative measurement errors. Accordingly, for each group of respondents that was differentiated by labor status, we estimated Model (3) in three ways: in a sample of all observations, including a group of observations that had no measurement error; in a sample of observations that had a positive measurement error (possible over-reporting in the survey); and in a sample of observations that had a negative measurement error (possible under-reporting in the survey). Notably, the sign of the estimates depends on the sign of the error. For example, a positive estimate in the analysis of positive measurement errors is indicative of a factor that increases the error, whereas a positive estimate in an analysis of negative measurement errors signifies a factor that lowers the error (toward zero).

Table 4 presents the estimates for the full sample. First, the goodness of fit (adjusted R-square) is 0.05 in the model covering all observations and 0.05 and 0.49, respectively, in the separate models for negative and positive errors (this is partly due to the inclusion of the 40% of observations that had no measurement error as this term is defined in Formula (2); in these observations, the explained variable has no variation). Consequently, a hypothesis that all explanatory factors have the same effect on errors in both directions is not supported by the data. Therefore, below we relate to models of positive and negative error separately, irrespective of the pooled model.

The level of income obtained from the administrative file affects measurement error by lowering positive errors and increasing (in absolute terms) negative ones, with elasticity of roughly 0.2. Going back to the gross income question during the interview and correcting the response reduces negative errors, as we would expect if respondents offered a net income value in their first response instead of one relating to gross income. The estimate of the response time was not significant in the models for any direction of error. Other factors controlled for, men tended to exhibit 7% larger positive errors and 11% smaller negative errors than women. Age affected measurement errors in both directions in a U shape. An increase in level of education was identified with an increase in positive error but had no effect whatsoever on negative error.

Being an employee as opposed to being self-employed had a downward effect on errors in both directions—18% on positive errors and 31% on negative ones. This finding is consistent with the general conclusion in Moore et al. (2001) that reporting errors are negligible, on average, in the reportage of wage income. Individuals who reported being afraid of losing their jobs exhibited a 7% larger negative error. Holding an additional job on the survey date increased positive error at a 5% rate. Finally, satisfaction with income is identified with a 2% increase in positive errors and a 7% decrease in negative errors.

Table 5, like the pooled model (employees and the self-employed together), shows that the goodness of fit of the employees-only model is very strong for cases of positive error (0.57) and relatively poor for cases of negative error and all observations (0.09 and 0.07, respectively). The findings concerning the effects of the aforementioned demographic and socioeconomic variables on the full sample are also valid for employees. The employees-only model is especially mindful of employment characteristics that are likely to affect measurement errors. Thus, the number of jobs held increases positive error (by 5%), as do overtime work (by 3%) and fringe benefits such as full pay for sick days (by 3%), contributions to advance-training fund (by 6%), company car (by 8%), coverage of transportation expenses (by 5%), and profit-sharing (by 7%). In contrast, employees who reported not having been promoted in their current job tend to report their income with smaller positive errors by 11%. Other factors were found to have significant effects on negative reporting errors. First, part-time employment reduces negative error by 28%. For the given duration (full- or part-time) of individual’s main job, the more hours worked in all jobs, the smaller the negative error is. The negative error tends to be smaller among employees who reported having received a raise in recent years at their current job (by 11%) and those who benefited from employer contributions to pension (by 21%) and reimbursement of transportation expenses (by 11%). A much larger negative error (68%) is typical of employees who believe they have not been promoted on the job.

Table 6 presents estimates for the self-employed only. In this model, too, goodness of fit is relatively strong for positive errors (0.32) and weak for negative ones (0.02). In this model, unlike the employees-only model, the respondents’ demographic and socioeconomic indicators had no effect, with the sole exception of academic education at the baccalaureate level. The AT variables, too, were not found significant. The only factors that had statistically significant effects (with a disclaimer for the small number of observations) were subjective variables. Thus, satisfaction at work and the feeling that the respondent’s standard of living has fallen in recent years are identified with an increase in positive error by 7%. Two variables related to expectations—that business turnover will increase and that the business will make progress generally—had a downward effect of 10% on positive errors.

4. Summary and Conclusions

This study analyzed errors in the measurement of labor income on the basis of a comparison of reportage in a social survey, by means of a banded question, with records in an administrative file obtained from income tax authorities. Following Abowd and Stinson (2011), we did not consider the contents of the administrative file as true reportage against which the data reported in a survey should be examined and any difference should be regarded as a reporting error in the survey. We treated discrepancies between the administrative source and the survey as errors in the measurement of a latent variable. These measurement errors may be affected by a gamut of factors including differences in definitions, reporting periods, timing of the recording and receipt of income, non-response in the survey, under-reporting to the income tax authorities, reporting errors, errors of record linkage, and so on.

We focused on the relation between measurement error and three groups of factors: those associated with question design and the cognitive process of answering a question; those related to differences between employees and the self-employed in the nature and definition of labor income, and those associated with employment characteristics that make it difficult to correctly reconstruct and evaluate the level of labor income reported in the survey.

Our main finding is that regression to the mean occurs in the income reported in the survey as against that obtained from the administrative source. Positive measurement errors were more common among low-income respondents and negative measurement errors were more common among those of high income.

A significant negative monotonic relation was found between the income recorded in the administrative file and the measurement error for employees, different in intensity between positive and negative measurement errors (elasticity of 0.2 and 0.5, respectively). In other words, the higher the individual’s income is according to the administrative source, the smaller h/her measurement error in the survey will be if the error is positive, and the greater it will be (in absolute terms) if the error is negative.

According to our research hypothesis, high-income persons who report their income in a social survey tend to “forget” income that they received from additional jobs, overtime, self-employed income, fringe benefits, and nonrecurrent gains such as bonuses and profit-sharing. As a result, they tend to under-report their labor income in the survey relative to the administrative data. In contrast, low-income workers, who hold part-time and/or irregular jobs, tend in surveys to report the income they receive in a full month of work, a level that may be not representative of their average income. Consequently, one expects them to over-report their income relative to the administrative data. The findings of our research confirm these hypotheses and, by so doing, prove that the factors related to negative measurement error are different from those that are associated with a positive measurement error.

Parsing the analysis of measurement errors by the respondents’ labor status—employees vs. self-employed—we found that the two populations should not be pooled into one model due to material differences in the conceptual definition of income, how income is measured, volatility in income level during the year, and reporting on income in the survey as against reporting it to the tax authorities. Notably, among the self-employed, positive errors were almost twice as frequent as negative ones but the average error was close to zero, with greater variance than in errors among employees.

Analyzing the response process by means of Audit Trail variables, we found that inserting a question about net income immediately after inquiring about gross income, as a logical way to allow respondents to control their responses, caused 22% of the respondents to go back and check the accuracy of their responses to the gross income question. Reversion to the gross income question during the interview and correction of the answer reduced measurement errors among both employees and the self-employed but did so more among the former than among the latter.

In sum, the analysis of income measurement errors in a survey as against an administrative file, with no distinction made between employees and the self-employed, between high-income and low-income respondents, and among factors that are unique to each group, may gravely bias the findings.

References

Abowd, J.M., and Stinson, M.H. (2011). Estimating measurement error in SIPP annual job earnings: a comparison of Census Bureau survey and SSA administrative data. *US Census Bureau Center for Economic Studies Paper*, No. CES-WP 11-20.

Azzalini, A., and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society*, Series B, vol. 65, 367-389.

Banerjee, A., Yakovenko, V.M., and Di Matteo, T. (2006). A study of the personal income distribution in Australia. *Physica A*, vol. 370, 54-59.

Bound, J., and Krueger, A.B. (1991). The extent of measurement error in longitudinal earnings data: do the two wrongs make a right? *Journal of Labor Economics*, vol. 9 (1), 1-24.

Central Bureau of Statistics (2008). *Social Survey methodology report - The survey population, sample framework and estimation method, data collection*. Central Bureau of Statistics, Jerusalem, Israel (available from: http://www.cbs.gov.il/www/skarim/social\_surv/metoda\_e.pdf)

Czajka, J.L., and Denmead, G. (2008). Income data for policy analysis: a comparative assessment of eight surveys. *Mathematica Policy Research*, *Inc*.

Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (2001). *Survey nonresponse*. John Wiley and Sons, Inc.

Martin, J., Cheesebrough, S., Dodd, T., Farrant, G., and McKernan. A. (1996). Asking the self-employed about their income. *Survey Methodology Bulletin*, 39, 11-15.

Moore, J. C., Stinson, L. L., and Welniak J.E .J. (2000). Income measurement error in surveys: a review. *Journal of Official Statistics*, vol.16 (4), 331–361.

Romanov, D., and Furman, O. (2006). Analysis of wage data from the 1995 census by using wage file of the National Insurance Institute. *Central Bureau of Statistics Working Papers*, 21, Jerusalem (in Hebrew).

Tourangeau, R., and Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, vol. 60, 275-304.

Yitzhaki, S. (2007). Pros and cons for using administrative records in statistical bureaus. *UN Economic Commission for Europe*, Conference of European Statisticians, Geneva.

**Table 1: Distribution of Gross Income in the Social Survey**

|  |  |  |  |
| --- | --- | --- | --- |
| **Band** | **Income (NIS per month)** | **Respondents (N)** | **Pct. respondents in sample** |
| 1 | Up to 2,000 | 381 | 11.2 |
| 2 | 2001–3000 | 249 | 7.3 |
| 3 | 3001–4000 | 336 | 9.8 |
| 4 | 4001–5000 | 448 | 13.1 |
| 5 | 5001–6000 | 378 | 11.1 |
| 6 | 6001–7500 | 378 | 11.1 |
| 7 | 7501–10000 | 453 | 13.3 |
| 8 | 10001–14000 | 340 | 10.0 |
| 9 | 14001–21000 | 233 | 6.8 |
| 10 | 21000+ | 221 | 6.5 |
| Total |  | 3,417 | 100.0 |

**Table 2: Response to Gross Income Question—Audit Trail Variables**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Employees** | **Self-employed** |
| **Mean** | **S.D.** | **Mean** | **S.D.** |
| Time spent responding to item (seconds) | 21.8 | 17.6 | 27.7 | 22.3 |
| Frequency of going back to correct gross income after net income question | 0.23 | 0.42 | 0.22 | 0.42 |
| Size of correction to response among respondents who made such correction (number of income bands) | 0.50 | 1.69 | 0.49 | 2.27 |

**Table 3: Distribution of Measurement Errors**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Respondent group | **Pct. of negative errors** | **Pct.** **of no error** | **Pct. of positive errors** | **Mean error** | **S.D. of error** |
| Full sample | 42.1 | 39.6 | 18.3 | 19.5- | 71.9 |
| Employees | 44.6 | 40.4 | 15.0 | 21.8- | 67.3 |
| Self-employed  | 24.0 | 33.8 | 42.2 | 2.3- | 97.7 |

**Table 4: Estimation Results, Full Sample**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All observations** | **Positive error** | **Negative error** |
| Intercept | -8.21(16.96) | 151.15(10.40)\*\*\* | -36.46(48.91) |
| (Ln of) income from administrative file | -4.97(0.54)\*\*\* | -17.06(0.93)\*\*\* | -23.55(5.26)\*\*\* |
| Correction of response about gross income during interview | -6.82(3.43)\*\* | -1.19(1.70) | -19.31(6.88)\*\*\* |
| Size of response correction | 4.36(1.37)\*\*\* | 0.64(0.58) | 7.09(3.61)\*\* |
| Response time | 0.13(0.07)\*\* | 0.03(0.03) | 0.15(0.13) |
| Male | 10.92(2.64)\*\*\* | 7.35(1.49)\*\*\* | 10.80(5.50)\*\* |
| Age | 1.11(0.70) | 0.79(0.34)\*\* | 7.22(1.56)\*\*\* |
| Age squared | -0.01(0.01)\* | -0.01(0.003)\*\* | -0.08(0.02)\*\*\* |
| Married | -5.12(2.96)\* | -0.30(1.50) | 0.54(5.90) |
| Arab | -0.91(3.91) | -6.40(1.81)\*\*\* | -3.35(7.53) |
| Immigrant from FSU | 8.32(3.37)\*\* | -1.24(1.85) | 14.64(6.69)\*\* |
| Education—high school with matriculation | -6.30(3.44)\* | 2.73(1.74) | -10.05(6.80) |
| Academic education, B.A. | 0.68(3.41) | 5.66(1.86)\*\*\* | 6.29(6.76) |
| Academic education, M.A. | -0.41(4.25) | 10.59(2.30)\*\*\* | -6.50(8.85) |
| Academic education, Ph.D. | 6.87(11.14) | 20.00(6.71)\*\*\* | 18.57(26.72) |
| Status at work: employee | -14.88(3.96)\*\*\* | -17.54(1.61)\*\*\* | 31.32(9.49)\*\*\* |
| Number of jobs held | 10.42(3.43)\*\*\* | 4.71(1.41)\*\*\* | 1.72(8.02) |
| Fear of losing job | -2.76(1.70) | -0.53(0.88) | -7.47(3.31)\*\* |
| Satisfied with job | -3.30(1.78)\* | 1.62(0.92)\* | -3.50(3.59) |
| Satisfied with labor income | 7.70(1.64)\*\*\* | 2.46(0.86)\*\*\* | 7.48(3.38)\*\* |
| Household’s standard of living has risen | 6.78(2.67)\*\* | 1.69(1.41) | 7.10(5.18) |
| N | 3097 | 595 | 1286 |
| Adjusted R-square | 0.05 | 0.49 | 0.05 |

Significance level: \*\*\*<0.01, \*\*<0.05, \*<0.10.

Standard deviations are presented in brackets

**Table 5: Estimation Results, Employees Only**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All observations** | **Positive error** | **Negative error** |
| Intercept | -21.94(9.97)\*\*\* | 185.04(10.51)\*\*\* | 232.90(46.58)\*\*\* |
| (Ln of) income from administrative file | -6.20(0.75)\*\*\* | -21.03(1.21)\*\*\* | -46.48(6.02)\*\*\* |
| Correction of response about gross income during interview | -10.15(3.49)\*\*\* | -0.99(1.97) | -20.91(6.36)\*\*\* |
| Size of response correction | 2.36(0.87)\*\*\* | 0.63(0.49) | 5.75(1.58)\*\*\* |
| Male | 7.30(2.86)\*\* | 7.74(1.74)\*\*\* | 10.10(5.25)\* |
| Arab | -8.97(4.01)\*\* | -7.01(2.04)\*\*\* | -14.29(7.08)\*\* |
| Academic education, B.A. | 6.46(3.37)\* | 4.76(2.08)\*\* | 9.53(6.02) |
| Academic education, M.A. | 3.78(4.12) | 15.77(2.56)\*\*\* | -5.18(7.86) |
| Academic education, Ph.D. | 9.43(12.71) | 32.33(9.99)\*\*\* | 1.90(28.57) |
| Number of jobs held | 8.90(3.69)\* | 4.52(1.61)\*\*\* | 3.75(7.51) |
| Usual weekly work hours  | 5.62(1.17)\*\*\* | 0.18(0.60) | 11.35(2.24)\*\*\* |
| Satisfied with household’s economic situation | 7.49(1.66)\*\*\* | 1.50(0.91)\* | 8.75(3.02)\*\*\* |
| Part-time main job | 13.87(6.35)\*\* | -2.27(3.67) | 27.97(10.77)\*\*\* |
| Has worked more hours | 0.23(2.98) | 3.62(1.72)\*\* | -1.83(5.27) |
| Has worked fewer hours | 8.79(4.79)\* | -0.64(2.92) | 10.12(8.57) |
| Received a wage raise | 6.54(2.93)\*\* | -2.22(1.66) | 10.78(5.19)\*\* |
| Received a wage cut | -12.31(6.29)\*\* | -2.65(3.79) | -10.86(10.54) |
| Has not been promoted | -42.96(9.49)\*\*\* | -10.52(5.52)\* | -67.96(15.60)\*\*\* |
| Receives full pay for sick days | 0.15(2.91) | 3.40(1.64)\*\* | 2.76(5.21) |
| Employer participates in pension insurance | -2.68(3.78) | -2.42(1.79) | 21.05(7.54)\*\*\* |
| Employer participates in advanced-training fund  | -16.56(3.20)\*\*\* | 6.43(1.92)\*\*\* | -2.67(5.84) |
| Participates in profit-sharing with employer | -3.27(4.16) | 6.81(2.84)\*\* | -6.59(7.48) |
| Has company car | -0.35(3.88) | 7.62(2.48)\*\*\* | -0.27(7.56) |
| Receives reimbursement of transportation expenses | 2.96(3.47) | 5.42(2.32)\*\* | 11.10(6.22)\* |
| N | 2403 | 344 | 1107 |
| Adjusted R-square | 0.07 | 0.57 | 0.09 |

Significance level: \*\*\*<0.01, \*\*<0.05, \*<0.10.

Standard deviations are presented in brackets

**Table 6: Estimation Results, Self-Employed**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **All observations** | **Positive error** | **Negative error** |
| Intercept | -6.21(30.10) | 144.04(15.08)\*\*\* | -20.82(320.65) |
| (Ln of) income from administrative file | -4.20(1.82)\*\* | -13.62(1.71)\*\*\* | -12.03(33.66) |
| Academic education, B.A. | 16.87(13.62) | 8.12(3.67)\*\* | 59.93(49.80) |
| Satisfied with job | -7.75(7.93) | 7.29(2.19)\*\*\* | -32.07(26.82) |
| Satisfied with labor income | 23.78(7.80)\*\*\* | 1.53(2.20) | 50.25(26.99)\*\*\* |
| Household’s standard of living has fallen  | 10.39(16.63) | 8.51(4.74)\* | -17.61(53.64) |
| Expects business turnover to increase | -0.37(13.25) | -10.60(3.70)\*\*\* | 10.41(44.45) |
| Expects business to make progress | -5.97(11.37) | -8.97(3.04)\*\*\* | -8.00 (41.12) |
| N | 369 | 156 | 86 |
| Adjusted R-square | 0.02 | 0.34 | 0.02 |

Significance level: \*\*\*<0.01, \*\*<0.05, \*<0.10.

Standard deviations are presented in brackets

**Figure 1: Gross Labor Income Reportage in Survey vs. Administrative Data**



Thick line—income from administrative file by survey reportage bands. Thin line—income reported in survey. Broken line—continuous estimator of income reported in survey based on fitted theoretical distribution

**Figure 2: Fit of Parametric Distribution to Respondents’ Gross Labor Income Data**

Estimated parameters of skew-t distribution: location=698.71, scale=7278.73, shape=13.50, degrees of freedom=3.89.

**Appendix: Definition of Variables, Means, and Standard Deviations\***

|  |  |  |
| --- | --- | --- |
| **Variable** | **Employees** | **Self-employed** |
| **Mean** | **S.D.** | **Mean** | **S.D.** |
| Measurement error, pct. | -21.82 | 67.31 | -2.29 | 97.68 |
| Income from administrative file in month preceding survey, NIS | 9034.2 | 9670.2 | 9649.2 | 13900.9 |
| Correction of response about gross income during interview (Yes=1, No=0) | 0.23 | 0.42 | 0.22 | 0.42 |
| Response time on gross income question, seconds | 21.83 | 17.60 | 27.73 | 22.33 |
| Size of response correction, number of income bands as specified in survey | 0.50 | 1.70 | 0.49 | 2.27 |
| Male | 0.50 | 0.50 | 0.72 | 0.45 |
| Age | 39.87 | 12.42 | 46.95 | 13.21 |
| Married | 0.67 | 0.47 | 0.77 | 0.42 |
| Arab | 0.12 | 0.32 | 0.13 | 0.34 |
| Immigrant from FSU | 0.19 | 0.39 | 0.07 | 0.26 |
| Education—high school with matriculation | 0.21 | 0.41 | 0.15 | 0.36 |
| Academic education, B.A. | 0.19 | 0.39 | 0.18 | 0.38 |
| Academic education, M.A. | 0.11 | 0.31 | 0.12 | 0.33 |
| Academic education, Ph.D. | 0.01 | 0.10 | 0.02 | 0.14 |
| Number of jobs held  | 1.10 | 0.34 | 1.11 | 0.37 |
| Usual weekly work hours  | 42.04 | 13.88 | 45.03 | 18.62 |
| Fear of losing job, subjective evaluation on scale of 1 (no fear) to 4 (acute fear) | 1.49 | 0.74 | 1.63 | 0.76 |
| Satisfied with job, subjective evaluation on scale of 1 (totally dissatisfied) to 4 (very satisfied) | 3.09 | 0.77 | 3.19 | 0.79 |
| Satisfied with labor income, subjective evaluation on scale of 1 (totally dissatisfied) to 4 (very satisfied)  | 2.42 | 0.86 | 2.53 | 0.82 |
| Satisfied with household’s economic situation, subjective evaluation on scale of 1 (totally dissatisfied) to 4 (very satisfied) | 2.49 | 0.83 | 2.63 | 0.78 |
| Household’s standard of living has risen in recent years, subjective evaluation (Yes=1, No=0) | 0.53 | 0.50 | 0.54 | 0.50 |
| Household’s standard of living has fallen in recent years, subjective evaluation (Yes=1, No=0) | 0.16 | 0.37 | 0.16 | 0.36 |
| Variables specified for employees only, on basis of self-reportage (Yes=1, No=0) |
| Part-time main job  | 0.06 | 0.25 |  |  |
| Has worked more hours than usual recently | 0.27 | 0.44 |  |  |
| Has worked fewer hours than usual recently | 0.10 | 0.30 |  |  |
| Received a wage raise on current job | 0.59 | 0.49 |  |  |
| Received a wage cut on current job | 0.06 | 0.24 |  |  |
| Has not been promoted on current job | 0.02 | 0.14 |  |  |
| Receives full pay for sick days | 0.56 | 0.50 |  |  |
| Employer participates in pension insurance | 0.75 | 0.43 |  |  |
| Employer participates in advance-training fund | 0.46 | 0.50 |  |  |
| Participates in profit-sharing with employer | 0.11 | 0.31 |  |  |
| Has company car | 0.14 | 0.35 |  |  |
| Receives reimbursement of transportation expenses | 0.17 | 0.38 |  |  |
| Variables specified for self-employed only |
| Expects business turnover to increase |  |  | 0.32 | 0.47 |
| Expects business to make progress |  |  | 0.52 | 0.50 |

\* The explanatory variables included are those found to have had a significant effect in at least one regression model.

1. \* Israel Central Bureau of Statistics, corresponded author, email dromanov@cbs.gov.il. [↑](#footnote-ref-1)
2. \*\* Israel Central Bureau of Statistics and Hebrew University of Jerusalem, email yuryg@cbs.gov.il [↑](#footnote-ref-2)