Tax Evasion in the Housing Market: Identification and Exploration

Danny Ben-Shahar, Roni Golan, and Eyal Sulganik*

Abstract

The real estate market is recognized as a fertile ground for tax fraudsters. Specifically, reporting a price lower than the true transaction price in order to avoid tax payments is a prevalent fraud technique. We propose an empirical method for identifying housing transactions that are suspected of under-reporting. Based on all reported housing transactions in Israel over the period 1998–2015, we conclude that about 8% of the transactions are under-reported, with an average price report of 33% below the projected true price. Also, the likelihood to under-report is positively associated with the total tax liability and positively (negatively) associated with the crime rate in (the socioeconomic level of) the area in which the transaction occurs. Compared to single unit owners, real estate investors are less likely to engage in under-reporting. Our empirical approach may serve tax enforcement authorities in promoting tax collection in the real estate market.

Current Version: July 2019

Key Words: under-report, false price, housing, tax evasion

JEL Codes: H26, R30

*Danny Ben-Shahar, Coller School of Management, Tel Aviv University, Tel Aviv 6139001, Israel; email: bdanny@post.tau.ac.il; Roni Golan, Coller School of Management, Tel Aviv University, Tel Aviv 6139001, Israel; email: rgolan@tauex.tau.ac.il; and Eyal Sulganik, IDC, Herzliya, 4610101, Israel; email: sulganik@idc.ac.il. The authors thank Yovav Sanders for assistance in generating the dataset. We also thank Meir Mizrachi and Eli Farjon, the Israel Tax Authority, and participants at Tel Aviv University Law School seminar, 2018 Israel Regional Science Association meetings, and 2019 Israel Law and Economics Association meetings, for helpful comments. We thank the Survey of Israel for financial support.

1. INTRODUCTION

The real estate market is often recognized as a source of attraction to tax fraudsters [Schneider (2004), Nelen (2006), and Unger *et al.* (2010)]. Of a number of fraud techniques, misreporting the transaction price and, specifically, reporting a price lower than the true transaction price (hereafter, under-reporting) is most prevalent [Center for Tax Policy and Administration (2007)]. According to Sullivan (2015), for example, under-reporting is a simple means used for money laundering in which "the launder purchases the property via a bank loan for the deflated price and then makes payments using dirty money" (page 33). Also, The State of Israel Comptroller (2007) argues that of a sample of about 20,000 housing transactions examined by the comptroller, in about 17% of the cases the price reported to the Israel Tax Authority was at least 30% lower than the Authority's appraised assessment.[1]

Due to tax considerations, both sellers and buyers are often incentivized to under-report a transaction to the tax authorities. While the buyer's incentive is to save on purchase tax payments, the seller wishes to reduce payments associated with capital gains tax.[2] However, despite the common belief in the significant presence of under-reporting in the real estate market, to the best our knowledge, there exists no evidence-based assessment of the scope of this phenomenon. For example, in an extensive study

[1] See the State Comptroller and Ombudsman of Israel, 2007. Also, the legislator in Israel even found a need to comment on the misreporting phenomenon, stating that "A person who, in order to evade tax...has committed one of the offenses enumerated below, shall be liable to imprisonment for a term of seven years or a fine ... twice the amount of tax he has concealed or intended to conceal: (1) Stated in a declaration under this Act, a statement or a false record.... (3) Prepared or…permitted to prepare…a false statement or a false contract or other false records" [Section 98 (c2) of the Land Taxation Law (Appreciation and Acquisition)].

[2] Real estate purchase tax (referred to in some countries as stamp duty or transfer tax), which is paid in some places by the buyer and in other places by both the buyer and the seller at the time of home purchase and is often ad valorem, i.e., computed as a rate of the purchase price. Among the many places where this tax is prevalent in different forms are the majority of U.S. jurisdictions, Canada, the UK, Australia, and Israel (see, e.g., http://datatoolkits.lincolninst.edu/subcenters/significant-features-property-tax/Report_Real_Estate_Transfer_Charges.aspx. Lincoln Institute of Land Policy and George Washington Institute of Public Policy. Real Estate Transfer Charges; accessed: 7/13/2017). It is important to note that under-reporting, while enabling the buyer to save on current purchase tax, may increase capital gains tax for a future sale. However, as we will later describe, the seller is often granted an exemption from capital gains tax: over the period 2012–2016, of the total tax revenue from home sales, 87% was purchase tax, compared to only 13% generated from capital gains. Also, as average mortgage loan-to-value in Israel is traditionally equal to only 55–60%, there is limited incentive to over-report the transaction price for the purpose of increasing the loan amount. Finally, as there is no property tax in Israel, there is no future supervision of the reported price following the tax authority's oversight at the time of the transaction.

of the real estate tax fraud experiences of eighteen OECD countries, researchers were unable to uncover evidence about the pecuniary extent of tax fraud [with the exception of limited evidence on Austria—see Center for Tax Policy and Administration (2007)]. This absence of financial estimates is consistent with the fact that, to the best of our knowledge, the literature does not propose any statistical method by which transactions could be identified as suspicious of under-reporting.[3]

In this study we address this issue by proposing an empirical method for identifying transactions that are suspected to involve under-reporting. Moreover, by observing all housing transactions reported to the Israel Tax Authority over the period 1998–2015, we empirically implement our proposed identification method to assess the scope of under-reporting in the market. We conduct a series of identification and robustness tests to further validate our approach. Finally, we examine transaction characteristics that are associated with the likelihood to under-report.

Our method for identifying under-reports is based on the subset of repeat-sale transactions. In order for a transaction to be classified as an under-report, we require that the reported sale price (relative to concurrent quality-adjusted average market prices) is significantly lower than the reported sale price at the other (adjacent) time that the asset is traded (relative to the quality-adjusted price at that time), controlling for a series of other explanatory factors. We use simulation, empirical, and analytical solution approaches to support our identification method.[4]

Based on all reported housing transactions in Israel over the period 1998–2015 (for each of which we observe at least one repeat-sale), we find that, under conservative

---

[3] Among the accepted methods for identifying tax fraud in the real estate market are risk analysis and risk profiling, data matching, and data mining (Center for Tax Policy and Administration, 2007). Also, note that in certain cases, buyers engaged in a real estate transaction are also incentivized to over-price report for the purposes of money laundering or for attaining a greater mortgage loan from the lender. Other common tax schemes in the real estate market include the use of false identities (Center for Tax Policy and Administration, 2007). Nelen (2006) further explains why the real estate market has become fertile ground for tax fraud. Among other causes, he identifies the large sums that are involved in the transactions and the large transaction volume, which allows for the concealment of large sums of illegal money; the generally limited transparency and regulation and smaller investment risk, as compared to financial markets; the social status that accompanies real estate ownership; and the possible separation between legal and economic ownership of the asset.

[4] Leung, Leung, and Tsang (2014), Kopczuk and Munroe (2014), and Best and Kleven (2017) study the non-linear tax policies to show irregularities around the tax liability thresholds. However, they do not directly explore the scope of under-reporting and tax fraud in the market. Also see Benjamin, Coulson, and Yang (1993) for the effect of transfer tax on house prices.

requirements on model classification outcomes, about 8% of the transactions are under-reported, with an average price report being 33% below the true price. We further find that the likelihood to under-report is positively associated with the amount of tax liability involved in the transaction and positively (negatively) associated with the crime rate in (socioeconomic level of) the area in which the transaction occurs. Finally, following List (2003, 2004), Dhar and Zhu (2006), Venezia and Shapira (2007) and others who present evidence on the effect of experience and professionalism on individual economic behavior, our evidence indicates that, compared to single unit owners, real estate investors are less likely to engage in under-reporting.[5] Our outcomes are robust to various model specifications.

The contribution of our study is threefold. First, to the best of our knowledge, this is the first research to present a systematic, empirically based approach for indicating individual real estate transactions suspicious of under-reporting. Moreover, our proposed method allows us to assess the extent of tax evasion in the market and, further, to identify transaction and individual characteristics that associate with the likelihood to under-report. Finally, our proposed method may serve tax enforcement authorities in promoting tax collection in the real estate market.[6]

The plan of the paper is as follows. Section 2 provides background and describes the data, including variable definition and related summary statistics. Section 3 introduces the under-reporting identification method and presents a related assessment of the scope of under-reporting in the market. Section 4 presents an identification and robustness test of the under-report classification method. Section 5 estimates transaction and individual characteristics that are associated with the likelihood to under-report. Finally, Section 6 provides a summary and concluding remarks.

2. BACKGROUND AND SAMPLE

Due to tax considerations, the parties (seller and buyer) involved in a real estate transaction are typically incentivized to under-report a transaction to the tax authorities.

[5] Genesove and Mayer (1997) and Genesove and Mayer (2001) further show, for example, that seller owner-occupants differ from investors by the time on the market and loss aversion, respectively.

[6] In that regard, we should note that our proposed method for identifying false reports is suggestive rather than conclusive: While our approach indicates transactions for which the reported price is considerably inconsistent with previous transaction reporting of the same asset, in some cases the inconsistency may of course be due to factors such as within-family transactions and "fire-sales" rather than false reporting.

While the buyer's incentive is to save on purchase tax payments, the seller wishes to reduce payments associated with capital gains tax (if applies). Our empirical test is based on transaction reported to the Israel Tax Authority.[7] Israeli law requires that both parties report to the Israel Tax Authority upon transaction closing. The report includes the closing price as well as information on fundamental attributes of the transacted asset. Thus, in order to report a price lower than the true transaction price, the seller and buyer must collude and coordinate the false price that will be reported to the Tax Authority.[8]

According to the Ministry of Finance (2017), the vast majority of the real estate government tax revenue in Israel comes from purchase tax, which comprises about 60%–65% of all real estate government tax revenues. Under the Israeli tax code, there are two types of purchase tax: one for purchasers of a single housing unit (including, over time, about 70%–80% of the transactions—hereinafter referred to as "single unit owner" transactions) and the other for purchasers who already own one or more housing assets (about 20%–30% of the transactions—hereinafter "investor" transactions). Figures 1A and 1B present the purchase tax brackets for single unit owners and investors, respectively. As shown in the figures, in general, the greater the reported price of the transaction, the greater is the required purchase tax payment. Moreover, for each dollar level of a transaction, tax brackets for investors exceed those of single unit owners.

Our dataset includes all housing transactions reported to the Israel Tax Authority in 58 cities with the greatest transaction volume in Israel over the period 1998–2015—a total of about 650,000 transactions.[9] Transaction data includes the closing price (denoted by *P*), transaction date, and a series of asset characteristics, including asset type—such as condominium, detached, duplex, etc. (*Type*1–*Type*7), size in square feet (*SqFt*), number of rooms (*Rooms*), structure age (*Age*), regulatory eligibility for additional building rights according to the National Outline Plan No. 38

[7] In general, over the examined period, an exemption from capital gains tax was granted in the case of a sale of a single residential apartment once every four years.
[8] Most commonly, housing units are sold in Israel via brokers or person-to-person (auctions are very rare). The commission of the broker and lawyers (one for each party) is often a function of the reported closing price—generating further incentive for the parties to collude around the under-report.
[9] The raw data include about 1.2 million transactions. However, we omitted observations with missing or erroneous data and in cities with low transaction volume, and thus were left with 58 cities with the greatest transaction volume of the total of 76 cities in Israel.

(*Rights*),[10] as well as characteristics of the statistical area in which the asset is located, including score on a socio-economic index (*SocEcon*) and average monthly income per standard person (*Income*).[11] Table 1 presents summary statistics of the sample. It follows that the typical transacted asset is a 3.6-room, 930-square-foot, 24-year-old condominium unit whose value is about $245K. The average score on the socio-economic index (ranging from -3 to +3) of the statistical area in which the asset is located is 0.3, the average number of years of schooling of household head is about 14, and average monthly income per standard person is equal to about $1,430.

As we further describe in the method section below, identification of under-reports is based on the sub-sample of repeat-sale transactions. Of the entire sample, we identify about 120,000 repeat-sale transactions of about 55,000 housing units, of which 89% are transacted twice, 10% three times, and 1% four times or more. Table 2 presents summary statistics of the repeat-sale transactions. It follows that the typical repeat-sale asset is a 3.4-room, 830-square-foot, 28-year-old condominium unit whose value is about $215K.[12] Characteristics of the statistical area in which the assets are located resemble those of the entire sample, as the average score on the socio-economic index is equal to about 0.3, average number of years of schooling of household head is about 14, and average monthly income per standard person is equal to about $1,400.

3. METHOD AND RESULTS

Consider the following empirical model for classifying under-reported transactions:

[10] The National Outline Plan No. 38, originally targeted at seismic strengthening of existing buildings, provides additional building rights to buildings constructed prior to 1980.

[11] A statistical area—the Israeli equivalent of a census tract—is the smallest geographic area examined by the Israel Central Bureau of Statistics. Each statistical area includes about 3,000–5,000 residents, and, as with census tracts, the division into statistical areas accounts for aspects of homogeneity with respect to population characteristics, economic status, and living conditions (see ICBS, 2013). The socio-economic index (provided by the Israel Central Bureau of Statistics) ranges from -3 to +3 and is generated by 16 indicators of the statistical area, clustered into 4 groups: standard of living, employment and welfare, schooling and education, and demography (see Israel Central Bureau of Statistics, 2013). Note also that in what follows, all U.S. dollar figures are translated from new Israeli shekels (NIS), where $1=4NIS. Finally, "income per standard person" is a measure used by the Israel Central Bureau of Statistics that is equivalent to "income per capita"—however, where the first person in the household weighs most heavily and weights gradually decrease with each additional person (see ICBS, 2013).

[12] It follows that the average size of units in the repeat-sale sample is slightly smaller than that of the entire sample. In the analysis in Section 6 below we control for unit characteristics.

$$(1) \qquad UnderReport_{jt} = \begin{cases} 1 & \text{if } \psi_j \ll 0 \\ 0 & \text{otherwise} \end{cases}$$

$$(2) \qquad \ln(P_{it}) = \alpha_{0,t} + \vec{\alpha}_{1,t} CHARACTERISTICS_{it} + \vec{\alpha}_{2,t} LOCATION_{it} + \varepsilon_{it} \text{ for all } t,$$

$$(3) \qquad \ln(\hat{P}_{it}) = \hat{\alpha}_{0,t} + \hat{\alpha}_{1,t} CHARACTERISTICS_{it} + \hat{\alpha}_{2,t} LOCATION_{it},$$

$$(4) \qquad \varepsilon_{jt} = \ln(P_{jt}) - \ln(\hat{P}_{jt}),$$

and

$$(5) \qquad \varepsilon_{jt} - \varepsilon_{jt'} = \beta_0 + \beta_1 \ln(\Delta T_j) + \beta_2 [\ln(\hat{P}_{jt}) - \ln(\hat{P}_{jt'})] +$$
$$\vec{\beta}_3 CHARACTERISTICS_j + \vec{\beta}_4 LOCATION_j + \vec{\beta}_5 TFE_{jt} + \vec{\beta}_6 TFE_{jt}^{t'} + \psi_{jt},$$

where the condition in (1) presents the under-reported transaction classification and equations (2)–(5) are auxiliary equations as described below. The indices *i* and *j* denote all transactions and the subset of repeat-sale transactions, respectively, and the indices *t* and *t'* denote a time period in which the asset is transacted and the pervious (adjacent) time period at which the same asset is transacted (for the repeat-sale transactions), respectively, where $t' < t$.

The left-hand side variable $UnderReport_{jt}$ defined and derived in (1) receives the value 1 when the reported transaction price at time *t* is classified as an under-report, and zero otherwise (see description below). The classification method in (1) is based on the following steps. We first estimate a hedonic price equation in (2) based on the universe of all housing transactions, where the dependent variable, $\ln(P_{it})$, is the log of the reported closing price of the transaction and the independent variables include *CHARACTERISTICS*, a vector of asset physical characteristics—including *TYPE*, a vector of structure type (see once again *Type*1–*Type*7 in Table 2); *Rooms*, a vector of dummy variables indicating the asset's number of rooms; ln(*SqFt*), the log of the asset area in square-feet; ln(*Age*), the log of the structure's age; *New*, a dummy variable that equals 1 for new assets (less than 1 year old) and zero otherwise; and *Rights*, a dummy variable that equals 1 for assets eligible for additional building rights according to the National Outline Plan No. 38 and zero otherwise. The independent variables in (2) further include *LOCATION*, a vector of locational attributes—including *CFE*, a vector of city fixed-effects and ln(*Income*), the natural logarithm of the average income-per-standard-person (an income measure provided by the Central Bureau of Statistics, which weighs most heavily the first person in the household and gradually less each additional person) in the statistical area in which the asset is located (on the definition of statistical areas, see once again footnote 10). Also, $\alpha_{0,t}$ and $\vec{\alpha}_{1,t} - \vec{\alpha}_{2,t}$ are estimated parameter and vector of parameters, respectively, and ε_{it} is a random disturbance term. We estimate Equation (2) for every quarter over the period 1998–2015 (a total of 72 estimations) using the universe of all housing transactions that occurred in Israel over the examined period—a total of about 650,000 observations.

Following the estimation of (2), for the sub-sample of repeat-sale transactions, equation (3) generates $\ln(\hat{P}_{it})$ and $\ln(\hat{P}_{it'})$, the projected log of the price of adjacent repeat-sale transactions, representing the associated average market price of same-quality assets at *t* and *t'*, respectively. These price projections are then used in equation (4) to derive ε_{jt} and $\varepsilon_{jt'}$, the differences between $\ln(P_{it})$ and $\ln(\hat{P}_{it})$ and between $\ln(P_{it'})$ and $\ln(\hat{P}_{it'})$, respectively. (Note that, as we estimate the *log* of the price in equation (2), ε effectively estimates the percentage difference between P and $\hat{P}$.) Hence, ε_{jt} and $\varepsilon_{jt'}$ indicate the extent to which the reported price *P* is different from the average projected price $\hat{P}$ of similar (quality-adjusted) assets at t and t', respectively. The parameter ε thus assesses the inconsistency between the reported price of an asset and the average market price of similar assets.

Finally, in equation (5), we estimate the difference between ε_{jt} and $\varepsilon_{jt'}$ on a set of independent control variables, including $\ln(\Delta T)$, the log of the elapsed time (in months) between two adjacent sales of the same asset (repeat-sales), controlling for a possible mismatch between ε_{jt} and $\varepsilon_{jt'}$ that follows from unobservable deterioration or improvement in the asset; $\ln(\hat{P}_{jt}) - \ln(\hat{P}_{jt'})$, the change between period *t* and *t'* projected log price of same-quality assets, controlling for the possible effect of *j*'s idiosyncratic change in the associated same-quality average market price level; and TFE and $TFE^{t'}$, vectors of time (year) fixed-effects at *t* and *t'*, respectively, controlling for time-varying economic effects such as changes in house price level and price dispersion. Also $\beta_0 - \beta_1$ and $\vec{\beta}_{1,t} - \vec{\beta}_{2,t}$ are estimated parameter and vector of parameters, respectively, ψ_{jt} is a random disturbance term, and all other variables are as described above.[13]

It follows that ψ in equation (5) represents a *controlled* measure of the inconsistency between the reported prices at time *t* and *t'* (each compared to the respective quality-adjusted projected price). In particular, according to the condition in (1), for a time *t* transaction to be classified as an under-report, it is required that $\psi_j \ll 0$, i.e., that the time *t* reported price, P_{jt} (relative to time *t* same-quality average market price, $\hat{P}_{jt}$) is statistically-significantly smaller than the time *t'* (the previous time at which the property was transacted) reported price, $P_{jt'}$, of the same asset (once again, relative to the time *t'* same-quality average market price, $\hat{P}_{jt'}$). Otherwise, the time *t* transaction is classified as a non-under-report.

For example, consider a time *t* reported transaction price P_{jt} that is equal to \$250K when the projected time *t* average market price of same-quality assets according to equation (3), $\hat{P}_{jt}$, is \$500K. If the price report at time *t'* (i.e., the previous time that the asset was transacted) is, say, \$100K when the projected time *t'* average market price of same-quality assets is \$200K, then, the estimated difference between ε_{jt} and $\varepsilon_{jt'}$ is zero (i.e., $\varepsilon_{jt} - \varepsilon_{jt'} = 0$) and, following (5), the time *t* price is classified in (1) as a non-under-report [recall that under the logarithmic functional structure of (2), ε effectively estimates the *percentage* difference between P and $\hat{P}$]. In other words, it is likely in this example that the reported price being consistently lower than the associated same-quality average market price in *t* and *t'* due to unobserved asset characteristics rather than being an under-report.[14] It follows that P_{jt} is classified as an under-report only if ε_{jt} is inconsistently (statistically significantly) smaller than $\varepsilon_{jt'}$.

Figure 2 schematically presents examples of potential classifications that follow from (1). Points *a*, *b*, and *c* present three hypothetical outcomes of ψ_{jt} that follow from the computation of (2)-(5). For point *a*, $\psi_{jt} > 0$ and it is thus classified as a non-under-report; for *b*, $\psi_{jt} < 0$, however, it falls above the lower bound of the confidence interval and is thus statistically insignificantly different from zero—hence, once again,

[13] Equation (5) is estimated by WLS to address heteroskedasticity in the difference $\varepsilon_{jt} - \varepsilon_{jt'}$. Outcomes of the estimation of (5) are robust to using a Robust Standard Error procedure.

[14] Of course, it could be that both P_{it}=250K and $P_{it'}$=100K represent under-price reports; however, for reasons of conservatism, they would not be identified as such under our approach in equation (1).

classified as a non-under-report; and for *c*, $\psi_{jt} < 0$ while it falls below the lower bound of the confidence interval and is therefore classified as an under-report.

Outcomes on the Extent of Under-Reporting

Following the classification method in equations (1)–(5), we assess the extent of under-reporting in the market.[15] Figure 3 presents the outcomes on the share of transactions classified as under-reports out of the repeat-sale transaction sample. Specifically, the unbroken dark line provides the estimated relative share of under-reports for different *threshold* rates below the adjusted projected market price, $\hat{P}_{jt} \times (P_{jt'}/\hat{P}_{jt'})$—i.e., the share of transactions j that conform to ψ_j being below a given threshold level. (Since ψ_j in (1) is derived from the logarithm of the price [in (2)], we compute the rate at which the time *t* reported price, P_{jt} is below its adjusted market price, $\hat{P}_{jt} \times (P_{jt'}/\hat{P}_{jt'})$, by the transformation $1 - e^{\psi_j}$.[16]) Accordingly, the grey unbroken line shows the *average* rate below the adjusted projected market price of that subsample classified as under-reports. As expected, the greater the threshold level $1 - e^{\psi}$, the greater the share of transactions classified as under-reports. For example, for a threshold of $1 - e^{\psi}$ = 22% below the projected adjusted market price, the share of under-reports equals 8%, where their average rate below the adjusted market price equals 33%. Finally, the dotted line depicts the *p*-value of the confidence interval around ψ_j that corresponds to the share of transactions classified as under-reports. For example, the subsample of 8% share of under-reports (average rate below the adjusted market price is equal to 33%) is associated with *p*-value equal to 0.1. Decreasing the *p*-value of the confidence interval naturally decreases the relative share of transactions that are classified as under-reports. Indeed, the cutoff *p*-value used for generating the confidence interval around ψ_j and the associated number of under-reported

[15] Results obtained from the 72 estimations of equation (2) (one for each quarter) are not reported and are available upon request. Average number of observations in each estimation is equal to 9,038 (min=4,020; max=18,732) and average R^2 is equal 0.82 (min=0.77; max=0.90). Also, classification results are robust to estimating equation (2) in a log-linear specification. Finally, results obtained from the estimation of equation (5) are not reported and are available upon request.

[16] Note that if P_{jt} maintains $P_{jt} = \hat{P}_{jt} \times (P_{jt'}/\hat{P}_{jt'})$ [implying that $(P_{jt}/\hat{P}_{jt}) = (P_{jt'}/\hat{P}_{jt'})$], then $\psi_{jt} = 0$. Instead, $P_{jt} \ll \hat{P}_{jt} \times (P_{jt'}/\hat{P}_{jt'})$ implies $\psi_{jt} \ll 0$.

transactions may be determined by practical requirement (and budget constraints) of tax enforcement authorities.[17]

The Extent of Forgone Purchase Tax Revenue

We assess the foregone tax revenue that results from under-reporting in the following way: (*a*) We first correct the price of the transactions that are classified as under-reports such that their associated non-under-report projected price, denoted by P_{jt}^{NU}, satisfies the condition $\psi_j = 0$ [i.e., following the requirement in equation (1), P_{jt}^{NU} is set such that the transaction would not have been classified as an under-report]; and (*b*) For each of the under-reports, we compute the purchase tax liability under both P_{jt}^{NU} and P_{jt}, and the difference between the resulting tax liabilities is the forgone tax.

Figure 4 presents the rate of incremental tax revenue as a function of the threshold level of $1 - e^{\psi}$ below which we classify the transaction as an under-report. As expected, the lower the threshold level of $1 - e^{\psi}$, the greater the estimated forgone tax revenue. Thus, for example, for a threshold level of 15% (25%), the loss from uncollected tax due to under-reporting is equal to about 5.5% (2.3%) of the total purchase tax revenue. In nominal terms, provided that purchase tax revenue in 2016 was about US $1.1 billion [equal to about 0.4% of GNP—see Ministry of Finance (2017)], the latter implies forgone annual taxes of about $60 ($25) million.

4. IDENTIFICATION AND ROBUSTNESS TESTS OF THE CLASSIFICATION METHOD

The outcomes of the classification in equation (1) are greatly dependent on the estimation of the price equation in (2). Ideally, in order to generate the projected average quality-adjusted market price that corresponds to each transaction, we would have wished to estimate (2) based solely on a sample of non-under-reports (i.e., net of the under-reports). However, the latter is of course not directly observed *ex ante* [as under-reports are only derived *ex post*, as an output of the classification process in (1)]. In this section, we therefore examine the robustness of our classification method to the inclusion of under-reports in the estimation of equation (2). The robustness checks include simulation and empirical approaches as described below.

[17] A related consideration is that under-reporting might be strategic in a way that the parties to the transaction would be cautious not to report below a certain price level that would become particularly suspicious.

Simulation Approach

Consider a housing unit j, j=1,…,N, that is transacted at two different time periods, t and t' ($t > t'$). We suppose that j's reported price at time s [s=(t, t')], P_{js}, conforms to $\ln(P_{js}) = \beta_0 + \beta_1 x_{js} + \beta_2 z_{js} + \varepsilon_{js}$, where, without loss of generality, x_{js} is a standard-normally distributed observable characteristic, where, for simplicity, $x_{jt} = x_{jt'}$; z_{js} is a standard-normally distributed unobservable characteristic (for simplicity, $z_{jt} = z_{jt'}$); β_0, β_1, and β_2 are parameters; and ε_{js} is a standard-normally distributed random disturbance term. Further, for simplicity, we assume that, except for ε_{js}, all variables and parameters are fixed over time (hence P_{js} may change from t' to t only as a result of a change in ε_{js}). Finally, we suppose that a random sample of the transactions in each period (i.e., at t and t'), size of which is equal to αN, $0 < \alpha < 1$, includes under-reports for which the true price sustains the expression $\ln(P_{js}) + \delta$, $\delta > 0$ (i.e., under-reports are δ% below the true transaction price).

Following this setting, we simulate repeat-sale reported prices of 10,000 housing units (that is, N=10,000 in each of the two periods) based on x_j, z_j, β_0, β_1, β_2, ε_{js}, α, and δ. We then apply the classification method in (1)–(5) for this set of simulated prices. The simulation allows us to test the validity of our empirical classification method, specifically examining whether the estimation of equation (2) that is based on all reported prices (including both true and under-reports) leads to an unbiased identification of under-reports.

Figure 4 presents simulation results of the classifications process for various levels of α, assuming δ=0.4 (i.e., under-reports are 40% below the true price) and $\beta_2 = 1$.[18] The dotted line depicts the share of simulated under-reports that are detected by the classification method [i.e., the identified under-reports in t out of the total simulated under-reports in t]. The dark solid line depicts the share of simulated under-reports that is detected by the classification method, net of those units that are simulated as under-reports in both periods t' and t. Finally, the lighter solid line depicts the share of

[18] We fix $\beta_0 = \ln(200k) \approx 12$ so that the average unit value in the simulation is about equal to that of our repeat-sale empirical data. Also, in order to maintain a sufficient cross-sectional price variance while avoiding negative prices, we fix $\beta_1 = 4$. Note that the larger the ratio of β_1/β_2 in the simulation, the greater the price effect of the observable (x_{js}) over the non-observable (z_{js}) variable. Simulation outcomes, however, are largely robust to changing β_1 while holding $\beta_2 = 1$. (Results are not reported but are available upon request.)

simulated non-under-reports that are erroneously identified as under-reports. It follows that for reasonable levels of α ($0.01 < \alpha < 0.15$), about 61%–95% of the simulated under-reports are correctly identified by the classification method in equation (1)–(5). Moreover, excluding those units that are simulated as under-reports in both periods, the classification method correctly detects 71%–95% of the simulated under-reports. Further, only about 1%–9% of the simulated non-under-reports are mistakenly identified as under-reports by the classification method.[19] Outcomes presented in Figure 5 are robust to reasonable levels of δ (i.e., 0.3 < δ < 0.4) and β_2 (i.e., 0.5 < β_2< 2).[20]

Empirical Approach

We empirically examine the robustness of the price estimation in (2) to the inclusion of under-reports by estimating equation (2) both before and after omitting the transactions that are classified as under-reports. That is, we first apply the classification process in equations (1)–(5) based on the original sample. We then omit those observations that are identified as under-reports and re-estimate equation (2). For each transaction in our sample, we then compare the projected price that follows from equation (3), pre- and post-omission of the under-reports.

Column 1 in Table 4 reports on the average percentage difference between the two projections (pre- and post-omission of under-reports) that follow from equation (3) across all sample observations. The results indicate that the price estimation in equation (2) is largely robust to the inclusion of under-reports. Specifically, it follows from column 1 that even under a relatively narrow confidence interval [that is based on a p-value of 10% around ψ_{jt}—see again the classification condition in (1)], the average absolute difference between the two price projections of the same asset is equal to only about 0.2%. As expected, the

[19] For example, for $\alpha = 10\%$, 70% (1.8%) of the under- (true-) reports are correctly (mistakenly) identified as under-reports. It follows that for $\alpha = 10\%$, the probability that an under-report classification is correct is equal to about 81% [i.e., $(10\% \times 70\%) / [(10\% \times 70\%) + (1 - 10\%) \times 1.8\%] = 81\%$].

[20] In the Appendix, we analytically derive the mean and variance of the reported prices, showing that while the inclusion of under-reports generates a somewhat lower price projection $\widehat{P}_{it}$ (depending of course on the share of under-reports in the population), at the same time the resulting confidence interval of $\widehat{P}_{it}$ widens—thus effectively compensating for the lower projections and maintaining the validity (and conservatism) of the identification process in (1)–(5).

difference decreases as we extend the confidence interval used in the under-report classification. Columns 2, 3, and 4 in Table 4 further present the standard deviation, 90th percentile, and median of the absolute difference between the two price projections (before and after). All figures support the conclusion that the estimation of (2) is practically robust to the inclusion of under-reports.

We further compare the classification outcomes that follow from (1) based on each of the two estimations of equation (2) (pre- and post-omission of under-reports). That is, we re-classify the entire sample based on the estimation of (2)—post-omission of under-reports—and compare the outcomes to the original classification [that is based on the estimation of (2) that includes all sample observations]. Column 1 in Table 5 presents the percentage of observations that are re-classified as under-reports out of those originally classified as under-reports. Results once again indicate that the classification method is robust to the inclusion of all price reports in the estimation of (2). In particular, it follows that, for a confidence interval with a p-value of 1%–10% around ψ_{jt}, about 99% of the transactions maintain their under-report classification—that is, in only about 1% of the cases, transactions originally classified as under-reports attain a non-under-report re-classification (see column 2 in Table 5).[21]

5. CHARACTERISTICS OF UNDER-REPORTING

We examine factors that are associated with the likelihood to under-report. Consider the following estimated equation:

(6)

$$UnderReport_{jt} = \gamma_0 + \gamma_1 Tax_{jt} + \gamma_2 Crime_j + \gamma_3 \Delta T_{jt} + \gamma_4 QP_{jt} + \vec{\gamma}_5 CHARACTERISTICS1_{jt} + \vec{\gamma}_6 LOCATION1_j + \vec{\gamma}_7 TFE_{jt} + \vec{\gamma}_8 TFE_{jt}^{t'} + \varphi_{jt},$$

where the indices *j* and *t* denote repeat-sale transactions and time periods, respectively, and where *UnderReport* on the left-hand side of equation (6) is a dummy variable that

[21] We further assess the robustness of the under-report classification to the hedonic price model specification in equation (2). We re-estimate equation (2) in a log-linear (rather than log-log) form, and reclassify the transactions according to equation (1). Results show that the overlap rate between transactions that were classified as under-report is equal to about 86%–88%, implying that outcomes are largely robust to the specification of the hedonic model. (Results of this robustness check are not presented but are available upon request.)

equals 1 for under-reports and zero otherwise—derived in equation (1). The independent variables in (6) include *Tax*, the log of the assessed purchase tax liability (in real terms), where we expect that a greater tax liability incentivizes under-reporting (see further description of *Tax* below); *Crime*, the total number of police records filed in the statistical area in which the transaction occurs over the sample period, where we arguably expect that an area with a greater overall crime rate associates with a greater likelihood of unlawful under-reporting;[22] ΔT, the log of the elapsed time (in months) between two adjacent sales of the same asset (repeat-sales), controlling for a possible mismatch between ε_{jt} and $\varepsilon_{jt'}$ that follows from unobservable deterioration or improvement in the asset (i.e., the greater ΔT is, the more likely that a difference develops between ε_{jt} and $\varepsilon_{jt'}$); QP, the price quantile to which the transaction price projection belongs (out of all properties transacted in period *t*), controlling for the transaction price level; *CHARACTERISTICS*, a vector of the transacted asset's physical attributes (including structure age, number of rooms, and type), controlling for asset characteristics; *LOCATION*1, a vector of locational attributes (including *CFE*, a vector of city fixed-effects and *SocEcon*, the score on the socio-economic index of the statistical area in which the asset is located); and TFE and $TFE^{t'}$ are vectors of time (year) fixed-effects at *t* and *t'*, respectively, controlling for time-varying effects (e.g., house price dispersion). Also, $\gamma_0 - \gamma_4$ and $\vec{\gamma}_5 - \vec{\gamma}_8$ are estimated parameters and vector of parameters, respectively, and φ is a random disturbance term.

As noted earlier, according to the Ministry of Finance (2015), there are two purchase tax codes: one for purchasers of a single housing unit ("single unit owner" transactions) including, over time, about 70%–80% of the total transactions and the other for purchasers who already own one or more housing assets ("investors")—see once again Figures 1A and 1B. As we do not directly observe the tax paid on each transaction, we assess the *Tax* variable on the right-hand side of (6) in the following way: (*a*) for all transactions whose reported price P_{it} satisfies the condition $\psi_j < 0$ [where the latter inequality is either statistically significant or statistically insignificant—see equation (1)], we generate a hypothetical non-under-report, denoted

[22] We observe the total number of police records by statistical area for the entire period 2011–2015 provided to us by the Israel police. These data, however, are available for only about 60% of the statistical areas for which housing transactions are observed. The number of police records per statistical area in the sample over this period ranges from 135 to 3,967 (average=1,025.3 and standard deviation=658.5).

by P_{jt}^{h}, that conforms to the condition $\ln(P_{jt}^{h}) - \ln(\hat{P}_{jt}) = \varepsilon_{jt'}$ (i.e., P_{jt}^{h} is computed such that its level, relative to the same-quality time *t* average market price, is consistent with $\varepsilon_{jt'}$—the associated difference between the reported and the same-quality average market price at time *t'*; (*b*) following the tax code in Figure 1A (that, as noted earlier, applies to the majority of the transactions), we compute the tax liability on P_{jt}^{h} (P_{jt}) for those transactions whose reported price P_{it} conforms to $\psi_j < 0$ ($\psi_j \geq 0$). In other words, *Tax* represents the theoretical tax liability for each transaction, where we use P_{jt}^{h} (rather than P_{jt}) for those transaction that are classified as under-reports.

We estimate equation (6) based on the sample of repeat-sale transactions over the period 1998–2015.[23] Table 6 presents the outcomes from the estimation of equation (6), where *UnderReport* is derived under different confidence intervals around ψ_j (see, once again, Table 3). The estimated parameters support our hypotheses. Specifically, it follows from column 1 in Table 6 that a 10% increase in the dollar value of the estimated purchase tax liability associates with a 2%–4% increase in the likelihood of an under-report (significant at the 1%-level); an increase of 1,000 police records filed in the statistical area in which the asset is located is associated with about 12%–23% increase in the likelihood of under-reporting (significant at the 1%-level); and a one-standard deviation increase in the socio-economic index of the statistical area in which the asset is located is associated with a 23%–37% decrease in the likelihood of under-price reporting (significant at the 1%-level).[24] Also, as expected, the coefficient on ΔT is positive and significant (at the 1%-level), implying that the greater the elapsed time between two adjacent sales of the same asset (repeat-sales), the more likely that a difference develops between ε_{jt} and $\varepsilon_{jt'}$ such that the transaction is classified as an

[23] Results obtained from the estimation of (6) are robust to the exclusion of repeat-sale transactions for which $\psi_j > 0$ [see again the classification condition in (1)]—thus maintaining only transactions for which $\psi_j < 0$ [in which case *UnderReport* equals one (zero) if the latter inequality is statistically significant (insignificant)]; these results are not reported but are available upon request.

[24] A coefficient equal to 0.16 on *Tax* in column 1 of Table 6 is interpreted such that a 10% increase in the tax payment associates with a 2% increase in the likelihood of under-reporting, i.e., 10%× [exp(0.17)-1]≈2%. Similarly, as each unit of the variable *Crime* represents 1,000 police records, a coefficient equal to 0.161 on *Crime* is interpreted such that 1,000 additional police records in the statistical area are associated with [exp(0.13)-1]≈14%, an increase in the likelihood of under-reporting. Finally, a coefficient equal to -0.34 on the variable *SocEcon* is interpreted such that a one-standard deviation increase in *SocEcon* associates with an 23% decrease in the likelihood of under-reporting, i.e., 0.78×[exp(-0.34)-1]=-23%, where 0.78 is the standard deviation of *SocEcon* (see Table 2).

under-report. Finally, it follows from columns 2 and 3 in Table 6 that the outcomes are robust to the derivation of *UnderReport* across the different confidence intervals around ψ_j.

Are Investors More or Less Likely to Engage in Under-Reporting?

A number of studies document the effect of experience and professionalism on behavior in various economic settings. For example, List (2003, 2004) finds that market experience is inversely related to the magnitude of the endowment effect; Venezia and Shapira (2007) document different trading patterns of amateur and professional investors during the days following a weekend; and Dhar and Zhu (2006) find that investor literacy about financial markets and trading frequency associate with variation in individual disposition effect. In line with these studies, we question whether real estate investors, as compared to single unit owners, are more or less likely to be involved in under-reporting. Consistent with the Israeli tax code, a purchaser-investor (-single unit owner) in our framework is defined as an individual who owns more than one housing units (a single housing unit) and who is then subject to the investor (single unit owner) tax brackets (see once again the tax brackets relevant to single unit owners and investors in Figures 1A and 1B, respectively).

Are investors more or less likely to engage in under-reporting? Intuitively, note that on one hand, compared to households that own a single housing unit, purchasing a number of units generates a greater incentive to under-report, as both the tax base and the marginal tax rate are greater—hence there is more to gain from tax evasion. On the other hand, purchasing a number of units may lead to more prudent behavior on the part of the investor, as both the risk of being caught and the expected penalty increase.

We observe the identity of the purchaser (whether an investor or a single unit owner) for the sub-sample that includes all housing transactions that occurred in the Central district of Israel (the greater Tel Aviv area, as defined by the Israel Central Bureau of Statistics) in the year 2015. Of the approximately 19,000 observations in this sub-sample, 3,647 are repeat-sale transactions [for which we can compute *UnderReport* according to equations (1)–(5)], about 24% of which include a purchaser who is identified as an investor (see summary statistics in Tables 1 and 2).

To examine the difference in the likelihood of engaging in under-reporting between investors, as compared to single unit owners, we estimate the following variation of equation (6):

(7)

$$UnderReport_{jt} = \delta_0 + \delta_1 Tax_{jt} + \delta_2 Crime_j + \delta_3 \Delta T_{jt} + \delta_4 Investor_{jt} + \delta_5 Investor_{jt} \times Tax_{jt} + \delta_6 QP_{jt} + \vec{\delta}_7 CHARACTERISTICS1_{jt} + \vec{\delta}_8 LOCATION_j + \vec{\delta}_9 TFE_{jt} + \vec{\delta}_{10} TFE_{jt}^{t'} + \eta_{jt},$$

where *Investor* is a dummy variable that equals 1 if the purchasing party is an investor and zero otherwise; and $Investor \times Tax$ is an interaction term for the variables $Investor$ and Tax, where Tax is, once again, the log of the purchase tax liability computed as suggested earlier [see description following equation (6)]—however, separately assessed for single unit owners and investors according to the tax code in Figures 1A and 1B, respectively. Also, $\delta_0 - \delta_6$ and $\vec{\delta}_7 - \vec{\delta}_{10}$ are estimated coefficients and vectors of coefficients, respectively, η is a random disturbance term, and all other variables are as described above.

Column 1 in Table 7 presents the outcomes from estimating Equation (7), where *UnderReport* is derived under a confidence interval with a *p*-value equal to 10%. It follows that the coefficients on $Investor$ and $Investor \times Tax$ are negative and positive, respectively (significant at the 1%- and 5%-levels, respectively).[25] However, as the tax code that applies to investors differs from that for single unit owners—generally implying considerably greater tax payments (see, once again, the tax codes in Figures 1A and 1B), we use the sample observations to interpret the implications of the estimated coefficients. Specifically, following the estimation of (7), for all *j*, we compute the ratio

(8a)

[25] As this estimation is based on a sub-sample for which we observe the variable *Investor* (i.e., transactions that occurred in the Central district of Israel in 2015), we estimate the model for confidence intervals with *p*-value equal to either 10% or 5%. The relatively smaller number of transactions categorized as under-reports under a 1 standard deviation confidence interval is likely the explanation for the decrease in the statistical significance of coefficients when moving from column 1 to column 2 in Table 7—note, however, that the coefficients are highly similar. Finally, note that the estimated signs on other independent variables obtained from the estimation of equation (6) are robust under the specification of (7).

$$R_{1,j} = \frac{\widehat{UnderReport}_{jt}\left(Investor = 1, Tax_j \middle| Investor = 0\right)}{\widehat{UnderReport}_{jt}\left(Investor = 0, Tax_j \middle| Investor = 0\right)},$$

where the numerator on the right-hand side of (8a) is the projected probability of under-reporting that follows from the estimated coefficients of (7), computed for *j* being an investor (i.e., $Investor_j = 1$) whose tax liability, however, is computed as if she is a single unit owner (i.e., $Tax_j|Investor_j$=0); and the denominator is the projected probability of under-reporting, calculated for *j* being a single unit owner (i.e., $Investor_j$=0) whose tax payment is computed as if she is a single unit owner (i.e., $Tax_j|Investor_j$=0). The idea underlying this exercise is to compare the likelihood of under-reporting between investors and single unit owners, *ceteris paribus* (i.e., while holding the tax code fixed across all *j*). Figure 6A plots the value of the ratio $R_{1,j}$ for all *j* in our sample. As is evident, $R_{1,j}<1$ for only 0.15% of sample observations with an average of $R_{1,j}$ equal to 0.04—implying that, *ceteris paribus*, investors are on average about 25 times *less* likely to be involved in under-reporting (for relatively low levels of tax liability, as commonly applies to single unit owner purchasers).

In Figure 6B we similarly plot the value of the ratio $R_{2,j}$ for all *j* in our sample, where

(8b)

$$R_{2,j} = \frac{\widehat{UnderReport}_{jt}\left(Investor_j = 1, Tax_j \middle| Investor_j = 1\right)}{\widehat{UnderReport}_{jt}\left(Investor_j = 0, Tax_j \middle| Investor_j = 1\right)}.$$

That is, in (8b) we compute Tax_j for all *j* under the assumption that *j* is an investor (i.e., $Tax_j|Investor_j$=1) and once again calculate the ratio between the projected probability of under-reporting for *i* being an investor (numerator) and single unit owner (denominator). It follows from Figure 6B that $R_{2,j}<1$ in 1.55% of sample observations with an average of $R_{2,j}$ equal to 0.41—implying that, *ceteris paribus*, investors are on average about 2.5 times *less* likely to engage in under-reporting (for relatively high levels of tax liability, as commonly applies to purchasers who are classified as investors). Finally, it follows from columns 2 in Table 7 that the outcomes are qualitatively robust to the derivation of *UnderReport* under a confidence interval with *p*-value equal to 5% around ψ_j.

6. SUMMARY

The real estate market is often recognized as a source of attraction for tax fraud. In order to avoid tax payments, reporting a price lower than the true transaction price is a prevalent fraud techniques. In this study, we propose an empirical method for identifying transactions that are suspected of being under-reports. Observing all reported housing transactions in Israel over the period 1998–2015, it follows from our method that, under reasonable assumptions, about 8% of the transactions are under-reported, with an average price report being 33% below the projected true price. We further find that the likelihood to under-report positively associates with the amount of the tax liability and positively (negatively) associated with the crime rate in (socioeconomic level of) the area in which the transaction occurs. Finally, compared to single unit owner purchasers, real estate investors are considerably less likely to engage in under-reporting. Our outcomes are robust to a series of robustness and identification tests. Our proposed empirical approach may be used by tax enforcement authorities to promote tax collection in the real estate market.

REFERENCES

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

Benjamin, J.D., N.E. Coulson, and S.X. Yang, Real estate transfer taxes and property values: The Philadelphia story. *Journal of Real Estate Finance and Economics*, 1993, *7*, 151–157.

Best, M.C., and H.J. Kleven, Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK, *Review of Economic Studies,* forthcoming

Center for Tax Policy and Administration, *Report on tax fraud and money laundering vulnerabilities involving the real estate sector*, 2007, Paris, OECD.

Dhar, R., and N. Zhu, Up close and personal: Investor sophistication and the disposition effect. *Management Science*, 2006, *52*:5, 726–740.

Israel Central Bureau of Statistics, *Characterization and classification of geographical units by the socio-economic level of the population 2008*, Press Release, Jerusalem, Israel, 2013.

Genesove D., and C. J. Mayer, Equity and time to sale in the real estate market. *American Economic Review*, 1997, *87*:3, 255-269.

Genesove D., and C. J. Mayer, Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics*, 2001, *116*:4, 1233-1260.

Kopczuk, W., and D.J. Munroe, *Mansion tax: The effect of transfer taxes on the residential real estate market*. Working paper. National Bureau of Economic Research, 2014.

Leung, C.K.Y., T.C. Leung, and K.P. Tsang, Tax-driven bunching of housing market transactions: The case of Hong Kong. *International Real Estate Review*, 2014, *18*, 473–501.

List, J.A., Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, 2003, *118*, 41–71.

List, J.A., Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica*, 2004, *72*:2, 615-625.

Ministry of Finance. State revenue report, 2009. http://mof.gov.il/ChiefEcon/StateRevenues/StateRevenuesReport/Pages/Report_2008.aspx (Hebrew).

Ministry of Finance. State revenue report, 2017. https://mof.gov.il/ChiefEcon/StateRevenues/StateRevenuesReport/DocLib/2015-2016/Report2015-2016_10.PDF (Hebrew).

Nelen, H., Real estate and serious forms of crime. *International Journal of Social Economics*, 2008, *35*:10, 751–762.

Roger, G., Optimal contract under moral hazard with soft information. *American Economic Journal: Microeconomics*, 2013, *5*:4, 55–80.

Schneider, S., Organized crime, money laundering, and the real estate market in Canada. *Journal of Property Research*, 2004, *21:*2, 99–118.

Shunnar, T., and M. Barry, Tracking fraudulent activities in real estate transactions. FIG Working Week, 2011.

Sullivan, K., *Anti-Money Laundering in a Nutshell: Awareness and Compliance for Financial Personnel and Business*, 2015. Apress.

The State Comptroller and Ombudsman of Israel, Annual Report 58a, Jerusalem, 2007 (Hebrew)

Unger, B., J. Ferwerda, J. Trouw, H. Nelen, and L. Ritzen, Detecting criminal investments in the Dutch real estate sector. Study prepared by the Dutch Ministry of Finance, Justice and Interior Affairs, 2010.

Unger, B., M. Siegel, J. Ferwerda, W. de Kruijf, M. Busuioic, K. Wokke, and G. Rawlings, The amounts and effects of money laundering. Ministry of Finance, The Hague, 2006. www.minfin.nl/binaries/minfin/assets/pdf/actueel/bijlage-nieuwsberichten/2006/02/06-011a. pdf.

Venezia, I., and Z. Shapira, On the behavioral differences between professional and amateur investors after the weekend. *Journal of Banking and Finance*, 2007, *31*:5, 1417–1426.

Acknowledgement

The authors thank Yovav Sanders for assistance in generating the dataset. We also thank Meir Mizrachi and Eli Farjon, the Israel Tax Authority, and participants at Tel Aviv University Law School seminar, 2018 Israel Regional Science Association meetings, and 2019 Israel Law and Economics Association meetings, for helpful comments. We thank the Survey of Israel for financial support.

Table 1: List of Full Sample Variables, Description, and Summary Statistics

Variable	Description	Mean	StD	Min	Max
P	Reported transaction price (US dollars)	244,843	176,178	5,808	5,140,603
Rooms	Total number of rooms in the housing unit	3.617	1.053	2	10
Area	Total area of the housing unit (SqFt)	937.9	376.5	214.0	3,210
Age	Structure age at time of transaction	24.41	17.93	0.0	100
Dum_New	Dummy variable that equals 1 if $Age \leq 1$ (new structure)	0.121	0.326	0	1
*Type*1	Dummy variable equals 1 if the transacted property is a condominium apartment; 0 otherwise (base category)	0.959	0.198	0	1
*Type*2	Dummy variable equals 1 if the transacted property is a ground-level apartment; 0 otherwise	0.006	0.077	0	1
*Type*3	Dummy variable equals 1 if the transacted property is a duplex apartment; 0 otherwise	0.002	0.049	0	1
*Type*4	Dummy variable equals 1 if the transacted property is a townhouse; 0 otherwise	0.002	0.049	0	1
*Type*5	Dummy variable equals 1 if the transacted property is a style 1 attached unit; 0 otherwise	0.019	0.137	0	1
*Type*6	Dummy variable equals 1 if the transacted property is a style 2 attached unit; 0 otherwise	0.008	0.089	0	1
*Type*7	Dummy variable equals 1 if the transacted property is a detached unit; 0 otherwise	0.003	0.056	0	1
SocEcon(SA)	The score on the socio-economic index of the Statistical Area in which the property is located	0.303	0.824	-2.462	2.893
Income(SA)	Average income (in US dollars) of standard person in the Statistical Area in which the property is located	1,436	559	192	4,342
Rights	Dummy variable equals 1 if the transacted property is eligible for National Outline Plan No. 38; 0 otherwise	0.194	0.396	0	1
Investor	Dummy variable equals 1 if the buyer is defined as an investor (owning more than one housing asset); 0 otherwise.	0.243	0.429	0	1

Note: The attribution of the property to a Statistical Area was provided by the Survey of Israel.

Table 2: List of Repeated-Sale Sub-Sample Variables, Description, and Summary Statistics

Variable	**Description**	**Mean**	**StD**	**Min**	**Max**
P	Reported transaction price (in US dollars)	214,095	139,064	6,370	3,200,000
Room	Total number of rooms in the housing unit	3.37	0.89	2	10
SqFt	Total area of the housing unit (SqFt)	829.9	293.2	224.7	3,210
Age	Structure age at time of transaction	28.1	16.7	0	100
Dum_New	Dummy variable that equals 1 if *Age*≤1 (new structure)	0.054	0.226	0	1
*Type*1	Dummy variable equals 1 if the transacted property is a condominium apartment; 0 otherwise (base category)	0.979	0.144	0	1
*Type*2	Dummy variable equals 1 if the transacted property is a ground-level apartment; 0 otherwise	0.005	0.073	0	1
*Type*3	Dummy variable equals 1 if the transacted property is a duplex apartment; 0 otherwise	0.002	0.041	0	1
*Type*4	Dummy variable equals 1 if the transacted property is a townhouse; 0 otherwise	0.001	0.034	0	1
*Type*5	Dummy variable equals 1 if the transacted property is a style 1 attached unit; 0 otherwise	0.008	0.089	0	1
*Type*6	Dummy variable equals 1 if the transacted property is a style 2 attached unit; 0 otherwise	0.004	0.063	0	1
*Type*7	Dummy variable equals 1 if the transacted property is a detached unit; 0 otherwise	0.001	0.034	0	1
SocEcon(SA)	The score on the socio-economic index of the Statistical Area in which the property is located	0.263	0.783	-2.462	2.893
Income(SA)	Average income (in US dollars) of standard person in the Statistical Area in which the property is located	1,401	518	192	4,342
Rights	Dummy variable equals 1 if the transacted property is eligible for National Outline Plan No. 38; 0 otherwise	0.275	0.446	0	1
ΔT	Number of months between two adjacent repeat-sales of the same property	57.64	37.27	1	189
Investor	Dummy variable equals 1 if the buyer is defined as an investor (owning more than one housing asset); 0 otherwise.	0.243	0.429	0	1

Table 3: The Rate of Under-Price Reports, 1998–2015

Column	(1)	(2)	(3)
p-value of confidence interval	10%	1%	1%
Rate of under-reported transactions	7.9%	4.4%	1.7%
Average percent gap from the associated same-quality average market price	32.7%	39.8%	53.3%

Table 4: Percentage Difference between Price Projections that Follow from Equation (3)—Pre- and Post-Omission of Under-Reports

***p*-value of confidence interval**	**Average absolute difference**	**Standard deviation**	**90th percentile of absolute difference**	**Median absolute difference**
	(1)	(2)	(3)	(4)
10%	0.17%	0.25%	0.43%	0.092%
5%	0.13%	0.22%	0.33%	0.07%
1%	0.09%	0.17%	0.20%	0.04%

Table 5: Re-Classification of Under-Reports

***p*-value of confidence interval**	**Percent of observations re-classified as under-reports**	**Percent of observations re-classified as non-under-reports**
	(1)	(2)
10%	98.8%	1.2%
5%	99.1%	0.9%
1%	99.5%	0.5%

Table 6: Regression Results from the Estimation of Equation (6)

Column	(1)	(2)	(3)
Dependent variable	*UnderPrice*	*UnderPrice*	*UnderPrice*
p-value of confidence interval for classifying $UnderPrice_i$	10%	5%	1%
Constant	0.029 (0.562)	-0.756 (0.718)	-0.041 (1.082)
QP	-0.125*** (0.040)	-0.201*** (0.053)	-0.115 (0.085)
Tax	0.163*** (0.014)	0.228*** (0.020)	0.328*** (0.033)
Crime	0.130*** (0.032)	0.145*** (0.042)	0.206*** (0.065)
ΔT	0.300*** (0.028)	0.459*** (0.038)	0.676*** (0.063)
SocEcon	-0.339*** (0.039)	-0.442*** (0.052)	-0.653*** (0.084)
$CHARACTERISTICS1$	Included	Included	Included
CFE (city fixed-effects)	Included	Included	Included
TFE (time fixed-effects)	Included	Included	Included
N	39,361	39,132	38,430
Pseudo R2	0.07	0.09	0.13
P(Chi2)	0.000	0.000	0.000

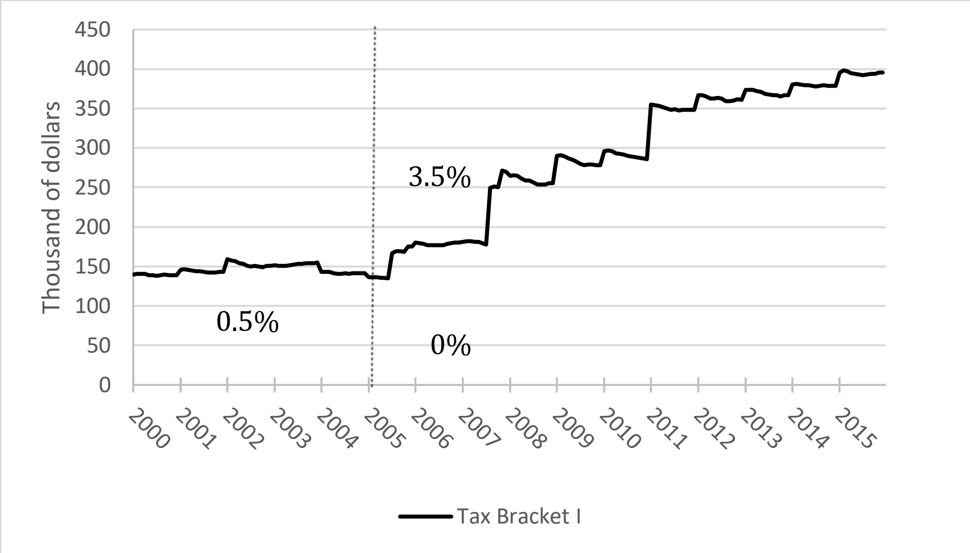
Notes: The table presents the results from the logit estimation of equation (6). Standard errors are shown in parentheses. One, two, and three asterisks denote significance at the 10%, 5%, and 1% levels. Number of observations decreases with the size of the confidence interval as a result of the omission of transactions in those localities where no transactions are identified as under-reports.

Table 7: Regression Results for the Estimation of Equation (7)

Column	(1)	(2)
Dependent variable	*UnderPrice*	*UnderPrice*
p-value of confidence interval for classifying $UnderPrice_i$	10%	5%
Constant	0.354 (2.798)	-2.1 (3.657)
QP	-0.221 (0.240)	-0.363 (0.331)
Tax	0.270*** (0.083)	0.389*** (0.123)
Investor	-8.528*** (3.158)	-11.002*** (4.169)
Investor×*Tax*	1.106** (0.444)	1.460** (0.579)
Crime	0.318** (0.162)	0.201 (0.240)
ΔT	0.115 (0.165)	0.352 (0.248)
SocEcon	-0.302 (0.269)	-0.558 (0.379)
$CHARACTERISTICS1$	Included	Included
CFE (city fixed-effects)	Included	Included
TFE (time fixed-effects)	Included	Included
N Pseudo R2 P(Chi2)	2,034 0.07 0.000	1,987 0.12 0.000

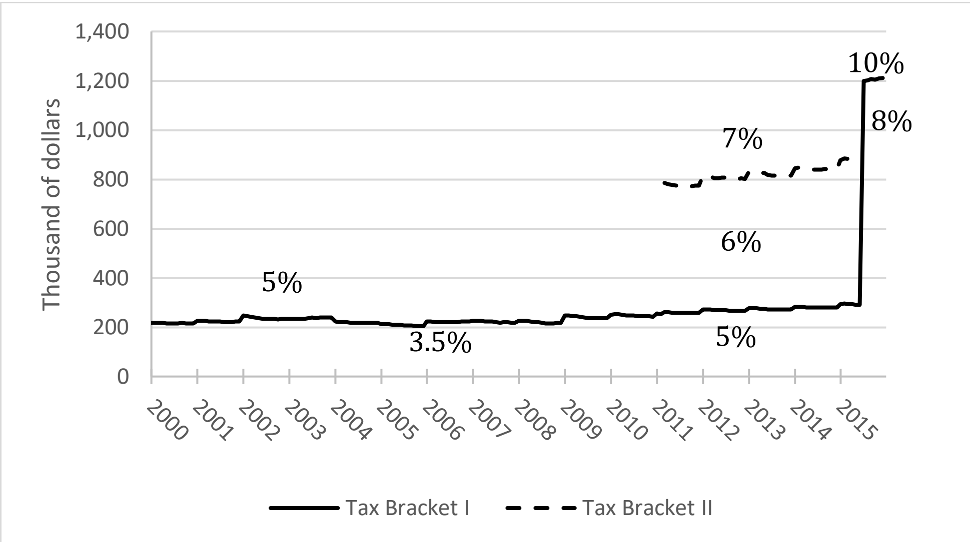
Notes: Results in Table 7 apply to the sub-sample of transactions that occurred in the Central district of Israel in the year 2015. The table presents the results from the logit estimation of equation (7) with the variable *Investor*. Standard errors are shown in parentheses. One, two, and three asterisks denote significance at the 10%, 5%, and 1% levels. Number of observations decreases with the size of the confidence interval as a result of the omission of transactions in those localities where no transactions are identified as under-reports. We omit the test under a confidence interval with *p*-value equal to 1% as only 8 investors are classified as under-reporters in this case.

Figure 1A: Purchase Tax Brackets for Owners of a Single Housing Unit (in 2014 prices), 2000–2015



Notes: The figure presents purchase tax brackets according to 2014 prices (on the y-axis) for an Israeli resident who *owns a single housing unit*. Note that before (after) 2005, the first tax bracket is 0.5% (0%). The next tax bracket is 3.5% (and, following 2013, an additional bracket is included for prices greater than \$1.2 million). The graph is based on the simulator of the Israel Tax Authority (https://www.misim.gov.il/svsimurechisha/frmFirstPage.aspx?cur=1#nbb, May 2017).

Figure 1B: Purchase Tax Brackets for Investors (in 2014 prices), 2000–2015



Notes: The figure presents purchase tax brackets according to 2014 prices (on the y-axis) for an Israeli resident who is not eligible for the "single housing unit ownership" category (i.e., *owning more than one housing asset*). Note that before (after) March 2011, the marginal tax rate of the first bracket is 3.5% (5%). Also, in March 2011, additional brackets are introduced, as depicted in the figure. In July 2015 the marginal tax rate of the first bracket increases to 8%. Post-2014 there is an additional tax bracket for prices exceeding about \$4M. The graph is based on the simulator of the Israel Tax Authority (https://www.misim.gov.il/svsimurechisha/frmFirstPage.aspx?cur=1#nbb, May 2017).

Figure 2: Examples of Typical Price Report Classifications

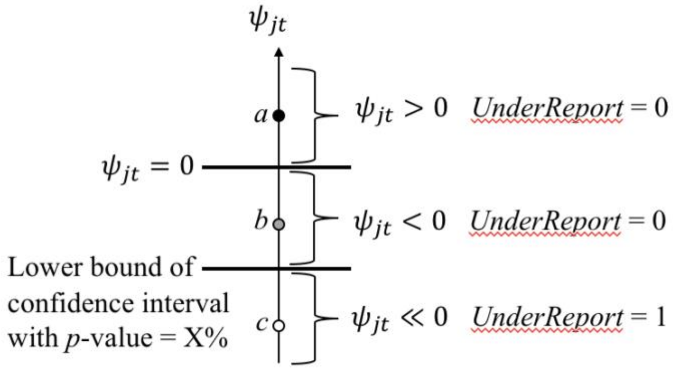


Figure 3: The Share of Under-Reported Transactions, Their Average Rate of Under-Report Relative to the Adjusted Market Price, and the Associated *P*-Value of Confidence Interval

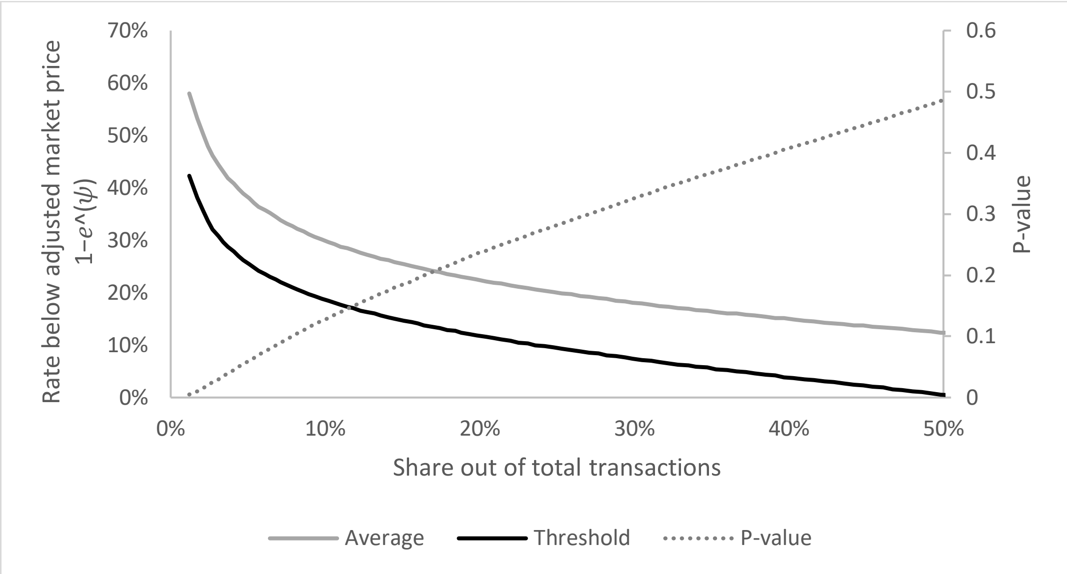


Figure 4: The Rate of Incremental Purchase Tax Revenue as a Function of the Threshold Rate of ψ Below Which Transactions Are Classified as Under-Reports

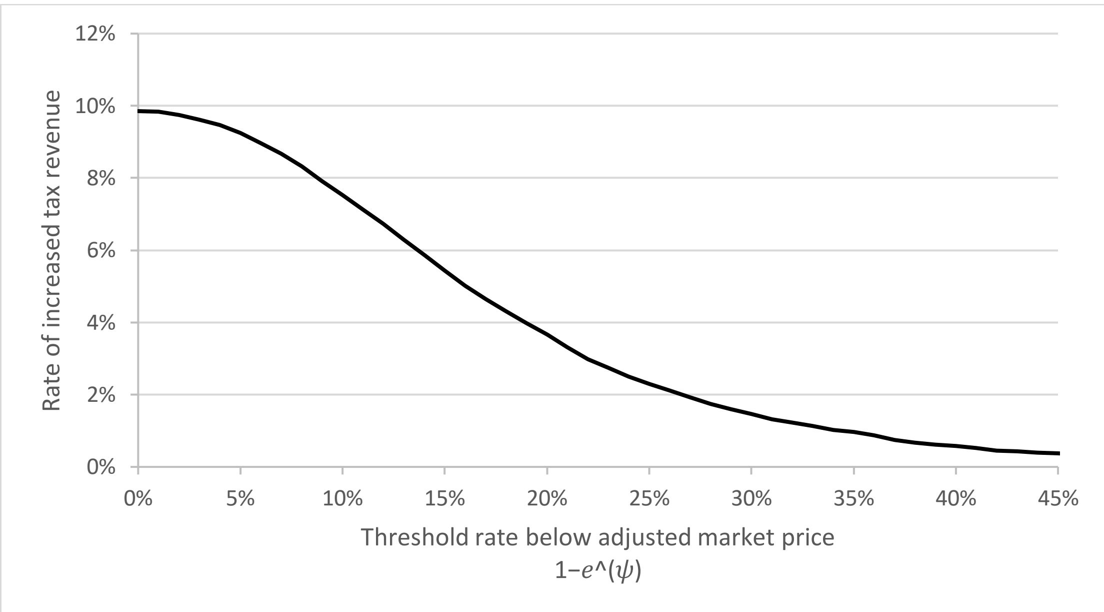
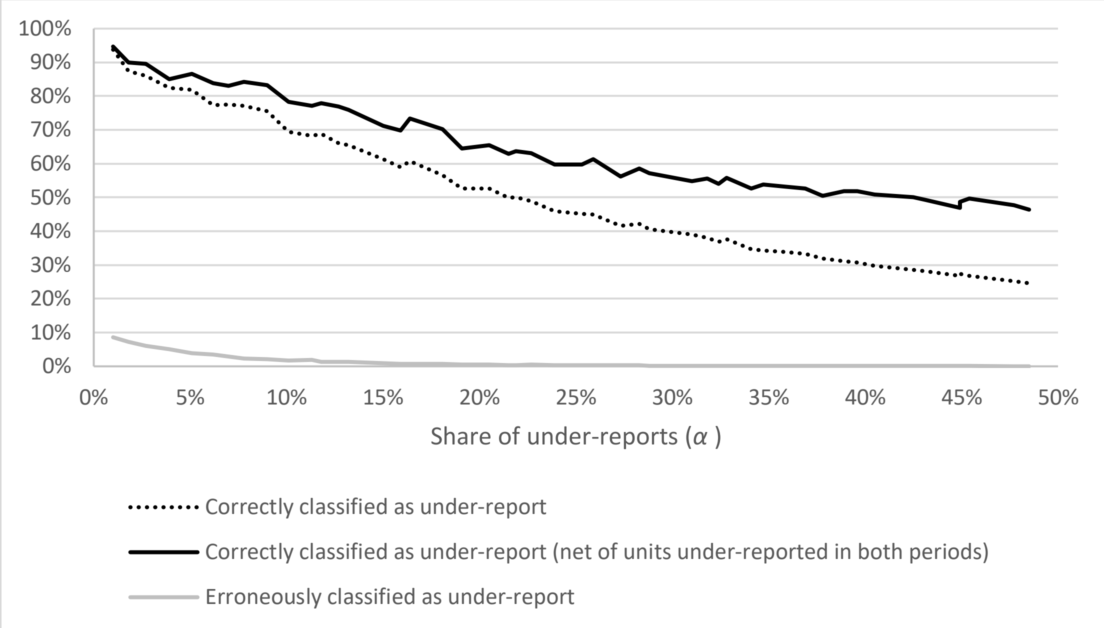
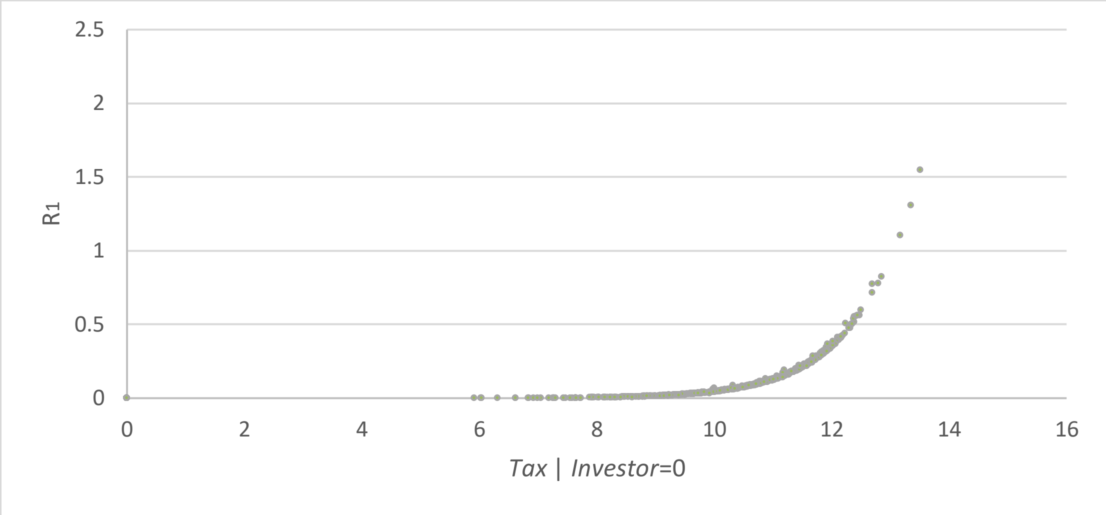


Figure 5: Simulation Results of Under-Report Classification (δ=0.4)



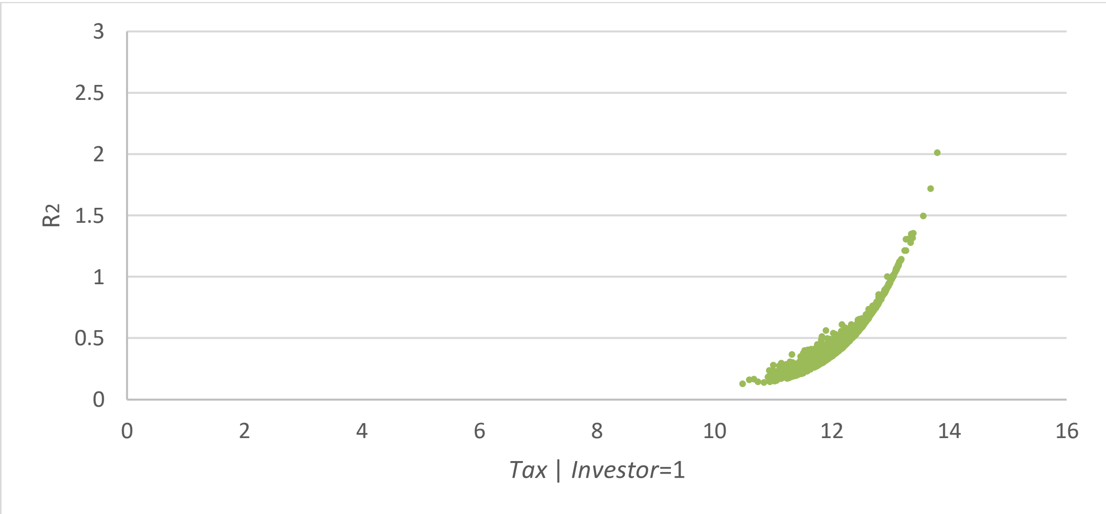
Notes: Figures 5 presents simulation results of the classifications process for various levels of α, assuming δ=0.4 and $\beta_2 = 1$. The dotted line depicts the share of simulated under-reports that are detected by the classification method [i.e., the identified under-reports in *t* out of the total simulated under-reports in *t*]. The dark solid line depicts the share of simulated under-reports that is detected by the classification method, net of those units that are simulated as under-reports in both periods *t'* and *t*. Finally, the lighter solid line depicts the share of simulated non-under-reports that are erroneously identified as under-reports. Outcomes presented in Figure 4 are robust to reasonable levels of δ (i.e., 0.3 < δ < 0.4) and β_2 (i.e., 0.5 < β_2< 2).

Figure 6A: The Ratio $R_{1,i}$ Across All Sample Observations



Notes: Figure 6A plots the value of the ratio $R_{1,i} = \frac{\widehat{UnderReport}_{it}(Investor\ =1, Tax_i|Investor\ =0)}{\widehat{UnderReport}_{it}(Investor\ =0, Tax_i|Investor\ =0)}$ for all *i* in our sample, where the numerator is the projected probability of under-reporting that follows from the estimated coefficients of (7), computed for *i* being an investor (i.e., $Investor = 1$) whose tax liability, however, is computed as if she is a single unit owner (i.e., $Tax_i|Investor$=0); and the denominator is the projected probability of under-reporting, calculated for *i* being a single unit owner (i.e., $Investor$ =0) whose tax payment is computed as if she is a single unit owner (i.e., $Tax_i|Investor$=0).

Figure 6B: The Ratio $R_{2,i}$ Across All Sample Observations



Notes: Figure 6b plots the value of the ratio $R_{2,i} = \frac{\widehat{UnderReport}_{it}(Investor\ =1, Tax_i|Investor\ =1)}{\widehat{UnderReport}_{it}(Investor\ =0, Tax_i|Investor\ =1)}$ for all *i* in our sample, where the numerator is the projected probability of under-reporting that follows from the estimated coefficients of (7), computed for *i* being an investor (i.e., $Investor = 1$) whose tax liability, however, is computed as if she is a single unit owner (i.e., $Tax_i|Investor$=0); and the denominator is the projected probability of under-reporting, calculated for *i* being a single unit owner (i.e., $Investor$ =0) whose tax payment is computed as if she is a single unit owner (i.e., $Tax_i|Investor$=0).

APPENDIX – Analytic derivation of the variance of the reported price:

Suppose that a reported price, P_i, is one of two types: under- or non-under-report. Denote the set of non-under-reports by $\{NU\}$ and the set of under-reports by $\{U\}$, and suppose that the (true) non-under-report of all $i \in \{U\}$ maintians $P_i + \delta$, $\delta > 0$. Further, suppose that the total number of transactions is equal to M and that the share of under-reports within the transaction population is equal to α (so that the number of under-price reports is αM). Finally, suppose that P_i, $i \in \{NU\}$ is normally distributed with mean μ, and variance σ^2 [i.e., $P_{i\in NU} \sim N(\mu, \sigma^2)$] and, correspondingly, that P_i, $i \in \{U\}$ is normally distributed with mean $\mu - \delta$ and variance σ^2 [i.e., $P_{i\in U} \sim N(\mu - \delta, \sigma^2)$.

It follows that the average reported price, $\bar{P}$, is equal to

(A1)

$$\bar{P} = (1-\alpha) \times \mu + \alpha \times (\mu - \delta) = \mu - \delta\alpha$$

and the variance of the reported price, SD^2, is equal to

(A2)

$$SD^2 = \frac{1}{M}\left[\sum_{i\in NU}(P_i - \bar{P})^2 + \sum_{i\in U}(P_i - \delta - \bar{P})^2\right].$$

Let us denote

(A3)

$$P_i^* = (P_i - \mu)/\sigma$$

and

(A4)

$$\bar{P}^* = (\bar{P} - \mu)/\sigma = -\delta\alpha/\sigma.$$

It follows that $P_i^* \sim N(0,1)$ where $\sum_i (P_i^* - 0)^2 = 1$ and $\sum_i P_i^* = 0$.

Note, however, that following (A3) and (A4), the first term inside the brackets on the right-hand side of (A2) can be developed into:

(A5)

$$\frac{1}{M}\sum_{i\in NU}(P_i - \bar{P})^2 = \frac{\sigma^2}{M}\sum_{i\in NU}(P_i^* - \bar{P}^*)^2 = \frac{\sigma^2}{M}\sum_{i\in NU}({P_i^*}^2 - 2P_i^*\bar{P}^* + {\bar{P}^*}^2) \cong (1-\alpha)\sigma^2 - 2\bar{P}^*\frac{\sigma^2}{M}\sum_{i\in NU}(P_i^*) + \sigma^2(1-\alpha){\bar{P}^*}^2 = (1-\alpha)\left(1 + {\bar{P}^*}^2\right)\sigma^2.$$

Similarly, the second term inside the brackets on the right-hand side of (A2) can be developed into

(A6)

$\frac{1}{M}\sum_{i\in U}(P_i-\delta-\bar{P})^2=\frac{\sigma^2}{M}\sum_{i\in U}(P_i^*-\delta/\sigma-\bar{P}^*)^2=\frac{\sigma^2}{M}\sum_{i\in U}[{P_i^*}^2-2P_i^*(\bar{P}^*+\delta/\sigma)+(\bar{P}^*+\delta/\sigma)^2]\cong\alpha\sigma^2-2(\bar{P}^*+\delta/\sigma)\frac{\sigma^2}{M}\sum_{i\in U}(P_i^*)+\alpha\sigma^2(\bar{P}^*+\delta/\sigma)^2=\alpha\sigma^2[1+(\bar{P}^*+\delta/\sigma)^2].$

It follows from (A5) and (A6) that

(A7)

$SD^2=(1-\alpha)\left(1+{\bar{P}^*}^2\right)\sigma^2+\alpha\sigma^2[1+(\bar{P}^*+\delta/\sigma)^2]=\sigma^2\left(1+{\bar{P}^*}^2+2\alpha\bar{P}^*\delta/\sigma+\alpha\delta^2/\sigma^2\right)=\sigma^2+\sigma^2{\bar{P}^*}^2+2\bar{P}^*\delta\sigma\alpha+\delta^2\alpha.$

However, following (A4), (A7) can be expressed as

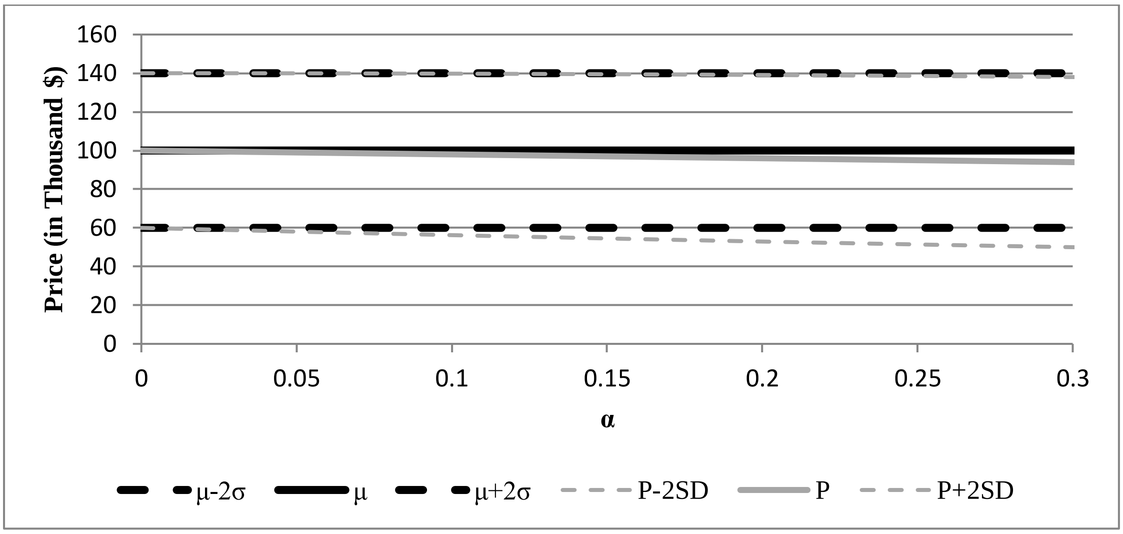
(A8)

$SD^2=\sigma^2+\alpha^2\delta^2-2\alpha^2\delta^2+\alpha\delta^2=\sigma^2+\alpha\delta^2-\alpha^2\delta^2.$

It follows from (A1) and (A8) that for $0<\alpha<1$ and $\delta>0$, not only is the average reported price (across both under- and non-under-reports) lower than the average non-under-report [which immediately follows from (A1)], but also that the variance of the reported price (across both under- and non-under-reports) is greater than the variance of the non-under-reports. Figure A1 presents the resulting confidence interval around the average price report for different shares of under-reports, α. The confidence interval is drawn for the populations that exclude and include under-reports (darker and lighter lines, respectively). It is assumed in the figure that the average non-under-report (μ) is equal to $100,000, with a standard deviation (σ) of $20,000 (i.e., $\sigma=20\%\times\mu$) and that the difference between under- and non-under-reports, δ, is equal to $20,000 (i.e., $\delta=20\%\times\mu$).

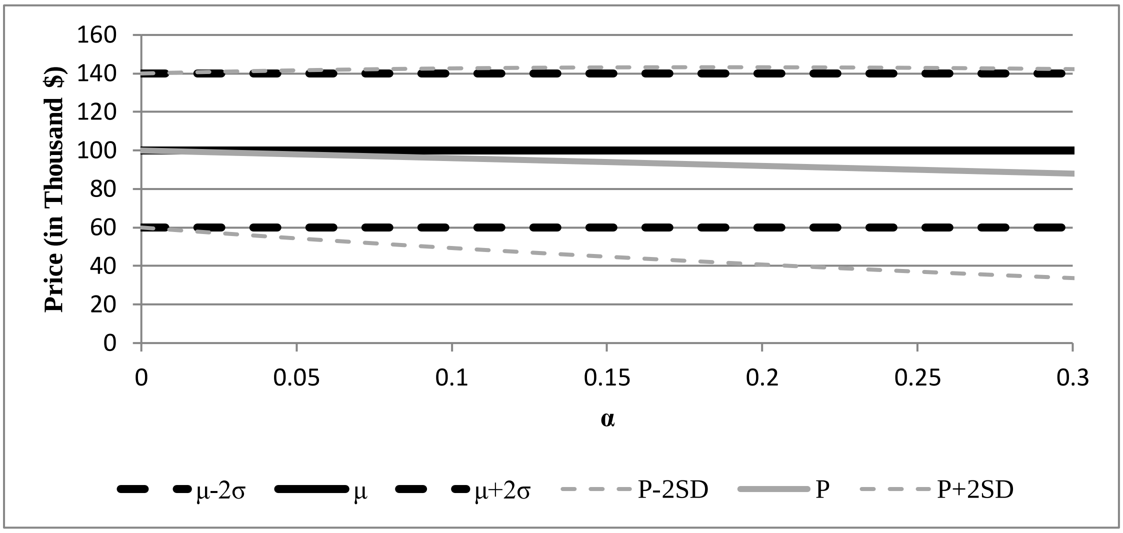
It follows from Figure A1 that while the average price report of the entire population (both under- and non-under-reports) is somewhat lower than that of non-under-reports, the lower bound of the confidence interval of the entire population is considerably below that of only the non-under-reports—for all $0<\alpha<0.3$. This implies that while the price projection that follows from the estimation of equation (2) is somewhat downward biased due to the inclusion of under-reports, at the same time the resulting confidence interval around $\hat{P}_{jt}$ and $P_{jt'}$ widens—thus providing further validity for the classification method in (1). Figure A2 further shows that the latter conclusion is robust to increasing the level of δ (i.e., when $\delta = \$40{,}000 = 40\% \times \mu$).

Figure A1: Confidence Interval around the Estimated Average Price Report for Different Levels of α (δ=20,000 and σ=20,000)



Notes: Figure A1 presents the resulting confidence interval around the average price report for different shares of under-reports, α. The parameters μ and σ are the mean and standard deviation, respectively, of the true prices (non-under-reports) and P and SD represent the mean and standard deviation, respectively, of the entire population (including both under- and non-under-reports).The confidence interval is drawn for the populations that exclude and include under-reports (darker and lighter lines, respectively). It is assumed in the figure that the average non-under-report (μ) is equal to \$100,000, with a standard deviation (σ) of \$20,000 (i.e., $\sigma = 20\% \times \mu$) and that the difference between under- and non-under-reports, δ, is equal to \$20,000 (i.e., $\delta = 20\% \times \mu$).

Figure A2: Confidence Interval around the Estimated Average Price Report for Different Levels of α (δ=40,000 and σ=20,000)



Notes: Figure A1 presents the resulting confidence interval around the average price report for different shares of under-reports, α. The parameters μ and σ are the mean and standard deviation, respectively, of the true prices (non-under-reports) and P and SD represent the mean and standard deviation, respectively, of the entire population (including both under- and non-under-reports).The confidence interval is drawn for the populations that exclude and include under-reports (darker and lighter lines, respectively). It is assumed in the figure that the average non-under-report (μ) is equal to \$100,000, with a standard deviation (σ) of \$20,000 (i.e., $\sigma = 20\% \times \mu$) and that the difference between under- and non-under-reports, δ, is equal to \$40,000 (i.e., $\delta = 40\% \times \mu$).