Transaction-tax Evasion in the Housing Market

José G. Montalvo
Amedeo Piolatto
Josep Raya

This version: January 2020
(March 2019)

Barcelona GSE Working Paper Series

Working Paper nº 1080

Transaction-tax evasion in the housing market

Revised Version January 2020

José G. Montalvo [1]	Amedeo Piolatto[2]	Josep Raya[3]
U. Pompeu Fabra-ICREA,	Autonomous U. Barcelona,	U. Pompeu Fabra,
Research Professor BGSE	BGSE, IEB, MOVE	ESCSE (Tecnocampus)

Acknowledgements: We gratefully acknowledges the financial support of the Spanish Ministry of Economy and Competitiveness (Montalvo: ECO2017-82696P, Piolatto: PGC2018-094348; Raya: ECO2016-78816R), the Government of Catalonia (Montalvo: ICREA-Academia and SGR2017-616; Piolatto: 2017SGR711), the Programa Ramón y Cajal (Piolatto: RYC-2016-19371), the Severo Ochoa Programme for Centres of Excellence in R&D of the Spanish Ministry of Economy and Competitiveness (Montalvo and Piolatto: SEV-2015-0563) and the Barcelona GSE.

[1]Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona. Tel.: +34 935 422 509. E-mail: jose.garcia-montalvo@upf.edu

[2][Corresponding Author] Autonomous U. Barcelona (UAB), Bellaterra Campus, 08193, Barcelona.
Tel.: Phone: +34 935 868 493 E-mail: amedeo.piolatto@uab.es

[3]Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona. Tel.: +34 931 696 501. E-mail: josep.raya@upf.edu

Abstract

We model the behaviour of a mortgagor considering to evade the real estate transfer tax. We build an observable measure of over-appraisal that is inversely related with tax evasion and conclude that the tax authority could focus auditing efforts on low-appraisal transactions. We include 'behavioural' components (shame and stigma) allowing to introduce buyers' and societal characteristics that explain individual and idiosyncratic variations.

Our empirical analysis confirms the predictions using a unique database for Spain, where we directly observe: real payment, value declared to the authority, appraisal, buyers' educational level and local levels of corruption and trust.

Keywords: *transfer tax, tax evasion, housing market, mortgage, appraisal, Loan-To-Value, tax-morale*

JEL classification: G21, H26, R21

1 Introduction

Transfer taxes (i.e. taxes on real property transactions) are common in most OECD countries and yet they remain understudied (Best & Kleven 2018). More generally, the empirical literature on tax evasion has been facing the challenge of obtaining reliable data.[1] Difficulty understanding and observing fraud has obvious consequences on the effectiveness of audits performed by tax authorities. Our paper aims to help fill these gaps by focusing on home buyers' strategic behaviour, whereby we test our prediction and provide some policy recommendations.

To this end, we present a model where a mortgagor decides their housing expenditure, together with the share of the latter they declare to the tax authority. The model embeds elements of behavioural economics identified by the most recent literature on tax evasion as potentially crucial to explaining tax evasion decisions. In particular, we decompose the behavioural component in what we denote as 'stigma', reflecting the unease that an agent may feel when others become aware of a fraudulent behaviour (hence, stigma is only suffered when being caught cheating) and, in parallel, we also include 'shame', which corresponds here to the feeling of guilt that an agent may suffer (regardless of whether their fraudulent behaviour is discovered). Both elements depend on societal characteristics (e.g. norms, trust or social capital). Furthermore, shame may vary with individual attributes.

The model uncovers the relation between tax evasion, access to cash (or other untraceable payment systems) and housing declared over-appraisal (that is, the ratio between declared appraisal and declared transaction).[2] In particular, when agents are liquidity-constrained, potential mortgagors and financial institutions have an incentive to collude and inflate the appraisal that is declared. Typically, those same potential mortgagors are unlikely to fraud.[3] Our result has an interesting policy implication: declared over-appraisal is observable, hence the tax authority could use it as a screening device to determine which transactions to audit.

Using a novel dataset, which includes second-hand private housing transactions that occurred in Spain between 2005 and 2011, we test the model's results empirically. Spain is a particularly interesting environment to study:

[1] Alm (2012), Scheneider & Enste (2000), Slemrod & Yitzhaki (2002), Esteller-Moré et al. (2018) provide extended surveys of the rich literature. Slemrod & Weber (2012) presents an analysis of the limits to the empirical study of tax evasion.

[2] Financial institutions are required to provide an appraisal, while buyers declare the value of the transaction. Both declared values may differ from the real one: for instance, appraisers may inflate the value of the real estate and expand its scope by adding items such as appliances, transaction or other costs.

[3] The part of the payment that remains undeclared in case of fraud cannot be financed through the mortgage, hence evading the tax requires access to liquidity.

while transfer taxes may be negligible in some countries, in Spain they amount to about 10% of the declared price: stakes are definitely high.[4] The data at our disposal is unique in that it includes both the real transaction price and that declared to the tax authority. Hence, we are able to observe the level of tax evasion without noise and to identify several strong patterns. Furthermore, for a subset of transactions, we also observe the socioeconomic characteristics of the buyer and some information about their mortgage. In particular, we detect a very robust negative effect of over-appraisal on tax evasion, as predicted by our theoretical model. We are also able to identify sources of heterogeneity in tax evasion both at the individual and the geographical level. We show that tax evasion decreases as the buyer's level of education rises while it varies depending on the local level of law compliance and trust, measured using different indicators of law enforcement.[5]

Our paper is related to the literature on appraisal, which suggests that over-appraisal was a generalised practice during the real estate bubble of the mid-2000s both in the U.S. and in Spain.[6] Nakamura et al. (2010) suggests that appraisals were subject to an upward bias, such that borrowers were able to obtain larger mortgages, driving excessively risky mortgage loans.[7] The (formally independent) appraiser should value homes objectively. However, appraisers' incentives were distorted in that their clients (money lenders) were often pressuring them to overstate the value of the property.[8] Analogously, in Spain during the housing boom, most agents un-

[4]For more on the transfer tax and the frictions it may generate, see Bradley (2018), Fritzsche & Vandrei (2019), Kopczuk & Munroe (2015) and the references therein.

[5]In our model, those factors enter through shame and stigma and our estimates are consistent with the model predictions. However, alternative explanations may exist. For instance, as kindly suggested by one referee, highly-educated buyers may have higher financial sophistication (Amromin et al. 2018), be more risk-averse, or work in industries where the pecuniary costs of being caught are significantly higher (e.g., law, auditing). Unfortunately, we don't have enough data to test the alternative hypothesis.

[6]In the U.S. the appraised price is (weakly) above the transaction price more than 95% of the time and an increase in inflated transactions was observed between 2000 and 2006 (Cho & Megbolugbe 1996, Loebs 2005, Nakamura et al. 2010, Ben-David 2011). The figures in the Spanish market are even higher (Montalvo & Raya 2012, Akin et al. 2014, Montalvo & Raya 2018). The institutional setting may play a crucial role relative to both the accuracy of appraisal and incentives to evade the transfer tax. The U.K. possibly represents the most extreme case documented in the literature in terms of compliance: appraisal tends to reflect the true value of the property (Cloyne et al. 2019), and evasion of the *Stamp Duty Land Tax* is minimal (Best & Kleven 2018).

[7]The underlying mechanism was the belief that housing prices would continue to grow strongly, reducing the risk of default. Appraisal prices lost validity as a risk assessment of the mortgage loan and gained validity as an element to be used for mortgage lending, since a higher appraisal price reduced the LTV ratio. LaCour-Little & Malpezzi (2003) finds a positive association between the quality of appraisals and mortgage defaults. Lang & Nakamura (1993) notes that, in this case, the bank would require a larger down-payment.

[8]Freybote et al. (2014) suggests that appraisers were influenced in their valuation. Although in the U.S., the deviation of the price from the real economic value was relatively

derestimated the risks of granting overly generous mortgages, assuming that house prices would grow without limits. Financial institutions were prone to open the market to borrowers with financial constraints. Meanwhile, appraisers were encouraged to upward bias their valuations, in turn used to produce artificially low LTVs, which ostensibly kept the credit risk of the mortgage portfolio under control (Montalvo & Raya 2018). Over-appraisal in Spain reached as high as 29% (Montalvo & Raya 2012), explained in part by the additional perverse incentive that more than half of the appraisals were performed by companies directly owned by financial institutions.

The theoretical model is in line with the long-standing literature that follows the seminal papers of Allingham & Sandmo (1972) and Yitzhaki (1974), where evasion has been modelled as a decision made under risk by expected utility maximising agents. The probability of being audited may depend both on the level of evasion and other idiosyncratic characteristics.[9] However, standard models of tax evasion have failed to explain certain empirical regularities.[10] Behavioural economists have consequently augmented the standard model in different ways.[11] The introduction of pro-social behaviours, the 'warm-glow effect' and feelings of stigma have proved extremely helpful in efforts to reconcile theoretical predictions and data.[12] Our model, as mentioned, follows this approach and includes both stigma and shame, which are meant to account for the different possible 'behavioural' components.

Studying the Spanish case is particularly interesting for at least three reasons. First, the empirical literature places Spain amongst the European Union countries with the highest levels of tax evasion, with estimates that range between approximately 20 to 25% of the GDP (Sardá 2014, Scheneider 2005, Medina & Scheneider 2017). Second, across the EU, urban development and construction are sectors where corruption vulnerabilities are usually high (Commission 2014). Fraud has been closely related to the housing market, particularly during the boom years. Perhaps the most common form of tax evasion in the housing market in Spain is under-declaring the

small (6,6% in Ben-David 2011).

[9] We abstract from the analysis of how the tax authority optimally sets the probability of audit. For more on this, see Reinganum & Wilde (1985), Macho-Stadler & Pérez-Castrillo (1997), Chander & Wilde (1998), Di Porto et al. (2013), Piolatto & Trotin (2016).

[10] Third-party reporting or specific institutional settings may significantly reduce the opportunity to evade of agents and, hence, explain the low level of tax evasion in some specific contexts (Kleven et al. 2011, Best & Kleven 2018). This is clearly not the case in the context that we analyse.

[11] A broad literature has developed around the idea of agents who follow the tenets of prospect theory. See, for example, Bruhin et al. (2010), Alm (2012), Hashimzade et al. (2013), Engström et al. (2015), Piolatto & Rablen (2017).

[12] Such additions may include aspects such as stigma (Gordon 1989, Kim 2003), social norms (Traxler 2010), intrinsic motivation like duty or tax morale (Dwenger et al. 2016), equity, fairness or trust (Falkinger 1995, Schildberg-Hörisch & Strassmair 2012).

purchase price to the tax authority. In this way, buyers reduce the burden of the real estate transfer tax, while sellers pay less taxes on capital gains. Finally, the strong ties between financial institutions and appraising firms and the volatility of the economy during the analysed period potentially leave more room for variation and thus allow to better identify different behavioural patterns.

The remainder of the paper is organised as follows. Section 2 presents the theoretical model that explains the evasion of the real estate transfer tax. The model's predictions are tested in Section 3. We begin by describing the institutional setting in Section 3.1, then present the data in Section 3.2, followed by our results in Section 3.3. Finally, Section 4 concludes. Proofs can be found in Appendix A, while complementary tables and robustness checks are included in Appendix B.

2 Model

We provide a stylised model where our focus is on the behaviour of the prospective buyer and their decision to (partially) evade the transfer tax. Notice that we are disregarding agents that are able to buy without borrowing. From a theoretical perspective, those buyers should behave according to the tenets of standard tax evasion theory (see Allingham & Sandmo 1972). Besides, those buyers do not appear in our data. The reader should bear in mind that such group of buyer is usually considered as the more likely to commit fraud, hence our fraud estimates should be interpreted as a lower bound.

Although agents other than mortgagors are involved (the tax agency, the mortgagee, the seller), their behaviour is introduced in a very reduced form, in order to keep the model tractable and to highlight a channel in the buyer behaviour that we then successfully test in the empirical analysis. While building the model, we had in mind the functioning of the real estate market in Spain, for which we had access to the data; however, we believe that our stylised model can be applied to many countries and, indeed, results are in line with previous studies that use data from other countries, including the U.S.

The timing of the process that we have in mind is the following:

1. Seller and buyer agree to carry on the transaction at a price H. The terms of the agreement include the amount that is declared (H^d) and the one paid cash (H^u). The agreement is conditional on the financial institution granting a loan of amount B.

2. The buyer asks the financial institution to grant a mortgage. The mortgagor has to inform the mortgagee of the (declared) transaction price H^d and they have to specify the size of the loan that they need to complete the transaction. By law, the bank must appraise the property. Denote by V the true appraisal and by $\hat{V}$ the appraisal made public: the loan cannot exceed a share α of the appraisal $\hat{V}$.[13] If the condition is not binding, the bank could declare the true value V, yet banks tend to be conservative and declare a smaller value $\hat{V} \in \left[\frac{B}{\alpha}, V\right]$. If the constraint is binding ($B > \alpha V$), the bank can either deny the mortgage and loose the client or they can accommodate and inflate V to make the transaction possible, in which case $\hat{V} = \frac{B}{\alpha}$.[14] We assume that mortgagee and mortgagor collude to guarantee that the transaction occurs. To simplify the analysis, we assume that the bank keeps $\hat{V}$ as low as possible,[15] hence, the bank sets $\hat{V} = \frac{B}{\alpha}$ even when the constraint is not binding.[16]

3. Once the buyer obtains a promise from the seller and from the mortgagee, all parts reunite in front of a public notary where, simultaneously, the declared transaction value H^d is registered (and the corresponding transaction tax is paid), the mortgage for an amount of B is signed and H is paid to the seller: the financial institution wires H^d while the buyer pays the off-the-books amount H^u.

The advantage for the buyer of paying part of the amount off-the-books is that they save the transaction tax.[17] For the fraud to occur, the buyer needs some liquid assets (for instance, cash or crypto-currencies). The tax authority investigates fraud: we take the level of enforcement effort e as exogenous. Yet, the probability of getting caught depends on the total amount that is evaded: collecting large amounts of liquidity generates suspicions and the probability of being noticed is increasing in the total amount of accumulated liquidity.[18] Therefore, we assume that the probability of being investigated

[13] The IMF assessed the impact of macroprudential policies in 46 countries and showed that by far the most common measure was a cap on the amount people could borrow as a multiple of the appraisal value of the property. More than half of the countries considered in the study imposed caps on LTV ratios. The objective of this macroprudential policies is to prevent money lenders to assume too large risks and endanger the financial system.

[14] The literature shows that financial institutions and buyers often put pressure on the appraiser to inflate the valuation and make the transaction possible. In Spain, this is facilitated by that often the financial institution owns the appraising firm.

[15] This is compatible with the fact that, everything else equal, financial institutions usually have a mild preference for setting the appraisal as low as possible.

[16] We will see later that this, if anything, implies that our empirical estimates represent a lower bound, because our channels are diluted in the observed data by the fact that some agents don't face a binding constraint. We disregard off-equilibrium cases of potential buyers whose transaction fail to occur because the bank cannot inflate $\hat{V}$ enough.

[17] The transfer tax in Spain is at least as high as 10%.

[18] All transactions are formally clean and occur in front of a public notary. Suspicions

(and caught) is convex in the total amount evaded.

The study of the preliminary bargaining that leads to the agreement on the selling price (period 1 in the timeline) remains beyond the scope or our analysis. Instead, we focus on the decision of the buyer who chooses the amount to spend in housing and the amount of the payment that is hidden.[19]

The ultimate objective of the model is to complement the literature that investigates the relationship between appraisal value ($\hat{V}$), declared transaction (H^d) and tax evasion (H^u). Taking into account that the two former are observable while the latter is not, being able to identify some co-movement of the three variables could be used to screen transactions and decide which ones to investigate. Suppose that we observe that appraisal is significantly larger than the declared transaction: a priori, two opposite stories are plausible. Either the declared value is biased downward or the appraisal has been inflated. The first explanation suggests that over-appraisal is a signal of fraudulent declaration. The alternative claim postulates the opposite: the spread is mostly due to the inflation of appraisal produced by financial institutions that try to please credit-constrained borrowers and a large spread is an indicator of low levels of evasion. While it is possible that the two channels co-exist, our model and empirical analysis will try to unveil the dominant effect. Notice that the latter mechanism can only appear as long as borrowers and financial institutions are willing and able to distort V (something that is largely documented in the literature).

The buyer chooses the amount to spend in housing (the remaining is used to consume a numeraire good) and which share of it is declared. Declaring a lower transaction price reduces the tax burden (*ad valorem* transfer tax) but the buyer faces the risk of being caught and paying a fine and they also bear an additional cost through social stigma and guilt. Liquidity is used to pay the undeclared cost of the property and also to complement the mortgage and pay the share of property that the mortgage cannot cover. We will show that buyers that are more liquidity-constrained use most of their liquidity to complement the mortgage and have little (or nothing) left to use to evade taxes.

The expected utility function of the agent is defined as

$$\mathbb{E}(U) = h(H) + \mathbb{E}(C) - \pi\left(H^u, e\right) s - \mu\left(H^u, \theta, n\right). \tag{1}$$

may be raised by the financial behaviour of the buyer in the period previous to the purchase (e.g. unusual withdraws from the bank) and of the seller after the transaction.

[19] This can be interpreted as the case of a buyer who anticipates the best deal that they can obtain from the financial institution and, with that in mind, makes a take-it or leave-it offer to the seller.

We assume that the utility from housing $h(H)$ is increasing and concave in H ($h'(H) > 0$, $h''(H) < 0$). $\mathbb{E}(C)$ is the expected value of consumption of the numeraire good.[20]

Agents suffer from social stigma s if caught. The (perceived) probability π of getting caught by the tax administration, as previously discussed, is increasing and convex in the amount evaded H^u and increasing in the exogenous enforcement level e ($\pi'_u > 0$, $\pi''_u > 0$ and $\pi'_e > 0$).

Contrary to social stigma, guilt (individual shame) always materialises and is increasing and convex in the amount evaded and increasing in the agent's type (individual characteristics) θ and in how socially unacceptable is to evade n ($\mu'_u > 0$, $\mu''_u > 0$ and $\mu'_\theta > 0$, $\mu'_n > 0$). Finally, $\frac{\partial^2 \mu}{\partial H^u \partial \theta} > 0$ and $\frac{\partial^2 \mu}{\partial H^u \partial n} > 0$. These two assumptions on the crossed derivative are quite natural: agents with more social capital tend to be more respectful of the law and guilt increases in environments where society doesn't tolerate evasion.[21]

The agent has some 'liquid' savings L, where liquidity is interpreted as money that can be hidden from the tax authority. We normalise to 0 all savings that the agent is unable to hide. By construction $H = H^d + H^u$ and $H^u \leq L$.

The agent can borrow an amount B against some value I that can be interpreted as the net present value of future income or some collateral. Then, $\mathbb{E}(C) = I - (1+i)B - \pi f H^u$, where i is the interest rate on borrowing, while f is the fine rate that is paid if caught cheating.

Finally, denoting t as the transfer tax on the declared housing value, restriction $(1+t)H^d + H^u \leq L + B$ guarantees that the agent spends on housing at most all their savings plus borrowing. Since borrowing money is costly, it is never optimal to borrow more than what is needed to purchase the house, therefore we can rewrite the previous restriction as $B = (1+t)H^d + H^u - L$. Using $H = H^d + H^u$, we can rewrite $B = (1+t)H - tH^u - L$.

Notice that, at any interior solution, this model is isomorphic to a two-period model in which the agent in period 1 borrows from period 2 and purchases the house, while in period 2 they pay back the debt and consume the numeraire good.[22]

[20] The monetary component in the utility function is linear, hence, our agent is risk neutral. The main reason for that is mathematical tractability. However, we believe that introducing some risk-aversion in this model should only reduce evasion and smooth results, but it should not have consequences on the mechanism.

[21] The empirical analysis provides support for these assumptions.

[22] While the two-period model is more intuitive, the one-period setting is more tractable. Either way, we are implicitly assuming that the house is kept forever and that the utility from its consumption corresponds to the net present (continuation) utility. Alternatively, one could consider that the property is sold eventually, and its value is consumed.

It is common to link the interest rate i to the risk profile of the operation, which depends on the share of the purchase that is financed through debt. This calls for a measure of the loan to value (LTV). One convenient way to do that is to set the interest rate to be $i\left(\frac{B}{H^d}\right) = i\left(\frac{(1+t)H - tH^u - L}{H - H^u}\right)$, with $i' > 0$ and $i'' > 0$.[23]

The maximisation problem of the agent is then

$$\max_{H,H^u} h(H) + I - (1+i)\left((1+t)H - tH^u - L\right) - \pi(fH^u + s) - \mu, \qquad (2)$$

which yields the first order conditions (FOCs) with respect to H (Eq. 3) and H^u (Eq. 4):

$$h'(H) = i' \frac{(L - H^u)\left((1+t)H - tH^u - L\right)}{(H - H^u)^2} + (1+i)(1+t) \qquad (3)$$

$$(1+i)t = i' \frac{(H - L)\left((1+t)H - tH^u - L\right)}{(H - H^u)^2} + \pi'_u(fH^u + s) + \pi f + \mu'_u \qquad (4)$$

The FOCs represent the maximum of the objective function if the problem is well-behaved. The following lemma defines the conditions under which this is the case.

Lemma 1 (Second order conditions)**.** *Let $D(H, H^u)$ denote the determinant of the Hessian matrix and define $\phi = i'' \frac{B}{H^d} + 2i' > 0$ and $\psi = \pi''_u(fH^u + s) + 2\pi'_u f + \mu''_u > 0$.*

Then, $D(H, H^u) = \frac{(L-H^u)^2}{(H-H^u)^3}\phi\psi - h''(H)\left(\frac{(H-L)^2}{(H-H^u)^3}\phi + \psi\right) > 0$ and all the second order conditions (SOCs) are satisfied.

Proof. See Appendix A. □

Eqs. (3) and (4) together define implicitly the optimal level for the two control variables H and H^u. Applying the implicit function theorem on the system of equations, we can study how the parameters of the model influence the control variables. For this, we denote the first order conditions, Eqs. (3) and (4), respectively as $F_1 = 0$ and $F_2 = 0$.

[23] Notice that the bank doesn't observe H, hence they cannot set the interest to be $i\left(\frac{B}{H}\right)$. One anonymous referee correctly pointed out that the institution could use the appraisal value V and have $i\left(\frac{B}{V}\right)$. Obviously, the institution is not willing to use the inflated value $\hat{V}$, which is however the only official appraisal value released. Even assuming that the mortgagee produces an additional unbiased estimate of V for internal purposes, it would seem risky to use it but indeed the institution could take it into account when setting i. Using H^d seems safer for the institution. As for the model, using H^d has two advantages: we don't have to assume the underlying relationship between V and H and we avoid introducing a further channel (through i) that would reinforce our results but would make it hard to then disentangle the part that goes through i from the remaining effect. Also notice that our results are robust to the use of H instead of H^u.

We start by looking at the impact of liquid savings L and obtain that

$$\frac{\partial H}{\partial L} = \frac{\frac{\partial F_1}{\partial H^u}\frac{\partial F_2}{\partial L} - \frac{\partial F_1}{\partial L}\frac{\partial F_2}{\partial H^u}}{D(H, H^u)} \tag{5}$$

and

$$\frac{\partial H^u}{\partial L} = \frac{-\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial L} + \frac{\partial F_1}{\partial L}\frac{\partial F_2}{\partial H}}{D(H, H^u)}. \tag{6}$$

The previous equations simplify to

$$\frac{\partial H}{\partial L} = \frac{(L - H^u)(H - H^u)\phi\psi}{D(H, H^u)} > 0 \tag{7}$$

and

$$\frac{\partial H^u}{\partial L} = \frac{-h''(H)}{D(H, H^u)} > 0 \tag{8}$$

Our empirical analysis confirms the result that an increase in liquidity, as expected, leads unequivocally to an increase in both house spending and evasion.

The tax authority, we will see later, may have an interest in combining the information on the transaction provided by the buyer with the one on the value of the property (appraisal) provided by the financial institution. Remember that V represents the real appraisal but that the financial institution releases instead $\hat{V}$. We also define $\tilde{V} = \frac{\hat{V}}{H^d}$, which is the only possible proxy for over-appraisal and takes values above 1 when the financial institution appraises the property higher than its declared value.

It is interesting to use over-appraisal for several reasons. First of all, there is an empirical literature showing a tendency to upward bias appraisals.. Second, $\hat{V}$ and H^d are publicly observable by the tax authority (and, hence over-appraisal too). Should we be able to identify a (possibly spurious) link between over-appraisal and evasion, the tax authority could use this observable as a proxy to identify cases where evasion is more likely to occur.

As previously discussed, $\hat{V}$ is often (by law) used to set an upper limit to the amount that can be borrowed. The financial institution has an incentive to distort upward the value V when the constraint is binding (to please constrained borrowers) and downward otherwise (to please share-holders). Assume that the financial institution can lend at most a percentage α of $\hat{V}$, then the institution will set $\hat{V} = \frac{B}{\alpha}$, hence $\hat{V} = \frac{(1+t)H - tH^u - L}{\alpha}$. Such practice is documented in the literature (see the introduction) and we also observe it in our dataset.[24] Then, α becomes a measure of how much an agent will

[24]We assume, for the sake of tractability, that the institution accommodates for the sake of completing the transaction and that the buyer doesn't pay an extra cost for manipulating V, however, we can imagine that financial institutions charge for it.

be allowed to borrow, which may depend on the legal setting, on individual characteristics and also possibly on some exogenous macroeconomic factors (e.g. GDP or unemployment).

We can now study how overappraisal is linked to liquidity.

$$\frac{\partial \tilde{V}}{\partial L} = \frac{(H - H^u)h''(H)}{D(H, H^u)} \left(L(H - H^u)^2 + (H - L)^2\phi + (H - H^u)^3\psi \right) < 0 \tag{9}$$

The following proposition puts together the results on the impact of a change in L, leading to our first policy implication and empirical question.

Proposition 1. *An increase of the liquid savings L induces an increase in the amount of both undeclared housing, $\frac{\partial H^u}{\partial L} > 0$, and total housing, $\frac{\partial H}{\partial L} > 0$, together with a decrease in the observed over-appraisal, $\frac{\partial \tilde{V}}{\partial L} < 0$. Furthermore, not surprisingly, an increase in audit probabilities π induces a decrease in the amount evaded: $\frac{\partial H^u}{\partial \pi} > 0$.*

Proof. See Appendix A. □

Overappraisal is increasing both in total housing ($\frac{\partial \tilde{V}}{\partial H} > 0$) and in the amount evaded ($\frac{\partial \tilde{V}}{\partial H^u} > 0$). Because $\frac{\partial H}{\partial L} > 0$ and $\frac{\partial H^u}{\partial L} > 0$, it follows that the indirect effect of L on $\tilde{V}$ is positive. Yet, the total (direct plus indirect) effect of L on $\tilde{V}$ is negative: $\frac{\partial \tilde{V}}{\partial L} < 0$. Corollaries 1 and 2 summarise the consequences of that.

Corollary 1. *If $\frac{\partial H^u}{\partial L} > 0$, $\frac{\partial \tilde{V}}{\partial H^u} > 0$ and $\frac{\partial \tilde{V}}{\partial L} < 0$, an increase in the amount of hidden-savings L (which is usually hard to detect) has opposite effects on tax evasion and over-appraisal. Therefore, the data should show a negative correlation between the level of evasion and over-appraisal.*

Corollary 2. *If $\frac{\partial H}{\partial L} > 0$, $\frac{\partial \tilde{V}}{\partial H} > 0$ and $\frac{\partial \tilde{V}}{\partial L} < 0$, an increase in the amount of hidden-savings L has also opposite effects on total spending and over-appraisal. It must be that housing H and over-appraisal $\tilde{V}$ are negatively correlated ($\frac{\partial H}{\partial \tilde{V}} < 0$). In our empirical analysis, we will use this relationship to support our result that $\frac{\partial H}{\partial L} > 0$.*

Proposition 1 and its corollaries shed light on a relevant question: do we expect any regularity to link overappraisal and tax evasion? If so, should we expect to observe higher levels of overappraisal amongst tax evader or amongst agents that declare honestly? Overappraisal, defined as $\frac{\tilde{V}}{H^d}$, is observable. On the opposite, tax evasion is not directly observable. Can we use this correlation as a first screening to select transactions that are more likely to be fraudolent?

We believe that the correlation discussed in Corollary 1 and Corollary 2, that

we will empirically test, can be used as an indicator of a possible fraud.[25] The logic behind Proposition 1 and Corollary 1 is straightforward: available liquid savings set an upper bound on how much an agent is able to evade, net of the down-payment of at least $(1-\alpha)\hat{V}$. Hence, any lack of liquidity has an impact on the amount that an agent can borrow. Agents with less liquidity than the down-payment must either buy a cheaper house or convince the financial institution to inflate $\hat{V}$. As a consequence, an agent with access to liquid savings can afford to evade and doesn't need to push for over-appraisal, whereas agents with less liquid savings are unable to evade and, eventually, they may even ask for an over-appraisal. It consequently follows that the level of evasion and over-appraisal are negatively correlated.

Proposition 1 crucially depend on that V doesn't reflect the objective value of the house but it is, instead, inflated by short-sighted financial institutions that try to please their clients, in order to attract as many as possible of them, without screening them based on their default probability. This makes it possible to observe inflated values $\hat{V}$, something that is supported both by our empirical analysis for Spain and the literature on other countries (see the introduction). At least in the case of Spain, financial institutions cared very little about clients screening and the appraisal value was, de facto, negotiated with the client.

Corollary 1 suggests a correlation between tax evasion and overappraisal (the latter being observable by the tax authority). This has a clear and important policy implication: tax authorities should focus their efforts on preventing evasion by auditing transactions that show low levels of over-appraisal. Of course, such strategy by the tax authority, if anticipated by tax-payers, would lead to a possible reaction that would limit the effectiveness of such policy. In particular, agents may inflate V in order to reduce the chances of audit. For the sake of tractability, in the model we assumed that there are no costs of pushing the financial institution to inflate $\hat{V}$. However, it is plausible that lenders will attach some costs to it. In which case, the borrower would face a trade-off between decreasing the chances of being audited and the cost of inflating $\hat{V}$. In that case, we would still have that overappraisal is negatively correlated with evasion, even if at a lower extent.

In the literature on tax evasion, there has appeared a growing interest in the role of behavioural components, such as stigma, in the decision to evade taxes (see the introduction for references). In our model, we decompose such behavioural features into two components: "stigma" corresponds to the dis-utility that an agent suffers when they are caught cheating.[26] We assume that stigma is a binary variable, in the sense that people will mostly

[25] Admittedly, in the long run buyers may react to the use of such a screening device and may reduce the power of the instrument.

[26] The idea being that society cannot stigmatise a tax evader if the latter is not caught.

remember the scandal but not the details. "Guilt" or "moral shame" corresponds to the disutility that an individual feels when they cheat, regardless of whether they are caught: agents are always aware of having cheated and hence guilt is present regardless of whether cheating becomes public. Shame goes with society and its culture and, as such, depends on the level of morality of the environment. It is, however, agent-specific, and thus also depends on individual characteristics (e.g. education). Shame, being an individual feeling, depends on the level of evasion: an individual's guilt will grow with the amount evaded.[27]

Proposition 2. *In this model, at any interior solution, stigma plays a role on the level evaded only as long as the probability of getting caught depends on the amount evaded. When* $\frac{\partial \pi}{\partial H^u} = 0$ *stigma may deter evasion (corner solution) but it does not affect the level of evasion, conditional on evading. As expected, the level of evasion is negatively affected by stigma:* $\frac{\partial H^u}{\partial s} < 0$.

Proof. See Appendix A. □

Proposition 2 suggests that the role of stigma, as defined here, is limited to when the probability of being caught depends on the level of evasion. The intuition is that stigma only matters if one gets caught. If the probability of being caught is orthogonal to the agent's behaviour, then stigma will only determine the extensive margin (the probability of evading) but not the intensive margin (how much to evade).

The decision to evade taxes is certainly affected by both the surrounding environment (e.g. the level of tax enforcement and the moral code of a society) and by individual characteristics (e.g. the level of education). In this model, the environment may enter through two channels (on top of stigma, which has already been discussed): it may directly affect the probability of being caught, through the level of enforcement e, or it may affect the level of shame, through n. Individual characteristics θ, instead only affect the model through shame. Proposition 3 discusses both of these elements.

Proposition 3. *The housing value that is hidden from the tax authority may vary locally, depending on the level of enforcement* e *and, through shame, on how much tax evasion is socially disapproved of* n. *As one may expect, both factors negatively affect the level of evasion:* $\frac{\partial H^u}{\partial e} < 0$ *and* $\frac{\partial H^u}{\partial n} < 0$. *Furthermore, law compliance varies at the individual level, through shame, due to individual characteristics* θ *(such as education), so that* $\frac{\partial H^u}{\partial \theta} < 0$.

Proof. See Appendix A. □

[27] The assumption of shame being continuous while stigma being binary can be easily relaxed.

Proposition 3 confirms that in this model society has an impact on the individuals' decision to evade. Indeed, the perceived enforcement affects the decision of a rational individual. Furthermore, living in a society that is less tolerant towards illegal behaviour produces more shame, which reduces the level of evasion. The empirical analysis confirms this result, showing that more evasion is observed in regions with higher levels of corruption and where social values are lower. Individual characteristics also matter: when the parameter θ increases, the level of evasion decreases. We also test this in our empirical analysis, where we observe that more educated agents are less prone to evasion.

The empirical analysis allows us to (coarsely) relate evasion with changes in macro-economic factors such as GDP or unemployment. While such parameters are not directly present in our theoretical model, we could expect some of our variables to be affected by them. In particular, a decrease in GDP or an increase in unemployment may affect, on average, our variable L (Ganong & Noel 2019). Should this be the case, we would expect a decrease in GDP to reduce the level of evasion H^u. Macro-economic factors such as GDP and unemployment may also have an impact on the availability of credit, which in turn could affect the variable α (the share of the valuation that an agent can borrow). Notice that $\tilde{V} = \frac{B}{\alpha H^d}$, thus, $\frac{\partial \tilde{V}}{\partial \alpha} = -\frac{B}{\alpha^2 H^d} < 0$. Therefore, since α is negatively correlated with over-appraisal, a credit restriction would tend to increase the tendency to over-appraise properties.

The theoretical model only focuses on interior solutions. Obviously, corner solutions occur and may lead either not to evade at all, or even not to buy. We focus on interior solutions for two reasons: on the one side, the mechanism at play that we highlight somehow looses interest when we are at a corner solution, where "nothing happens". In the data, we observe that a fair share of the population does not cheat, yet the theoretical analysis of the corner solution would bring little insight. On the other hand, a serious and full analysis of corner solutions would call for a much more sophisticated (general equilibrium) model, which should include the outside option of borrowing, a full specification of the housing market (demand and supply) and also of the financial markets. Such a model would, most likely, loose its tractability, while our model is able to generate clean predictions that, as we will see in Section 3, is fully consistent with the data that we have.[28]

[28]Gete & Reher (2016) provides a theoretical model that could be considered to be complement to ours, where they endogenously treat financial markets and highlight the connection between Loan-to-value, housing tenure choice, mortgage markets and renting. Their focus is on financial markets and default. Regrettably, it would be nearly impracticable to augment their model to account for tax evasion.

3 Empirical analysis

In this section, we test the previous results using a novel dataset on about 1,500 real estate transactions that occurred in Spain during the period 2005-2011. The dataset is particularly unique in that it includes both the value declared to the tax authority and the amount effectively paid. In what follows, we begin by presenting the institutional framework, we then describe our data and finally, report the results.

3.1 Institutional framework

Over the first decade of the twenty-first century, Spain experienced one of the largest housing booms of any developed economy.[29] The construction sector alone was responsible for approximately 20% of the GDP growth. This housing boom led to a housing price bubble (housing prices tripled between 1998 and 2008) that began to burst in 2008. At the time, an average of approximately 1.1 million mortgages per year were approved.[30]

The lending market was extremely competitive. Spanish financial institutions offered the lowest mortgage rates of the Euro area. In fact, over the 2003-06 period, the average mortgage rate in the Euro zone was 21% higher than in Spain. Financial institutions attempted to compensate for the reduced per-mortgage margin with an increasing number of transactions, which contributed to the sharp increase in the number of mortgages. The excessive dependence of the Spanish economy on the real estate market, together with loose credit standards (Akin et al. 2014), largely explain why the financial crisis hit Spain more severely than most other economies.

The attempt to increase the number of transactions led to a softening of credit standards. Yet financial institutions were constrained by internal policies on the LTV ratio. These constraints were relaxed by pushing appraisers to over-value properties whenever the borrower did not have sufficient resources for the down-payment or preferred to borrow more for a different reason. Montalvo & Raya (2018) find evidence consistent with financial intermediaries encouraging appraisal firms, most of them owned by banks themselves, to introduce an upward-bias in their valuations by approximately 30% to meet the LTV recommendations, so as to be able to

[29] During this period, more dwellings were built in Spain than in Germany, France and Italy put together. According to the official statistics of the Department of Public Works, housing initiations reached as high as 860,000 dwellings in 2006.

[30] Note that there were approximately 15.5 million households in Spain. Over the considered period, the average number of transactions realised per year and region was approximately 20,000.

use them as collateral for covered bonds (the limit LTV for this pool of collaterals is 80%) and to reduce their capital requirements. Indeed, 40% of mortgages in the researchers' sample are bunched at the LTV threshold.

It is important to note that Spain has only been a democracy since 1975. Young democracies are particularly vulnerable to illegal activities (Treisman 2000) and it is well known that different kinds of criminal behaviour, from tax evasion to black markets and corruption, are positively correlated (Fortin et al. 2000). It is perhaps not surprising then that Spain ranks third in Europe in terms of the percentage of citizens (95%) who believe that corruption is widespread (Commission 2014). Various cases of corruption have, in fact, recently been uncovered, many of which relate to the real estate sector and involve politicians at all levels.[31] Real estate transfer tax in Spain is at the order of magnitude of 10% of the declared value.[32] The most common way to reduce the tax burden related to real estate transactions is to under-declare the transaction value to the tax authority. The seller may also occasionally benefit if the sale is classified as speculative and, therefore, subject to the capital gain tax.

3.2 Our data

Data on either the Spanish housing boom or related aspects is scant. One of the main reasons is a lack of reliable statistical information on housing values. Indeed, prior to 2007, the Spanish official house price index was computed based on appraisals, which were highly unreliable, as mentioned earlier. Since 2007, the price index has been based on the Property Registry values, that is, the transaction value declared for tax purposes by the buyer. As we will show, for many transactions these prices do not correspond to the actual market price either.

Our dataset is the first to include actual market prices. For one-fourth of the dwellings in the sample, we also have individual characteristics of the mortgagor. This unique dataset was obtained from a real estate intermediary[33] that operates across most Spanish provinces and that also runs its own mortgage brokerage business. The intermediary has a 3-5% market

[31] Corruption and illicit practices are common in urban planning and spatial development in Spanish cities. Benito et al. (2015) cite 676 cases of urban corruption that have been documented in the media. Of the corruption cases that occurred during the period of analysis, some relate to the illicit funding of political parties, or tax fraud and embezzlement by members of the government.

[32] Contrary to the U.K., where the tax rate increases with the value of the property (Best & Kleven 2018), in Spain the tax rate is flat.

[33] We signed a non-disclosure agreement prohibiting the disclosure of the company's name.

proportion of realised sales (depending on the year).[34]

We merged the dataset obtained from the real estate intermediary with information from other sources. Data from the intermediary include the actual transaction price (i.e. the amount effectively paid by the buyer, on which the intermediary computes their fees), and the characteristics of the properties from a random sample of their sales. From the Property Registry (*Registro de la Propiedad*) we obtained information on the amount of the mortgage, the appraisal value and the buying price declared to the tax authority. W also obtained the cadastral reference (a unique identifier for each property) from the cadastre (*catastro*). Matching the cadastral reference, the declared appraisal value and the registry value from different data sources minimises the risks of measurement error or mismatch of the observations. Measurement errors concerning the real spending (H) are also very unlikely: the real estate intermediary that collected (and shared with us) the data charges the buyer proportionally to spending H.[35] The sample period runs from 2005 to 2011. The merged data allow to compute the amount that was not declared for 1,445 transactions of existing housing units (apartments). We refer to this set of data as the 'whole sample'.

For a subset of 430 observations, we were able to merge previous data with information provided by financial intermediaries. We use also this information to double check on the amount of the mortgage and the declared appraisal. Thus, this subset includes, among other variables, individual characteristics of the buyer, such as the number of owners of the property and their respective levels of education. We refer to this subset as the 'sample with individual characteristics'. For these 430 observations, certain financial information (e.g. appraisal prices or the amount of the mortgage) was present in several different datasets. We used such redundant information as a further check of the reliability of the merging process. Table 1 displays the descriptive statistics of our dataset both for the 'whole sample' and for the 'sub-sample with individual characteristics'.[36]

[34]Notice that most of the existing home sales in Spain are sold directly by the owner.

[35]Difficulties matching the data unfortunately caused the loss of some information. We were very strict, requiring an exact match on redundant information from different sources. Furthermore, the Spanish registry is organised in the format of a 'continuous roll': successive owners of a given property are sequentially added on a single document recorded by the original address at the time of building. However, due to many political upheavals (including two dictatorships, the republic and two monarchies), street names have changed several times over the last century. In order to match the data, it was necessary to match the address of the estate at the moment of construction with that when it was sold.

[36]It is worthy to notice that the sub-sample is not a random selection of the larger sample. Indeed, we have socio-economic data for buyers who financed their purchase through the financial department of the real estate agency. These clients are likely to be more liquidity constrained that the average and may have experienced problems getting a mortgage directly from banks, which are always cheaper than the financial subsidiary of

Nearly half of the transactions included some undeclared money, with a mean value for the percentage of undeclared money of 7.64% (13,847 euros). Conditional on fraudulent behaviour, this percentage rises to 15.1% (27,409 euros). Figure 1 presents a histogram of the share of undeclared money (H^u/H), conditional on evasion. The percentage of undeclared money over the actual transaction price was lower than 20% in 76.03% of the fraudulent transactions.

For our main measure of corruption, we identified municipalities where politicians in power have been accused of corrupt behaviour, following the definition of corruption in Fernández-Vázquez et al. (2016).[37] We combined several databases on corruption scandals reported in local, regional and national newspapers, as well as in reports written by non-governmental organisations, think tanks and public advocacy groups. We focus on the 26 municipalities for which we have 10 or more observations, resulting in a sample size of 1,233 observations. In 14 municipalities, we identified at least one case of corruption.[38]

We also construct a measure of the shadow economy. For that purpose, we use data from Sardá (2014)[39] on the mean shadow economy in Spain from 2004 to 2011 at the province level, merging the latter with our dataset. For 1,432 of the observations in our 'whole sample',[40] we use the estimated percentage of the shadow economy at the province level over the 2004-2011 period.[41]

The period we analyse saw both a bubble and a burst in the housing market, which also had an impact on GDP, unemployment and the economy in general. Table 5 in Appendix B shows the evolution of tax evasion over time. The share of fraudulent transactions steadily decreased over the considered

real estate agencies. Hence, over-appraisal should be more pronounced than in the whole sample.

[37] In particular, our corruption dummy takes value 1 when four conditions are simultaneously met at the municipal level: 1) the mayor or another member of the municipal executive branch is involved in the scandal; 2) the accusation involves criminal charges related to corruption and abuse of public office; 3) charges are brought by a non-partisan actor and 4) claims about misbehaviour were in the press between 2004 and 2010.

[38] As a robustness test, in Appendix B we use two alternative measures of corruption. Those data sources are described together with the corresponding tables.

[39] To measure the size and development of the shadow economy, we adopt a 'Multiple Indicators Multiple Causes' (MIMIC) approach (Weck-Hanneman & Frey 1985), a special case of the general LISREL model. A MIMIC model consists of two parts, the structural equation and the measurement equation system. The structural model examines the relationships between the latent variable (output of the shadow economy) and the causes, while the measurement model links indicators and the latent variable.

[40] Sardá (2014) does not report the estimation of the shadow economy for Vizcaya Province.

[41] The mean value of the shadow economy in Spain during these years is 19.63%. The maximum value is 23.3% (Zamora), while the minimum is 13.8% (Madrid).

period. However, note that around 2008 and conditional on fraud, the share that remained undeclared begins to increase. One possible interpretation, consistent with the discussion at the end of Section 2, is self-selection. When the crisis hit, many citizens were impoverished. The decrease in GDP and the increase in unemployment resulted in less buyers having some liquid savings to use for purchasing. The probability of having sufficient savings to make any cash side-payments decreased. Meanwhile, the decline in housing prices that followed the bubble burst meant that agents who had access to liquid savings could use them to pay a larger share of the total value. To this regard, Section 3.3 shows how the share of fraudulent transactions is decreasing in unemployment, while the share that is undeclared (conditional on fraud) is increasing in unemployment.

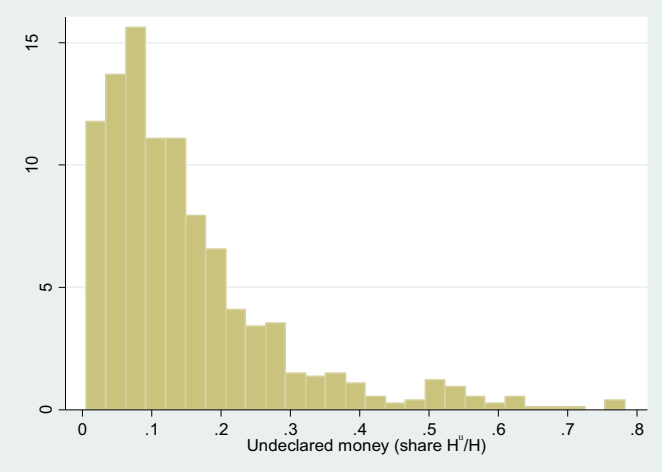

Figure 1: Histogram: share of undeclared money, conditional on evading

Table 5 in Appendix B also displays how agents' behaviour is heterogeneous across the country. We immediately observe a large spread both in terms of the share of illegal transactions carried out (up to 23 percentage points) and in terms of the share of the price that remains undeclared (up to 9 percentage points).

3.3 Results

In this section, we test the predictions of the theoretical model. Tables 2 and 3 include the result of the main regressions using respectively the whole sample and the sample with individual characteristics. For each specification, we estimate a Logit model for the determinants of the probability of a fraudulent transaction, that is $Pr(H^d < H)$, as well as a Tobit model for the

	Whole sample		Sample with individual charact.	
	Mean	Std. Dev.	Mean	Std. Dev.
# Fraudulent transactions	0.51	0.50	0.53	0.50
Undeclared share (H^u/H)	0.08	0.12	0.09	0.15
Undeclared amount (H^u)	13.847	25.329	16.524	31.172
Overappraisal ($\tilde{V}$)	1.29	0.25	1.31	0.25
Spread			0.86	0.45
Year				
2006	0.27	0.44	0.12	0.33
2007	0.18	0.39	0.20	0.40
2008	0.14	0.34	0.12	0.32
2009	0.18	0.38	0.26	0.44
2010	0.18	0.39	0.30	0.46
2011	0.04	0.15	-	-
Region				
Andalusia	0.26	0.44	0.30	0.46
Aragon	0.12	0.33	0.09	0.29
Castile La Mancha	0.03	0.17	0.04	0.18
Castile and León	0.02	0.14	0.02	0.14
Catalonia	0.19	0.39	0.13	0.33
Community of Madrid	0.31	0.46	0.36	0.48
Valencian Community	0.06	0.24	0.05	0.22
Others	0.01	0.10	0.01	0.10
Educational level				
Primary			0.45	0.54
Secondary			0.40	0.49
Tertiary			0.15	0.35
Number of mortgagors				
One			0.53	0.55
Two			0.41	0.49
Three or more			0.06	0.24
Labour situation				
Non-Occupied			0.07	0.25
Occupied: private sector			0.73	0.44
Occupied: public sector			0.14	0.34
Self-employed			0.06	0.24
Observations	1,445		430	

Source: Own elaboration.

Table 1: Descriptive statistics.

determinants of the amount of undeclared money (that is, the total transaction value that is hidden from the authority: H^u). In the case of Logit, tables report the estimated coefficients, while marginal effects are reported and discussed through the text. Column (1) includes no fixed effects, while (2) includes both year and regional fixed effects. Column (3) controls for the level of corruption while Column (4) controls for the size of the shadow economy.

Proposition 1, together with Corollaries 1 and 2, predicts that over-appraisal ($\tilde{V}$) and tax evasion (H^u) are inversely related. The same is true for total house spending (H), which is negatively correlated with over-appraisal, which implies that liquid savings (L) and house spending are positively related.

Table 2 confirms the prediction of the theoretical model and shows that tax evasion and over-appraisal are strongly, negatively related in all our specifications.[42] In particular, focusing on specification (2), we observe that an increase in over-appraisal by one point (that is, the appraisal value doubles the transaction price) coincides with a decrease in the probability of fraud of 66.5% and a decrease in the amount that remains undeclared of 26,660 euros. Using the 29% mean over-appraisal in Spain computed in Akin et al. (2014), our results would suggest that liquidity constraints correspond to a reduction of 7,731 euros in the amount that is hidden from the tax authority.[43]

Complementing the discussion following Proposition 1, buyers try to minimise over-appraisal and the amount that they borrow. The use of over-appraisal to increase the amount that can be borrowed is a last recourse for a buyer, used only when they have no alternatives. Over-appraisal becomes a signal of liquidity constraint, which is unlikely to occur for agents who

[42] A linear-logarithmic model would yields qualitatively very similar results (available upon request).

[43] In the case of Table 3, the marginal effect in our specification (2) suggests that we observe a decrease in the probability of fraud by 16.61% when over-appraisal increases by one point. Hence, audit activity is less likely to be effective in discovering evasion. The difference in results between the whole sample and the reduced one conforms with our expectation, as mentioned. We replicated Table 2 with the restricted sample of observations in order to have a better understanding of the difference across sub-samples (table available upon request) and, indeed, we noticed that the magnitude of the effect of over-appraisal is much larger in the restricted sample, while Corruption and Shadow Economy loose, at least partially, their significance. As we argued before, there is a selection bias in the reduced sample: individuals in there have signed a mortgage through the real-estate agency, which suggests that they have a riskier than average borrowing profile. In the reduced sample we have more liquidity constrained agents, which explains why the magnitude of the effect of over-appraisal is larger. Individual characteristics may be correlated with "Corruption" (s) and "Shadow economy" (n), which could explain why these variables loose significance when we control for individual characteristics in the reduced sample.

	(1)		(2)		(3)		(4)	
	Logit	Tobit	Logit	Tobit	Logit	Tobit	Logit	Tobit
Overappraisal ($\tilde{V}$)	-0.799***	-30314.943***	-0.665***	-26660.974***	-0.837***	-32140.544***	-0.739***	-29213.131***
	(0.215)	(4523.444)	(0.246)	(4714.932)	(0.262)	(5456.640)	(0.238)	(4845.255)
Spending (H)	0.000**	0.103***	0.000***	0.158***	0.000	0.119***	0.000	0.103***
	(0.000)	(0.016)	(0.000)	(0.020)	(0.000)	(0.020)	(0.000)	(0.019)
Corruption (s)					0.814***	17872.286***		
					(0.128)	(2671.231)		
Shadow economy (n)							0.044**	1419.329***
							(0.021)	(434.141)
Intercept	0.742**	19834.496***	-1.432**	-5985.385	-0.268	13057.867	-0.644	-7584.225
	(0.344)	(7143.754)	(0.721)	(14121.863)	(0.525)	(10934.438)	(0.626)	(12918.334)
N. Obs.	1,445		1,445		1,233		1,432	
Year F.E.	No		Yes		Yes		Yes	
Region F.E.	No		Yes		No		No	

* p <0.10. ** p <0.05. *** p <0.01

Table 2: Estimated model: whole sample

have liquid savings that can be used for side-payments. Tables 2 and 3 thus confirms the model's prediction and, accordingly, has a strong policy implication. Since declared over-appraisal is much easier to assess and observe than access to liquid savings or fraud, it should be used as an indicator for the likelihood of fraud. In particular, the tax authority should focus their audit efforts on transactions where the appraisal is relatively low.[44] Results in Table 3 reinforce the argument that liquidity constraints matter. Indeed, when the purchase is made by three or more buyers, the probability of fraud increases.

Corollary 2 also suggests a positive relation between liquidity (L) and total spending (H) and, therefore, a negative relation between total spending (H) and over-appraisal ($\tilde{V}$). We observe that the two variables are, in fact, negatively correlated. The left picture in Fig. 2 presents the scatter-plot of the relation between total spending and over-appraisal. Confirming our expectation, correlation is -0.35. The right picture does the same, after removal of the outliers. In such case, correlation is -0.34.

Table 6 in Appendix B summarises the results of the estimation of the bivariate model of overappraisal and underreporting, both for the case of bivariate probit and bivariate tobit. The bivariate model does not imply simultaneity, only that both variables are potentially driven by an unobserved third factor, namely liquid assets. We present the estimations only for the whole sample. Parameter ρ shows the correlation between the unobservables of

[44] The caveats discussed in Section 2 about possible strategic reactions to such a policy by the tax authority hold.

	(1)		(2)		(3)		(4)	
	Logit	Tobit	Logit	Tobit	Logit	Tobit	Logit	Tobit
Overappraisal ($\tilde{V}$)	-1.371***	-55709.517***	-1.423***	-53073.802***	-1.537***	-60034.674***	-1.472***	-57941.671***
	(0.469)	(11126.823)	(0.538)	(11063.466)	(0.551)	(12290.175)	(0.511)	(11479.299)
Spending (H)	0.000**	0.116**	0.000**	0.143**	0.000*	0.139**	0.000*	0.128**
	(0.000)	(0.049)	(0.000)	(0.058)	(0.000)	(0.059)	(0.000)	(0.056)
Corruption (s)					0.334	12130.767**		
					(0.257)	(5691.337)		
Shadow economy (n)							0.034	59.553
							(0.042)	(947.049)
Spread	-0.246	-5470.717	-0.289	-6924.125	-0.152	-3740.194	-0.130	-2689.968
	(0.233)	(5484.936)	(0.262)	(5484.162)	(0.247)	(5619.510)	(0.244)	(5545.172)
Educational level								
Secondary	-0.508**	-12622.429**	-0.543**	-14245.931**	-0.423	-13391.555**	-0.499**	-14835.838***
	(0.231)	(5498.717)	(0.259)	(5509.045)	(0.265)	(5958.282)	(0.251)	(5638.023)
Tertiary	-0.785**	-16110.889**	-0.838**	-17890.235**	-0.833**	-21431.167**	-0.809**	-18077.244**
	(0.336)	(8031.137)	(0.372)	(7942.052)	(0.385)	(8760.025)	(0.363)	(8188.559)
Number of mortgagors								
Two	0.135	8643.709	0.063	7503.354	0.059	9513.963	0.145	9787.412*
	(0.223)	(5332.197)	(0.252)	(5316.324)	(0.261)	(5857.201)	(0.246)	(5503.371)
Three or more	1.139**	9339.788	1.490**	9870.198	1.655**	12862.928	1.611**	13251.708
	(0.546)	(10367.686)	(0.705)	(10352.859)	(0.708)	(10888.759)	(0.698)	(10676.075)
Employment								
Private sector	-0.269	-15706.120	-0.421	-16432.551	-0.350	-16889.565	-0.667	-21585.969**
	(0.409)	(9732.964)	(0.489)	(10129.825)	(0.526)	(11768.959)	(0.465)	(10282.032)
Public sector	-0.138	-12672.932	-0.148	-10428.781	0.015	-8895.414	-0.310	-15383.452
	(0.493)	(11755.980)	(0.563)	(11856.870)	(0.613)	(13827.114)	(0.548)	(12252.472)
Self-employed	0.834	4009.609	0.405	-2559.532	0.663	529.552	0.215	-5562.931
	(0.606)	(13065.837)	(0.677)	(13076.494)	(0.751)	(15543.769)	(0.653)	(13442.981)
Intercept	1.666**	72647.077***	1.129	75604.598***	2.743**	113103.292***	1.557	101634.520***
	(0.841)	(19486.696)	(1.223)	(24901.393)	(1.195)	(25095.229)	(1.376)	(30833.770)
N. Obs.	430		430		351		385	
Year F.E.	No		Yes		Yes		Yes	
Local F.E.	No		Yes (Region)		No		No	

References. Education: Primary; Mortgagors: One; Labour: Non-Occupied
* p <0.10. ** p <0.05. *** p <0.01

Table 3: Estimated model: sample with individual characteristics

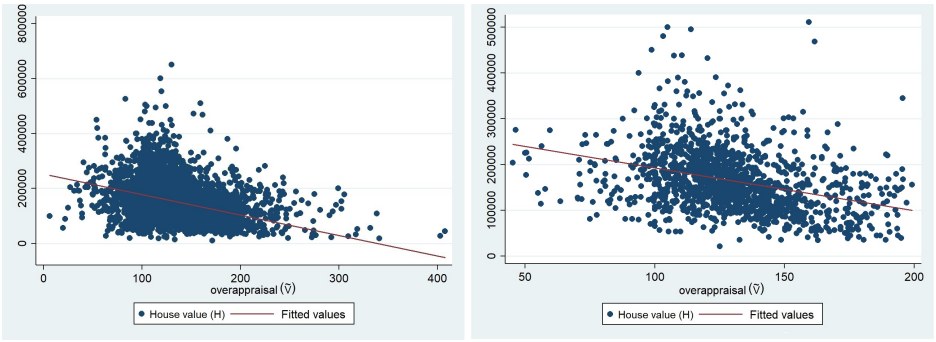

Figure 2: Correlation between Total Spending and Over-Appraisal

both equations, it is significant in all bitobits showing that the unobservable (i.e liquid assets) of the amount of undeclared money and overappraisal are negatively correlated.

3.3.1 Tax-morale: idiosyncratic versus individual differences

Table 5 in Appendix B evidences large differences in evasion across regions.[45] It has been well documented[46] that a social component that involves information, trust, social capital may be responsible for people restraining themselves from acting illegally. To this regard, our data contributes to explain idiosyncratic differences by showing that the environment and social values indeed may explain part of the variance in fraudulent behaviour.

The theoretical model distinguishes between what we (loosely) call stigma and shame. Both of them have a negative impact on tax evasion (propositions 2 and 3). According to our definitions, the difference between these two is that an agent suffers stigma conditional on being caught, whereas shame is a feeling of guilt that is independent of being exposed. Thus while stigma is the result of being judged by others, shame is an individual perception, although it may also be affected by idiosyncratic elements such as societal tolerance of illegal behaviour. Unfortunately, we do not have access to audit

[45] In particular, tax evasion in Andalusia and the Valencian Community is significantly larger then elsewhere, in terms of both the extensive and intensive margin. In the estimation using the sample with individual characteristics, the same is true for the Community of Madrid. The quantitative interpretation of the Logit results originates from marginal effects; for Andalusia and Valencian C., the probability of under-declaring money increases by 0.34 and 0.29 points, respectively. Considering a mean probability of 51%, these effects represent an increase close to 70% and 60%, respectively. In addition, in Andalusia and the Valencian Community, the amount of undeclared money is 28,107 euros and 31,105 euros, respectively.

[46] See, for example, (Alesina & La Ferrara 2000,0, Boffa et al. 2016) and the literature therein.

probabilities and our sample of individual characteristics is also small, we have to rely on imperfect proxies to test our predictions and further analysis is warranted.

In order to better understand whether moral values and the social environment actually affect the amount of fraud observed, we used several indicators of social values, trust in government and feelings of reciprocity. Our argument being twofold: on the one side, indicators of other illegal behaviours are a proxy for the level of law enforcement and, therefore, of audit probability and hence of stigma. On the other side, low social value indicators, together with individual characteristics, define the degree of morality of individuals and, hence, the impact of shame on their decision to commit fraud.

To formally test this argument, as previously discussed, we have constructed a measure of corruption. Results are summarised in specification (3) of Tables 2 and 3. Corruption has a significant and positive effect on the probability of undeclared money and on the percentage of undeclared money with respect to the appraisal price. Using our measure of corruption, the model suggests that dishonesty at the local level significantly increases the probability of engaging in a fraudulent transaction: in our specification (3) of Table 2, the marginal effect suggests that increasing corruption by one point increases the probability of fraud by 20.0%.[47] Moreover, the amount of value that remains undeclared increases by 17,872 euros (considering the estimation using the whole sample). For robustness, Table 7 (in Appendix B) replaces our index of corruption with two specifications of the Global Transparency Index (GTI), published by Transparency International, which measures the level of transparency of public institutions and the 'corruption' variable in the Quality of Government (QoG) data from the Quality of Government Institute.[48] Results are consistent with the previous ones.

Alm et al. (2004) and Alm & Torgler (2006) find a negative correlation between tax morale (shame, in our model) and the size of the shadow economy. In our theoretical framework, the shadow economy would enter through parameter n. In specification (4) of Tables 2 and 3, we then control for the level of shadow economy. Results again confirm our expectations. A larger shadow economy has a positive effect on the probability of underdeclaring the value of the transaction to the tax authority, as well as on the amount of undeclared money. In particular, looking at marginal effects, a rise of one percentage point of the shadow economy increases the probability of fraud by 1.1% and the amount of the transaction price that remains un-

[47] The analysis using the reduced sample is affected by the reduction in the number of observations. Results on corruption loose significance at the usual levels. The marginal effect would suggest that increasing corruption by one point increases the probability of fraud by 8.3%.

[48] For the sake of consistency in results, original data were treated to have all proxies being increasing in corruption.

declared by 1,419 euros.[49] Table 8 (in Appendix B) checks the robustness of the result, replacing our indicator of the shadow economy with the European Social Value (ESV) index, a large-scale, cross-national, longitudinal survey research programme on basic human values.[50]

Proposition 3 also predicts that individual characteristics matter, for they affect the level of shame. Our data include socio-economic information for the subset of agents for whom we have individual characteristics collected by the financing institution. We test the level of evasion for these agents in Table 3, controlling for education and type of employment. While most characteristics in our possession have little explanatory power, education appears to be strongly connected to the level of evasion, both on the extensive and intensive margin. Indeed, the higher the educational attainment, the lower the probability of fraud and the amount of the transaction's value that remains undeclared (20.5% and 17,890 euros respectively). This result is consistent with the prediction of the theoretical model and is in line with theories of pro-social behaviour: better educated citizens are more affected by shame and, therefore, engage in less tax evasion.

Individual characteristics may be correlated with unobservables: more educated agents may be financially more sophisticated (Amromin et al. 2018), have different risk aversion, larger wealth and more liquidity.[51] Using the Spanish Survey of Households Finances (EFF)[52] we found a significant correlation between liquidity and both education (0.02***) and monthly mortgage payments (0.17***). Table 9 in Appendix B shows that liquid assets are correlated with higher educational attainments even after controlling for the number of mortgagors, net wealth and the monthly mortgage payment.

Table 10 replicates specification 2 of Table 3, however in Table 10 we do not control for overappraisal. Basic results remain unchanged. The number of mortgagors (three or more) is a predictor of the probability of fraud: our interpretation is that buyers (who, in our sample, are mostly households) decide to join forces precisely to increase liquidity. The type of job is also a predictor of the amount of undeclared money: employees in the private sector show a lower amount of undeclared money compared to public employees

[49] For completeness, we run the same analysis also in Table 3, where again we loose some observations and significance at the usual levels. In this case, marginal effects suggest that a rise of one percentage point of the shadow economy increases the amount of the transaction price that remains undeclared by 60 euros.

[50] Our proxies (shadow economy and ESV) may suffer from omitted variable bias. However, it should be noticed that using ESV (computed as the difference between two consecutive years) makes it much less likely.

[51] We are grateful to the editor and referees for pointing this out.

[52] The EFF is a survey conducted by the Spanish Central Bank (BdE) that provides detailed information on income, assets, debt and spending of Spanish households. We used the question "What is your balance at the moment in accounts that can be used to make payments?", which could be a proxy for liquid assets.

and self-employed.

We surmised at the end of Section 2 that macro-economic variables, such as unemployment, may affect liquidity and hence, evasion. We test this conjecture in Table 4.

	Fraud Extensive margin	Fraud Intensive margin
	(1)	(2)
Unemployment	-0.033***	0.003***
	(0.007)	(0.001)
Intercept	0.439***	0.111***
	(0.107)	(0.009)
N. obs	1,445	730
* p <0.10. ** p <0.05. *** p <0.01		

Table 4: Tax evasion and Unemployment

On the extensive margin (column 1), an increase in unemployment induces a reduction in the number of fraudulent transactions. Interestingly, unemployment instead has the opposite effect when it comes to the intensive margin (column 2). Indeed, when unemployment increases, the share of the final price that is hidden to the tax authority increases, conditional on fraud. Our interpretation of these results is that the economic crisis affected most people, and this meant a reduction in the share of agents that were able to evade (due to liquidity constraints). It is, however, common to observe in periods of crisis an increase in inequality, with some people suffering more than others. Meanwhile, prices are more likely to decrease in those markets where unemployment is most severe. Taken together, these two effects may mean that those who are not constrained, and hence are able to commit fraud, can actually evade a larger share of the total price. This result is in line with Carozzi (In press), who shows that the 2008 crisis in the UK affected the housing market more relative to the units at the lower end of the market. The reason being the tightening of the credit market, which made the liquidity constraint more stringent for younger or financially weaker potential buyers.

4 Final remarks

This paper contributes to the existing literature on tax evasion by modelling and estimating the determinants of the undeclared money in home purchases. Because tax evasion is usually not observable, the empirical literature has typically relied on imperfect proxies for the level of evasion. We

were able to construct a unique dataset, in which we combine the true transaction price with that declared to the tax authority for sales that occurred in Spain between 2005 and 2011. The results elucidate the determinants of a previously undocumented type of tax evasion: declaring a purchase price below that actually paid in order to avoid the real estate transfer tax. This kind of tax evasion is of particular interest in several countries (e.g. Spain) where this levy is especially heavy and the real estate sector represents a large proportion of the total economy. And where, furthermore, urban development and construction are characterised by a high level of corruption, money embezzlement, illegal workers and other sorts of misconduct.

In contrast to other types of fraud, Akin et al. (2014) suggests that undeclared money is negatively correlated with both the economic crisis and the over-appraisal mechanism used during the boom years in Spain to allow financial institutions to extend borrowing to agents with a low credit score. Indeed, as we show through our analysis, agents who want to evade the transfer tax need access to some 'liquid' savings (i.e. that can be hidden from the tax authority); over-appraisal is instead used by agents who have severe liquidity constraints, in order to be able to borrow larger sums of money. Our model, then, explains how over-appraisal and tax evasion are negatively related. Moreover, in highlighting that agents who resort to over-appraisal are those who are less likely to engage in fraudulent activities, our results have an important policy implication. It is advisable that tax authorities target transactions with low appraisal values with respect to the declared price, so as to increase their auditing performance. This approach is also advantageous in that appraisals are much easier to observe than any other element, such as access to cash or fraudulent behaviour itself. Our results relate to and reinforce Artavanis et al. (2016) in that the tax authority could use the observable behaviour of financial institutions as a screening device to identify subjects that are more likely to commit fraud. To that extent, it is advisable to create a public repository where financial institutions are required to release all appraisals and to ensure its access to the tax authority.

Our empirical analysis shows that "declared" over-appraisal[53] is indeed strongly significant in explaining tax evasion. Previous literature on household borrowing and mortgages has shown that LTV is a crucial element that heavily affects constrained borrowers (Di Maggio et al. 2017, Ganong & Noel 2018, Cloyne et al. 2019). Yet, to the best of our knowledge, this is the first paper that estimates its impact on tax evasion. Interestingly, tax evasion reduces the effective tax rate and, according to our interpretation of the results, less constrained borrowers are those who are more likely to

[53]By "declared" over-appraisal, we mean the ratio between declared appraisal and declared transaction price.

evade. Evading the transfer tax thus has a clear regressive effect in terms of inequality and redistribution, going against what would be desirable. As shown in Best & Kleven (2018), ideally the tax should be lower for constrained households.

Our theoretical model suggests that differences in the level of fraud may originate from various attitudes towards illegality both at the societal and individual levels. Hence, geographical and individual idiosyncrasies in the share of fraudulent transactions (extensive margin) and in the proportion of the transaction value that is hidden from the tax authority (intensive margin) may be due to a different impact of stigma and shame, which are, in turn, affected by the level of social capital and individual characteristics. To this regard, the data show two types of heterogeneity. At the individual level, we observe that education matters and that behaviour differs across regions. We conclude, for extensive margins, that less educated citizens are more prone to tax fraud, as are agents who live in areas with lower social values (high corruption, low transparency and a larger informal economy). Furthermore, for intensive margins, these same agents are also prone to evade larger amounts. These results have two policy implications. On the one hand, increasing trust in society (through greater transparency and strictness towards corrupt prominent people) has a positive effect on the level of fraud committed by citizens; prominence may hence become a criterion for auditing when the tax agency has limited resources. On the other hand, education plays an important role in terms of the level of fraud; hence, long-run policies could also use this channel to increase compliance.

Results are robust to several definitions of corruption at the municipal level or to the use of transparency indices at the province or regional level. The 'guilty feeling' and the loss of reputation of a defrauder decrease when corruption is widespread. This link between individual and collective reputation also helps to explain long-run tax fraud. A short-run increase in corruption due to a housing bubble, as it was the case in Spain, may hurt collective reputation as well as have long-lasting effects in terms of tax fraud. Once again, there are clear policy implications: governments should promote anti-corruption policies,[54] but also educate their citizens. Well-educated citizens who observe responsible governments are less prone to engage in tax evasion.

To the best of our knowledge, this paper is the first to document this phenomenon in such depth, in part made possible by the richness of the available database, that allows us to directly observe both the amount paid by the buyer and the one declared to the tax authority. Further research is needed to fully understand this type of tax fraud and its determinants. Data avail-

[54] Anti-corruption policies include bureaucratic incentives - e.g. punishment, monitoring, compensation and selection - and non-bureaucratic ones - e.g. reducing intermediaries, incentivising wrong-doing reports or facilitating job rotation (Burguet et al. 2016).

ability remains a considerable hurdle. For instance, available measures of social capital or corruption are quite imperfect.[55]

[55]Corruption cases, for instance, are not equally perceived by voters, and information circulates better in some environments than in others (Fernández-Vázquez et al. 2016). Time and geographical differences are better understood with greater knowledge of how different types of illicit behaviours produce externalities on the surrounding community.

Appendix A Proofs

Proof of Lemma 1 . Denote the first order conditions, Eqs. (3) and (4), respectively as $F_1 = 0$ and $F_2 = 0$. The second order conditions require $\frac{\partial F_1}{\partial H} < 0$, $\frac{\partial F_2}{\partial H^u} < 0$ and the determinant of the Hessian matrix is positive: $D(H, H^u) = \frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial H^u} - \frac{\partial F_1}{\partial H^u}\frac{\partial F_2}{\partial H} > 0$.

Define $\phi = i''\frac{B}{H^d} + 2i' > 0$ and $\psi = \pi_u''(fH^u + s) + 2\pi_u' f + \mu_u'' > 0$.

Then, it is immediate to obtain that:

$$\frac{\partial F_1}{\partial H} = h''(H) - \frac{(L - H^u)^2}{(H - H^u)^3}\phi < 0 \tag{10}$$

$$\frac{\partial F_2}{\partial H^u} = -\frac{(H - L)^2}{(H - H^u)^3}\phi - \psi < 0 \tag{11}$$

$$\frac{\partial F_1}{\partial H^u} = \frac{\partial F_2}{\partial H} = -\frac{(H - L)(L - H^u)}{(H - H^u)^3}\phi < 0 \tag{12}$$

It is a matter of simple algebra to show that

$$\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial H^u} - \frac{\partial F_1}{\partial H^u}\frac{\partial F_2}{\partial H} = \frac{(L - H^u)^2}{(H - H^u)^3}\phi\psi - h''(H)\left(\frac{(H - L)^2}{(H - H^u)^3}\phi + \psi\right) > 0. \tag{13}$$

□

Proof of Proposition 1 . Eqs. (5) and (6) are a direct application of the implicit function theorem, applied to a system of two FOCs. For the problem to be well-behaved, the SOCs impose $D(H, H^u) > 0$.

As for the numerator, notice that:

$$\frac{\partial F_1}{\partial L} = \frac{(L - H^u)}{(H - H^u)^2}\phi \tag{14}$$

$$\frac{\partial F_2}{\partial L} = \frac{(H - L)}{(H - H^u)^2}\phi. \tag{15}$$

Eqs. (7) and (8) immediately follow. Since, by assumption, $h''(H) < 0$, the sign of Eq. (8) is unambiguous.

From the definition of $\tilde{V}$, it is immediate to compute how overappraisal changes with H and H^u:

$$\frac{\partial \tilde{V}}{\partial H} = \frac{\alpha(L - H^u)}{(H - H^u)^2} \geq 0 \tag{16}$$

$$\frac{\partial \tilde{V}}{\partial H^u} = \frac{\alpha(H - L)}{(H - H^u)^2} \geq 0. \tag{17}$$

□

Proof of Proposition 2 . We apply again the implicit function theorem to the system of FOCs and have:

$$\frac{\partial H^u}{\partial s} = -\frac{\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial s}}{D(H, H^u)} = \frac{-\frac{\partial F_1}{\partial H}}{D(H, H^u)}(-\pi_u'). \tag{18}$$

It immediately follows that $\frac{\partial H^u}{\partial s} < 0$ as long as $\pi_u' > 0$, while $\frac{\partial H^u}{\partial s} = 0$ as long as $\pi_u' = 0$. □

Proof of Proposition 3 . We apply again the implicit function theorem to the system of FOCs and have:

$$\frac{\partial H^u}{\partial e} = -\frac{\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial e}}{D(H, H^u)} = \frac{-\frac{\partial F_1}{\partial H}}{D(H, H^u)}\left(-\frac{\partial^2 \pi}{\partial H^u \partial e}(fH^u + s) - \frac{\partial \pi}{\partial e}\right) < 0 \tag{19}$$

$$\frac{\partial H^u}{\partial n} = -\frac{\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial n}}{D(H, H^u)} = \frac{-\frac{\partial F_1}{\partial H}}{D(H, H^u)}\left(-\frac{\partial^2 \mu}{\partial H^u \partial n}\right) < 0 \tag{20}$$

$$\frac{\partial H^u}{\partial \theta} = -\frac{\frac{\partial F_1}{\partial H}\frac{\partial F_2}{\partial \theta}}{D(H, H^u)} = \frac{-\frac{\partial F_1}{\partial H}}{D(H, H^u)}\left(-\frac{\partial^2 \mu}{\partial H^u \partial \theta}\right) < 0 \tag{21}$$

□

Appendix B Complementary tables

	Transactions with undeclared money	Undeclared money	Undeclared money, conditional on fraud
2005	80.89%	12.59%	14.17%
2006	66.07%	9.58%	14.50%
2007	59.77%	7.91%	13.23%
2008	48.24%	7.34%	15.21%
2009	38.67%	6.13%	15.86%
2010	34.72%	6.24%	17.97%
2011	31.15%	6.73%	21.62%
Andalusia	59.36%	10.90%	18.36%
Aragon	44.07%	4.53%	10.27%
C. La Mancha	41.86%	6.38%	15.23%
C. León	41.38%	7.26%	17.54%
Catalonia	37.59%	5.27%	14.01%
C. Madrid	53.51%	6.97%	13.03%
Valencian C.	61.80%	11.84%	19.15%

Table 5: Evolution of the undeclared money (shares) over time and space.

	(1)		(2)		(3)		(4)	
	Probit	Tobit	Probit	Tobit	Probit	Tobit	Probit	Tobit
Eq. undeclared								
Spending (H)	0.000***	-0.000	0.000***	0.000*	0.000**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Corruption (s)					0.487***	0.009		
					(0.077)	(0.009)		
Shadow economy (n)							0.028**	0.006***
							(0.013)	(0.001)
Intercept	-0.277***	0.181***	-1.437***	0.240***	-0.914***	0.229***	-1.101***	0.116***
	(0.081)	(0.010)	(0.365)	(0.047)	(0.211)	(0.029)	(0.311)	(0.038)
Eq. overappraisal								
Spending (H)	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Corruption (s)					-0.088	0.009		
					(0.110)	(0.012)		
Shadow economy (n)							-0.024	0.001
							(0.019)	(0.002)
Intercept	1.847***	1.509***	6.912	1.459***	2.387***	1.501***	2.895***	1.493***
	(0.119)	(0.013)	(354.870)	(0.055)	(0.431)	(0.032)	(0.540)	(0.049)
ρ	-0.136**	-0.234***	-0.049	-0.266***	-0.103	-0.266***	-0.100	-0.270***
	(0.057)	(0.054)	(0.065)	(0.058)	(0.067)	(0.063)	(0.062)	(0.058)
N. Obs.	1,445		1,445		1,233		1,432	
Year F.E.	No		Yes		Yes		Yes	
Region F.E.	No		Yes		No		No	

* p <0.10. ** p <0.05. *** p <0.01

Table 6: Bivariate model

Tax morale

For robustness, Table 7 uses two indicators of corruption other than the one we constructed and used in Tables 2 and 3. Columns (1) to (4) use the Global Transparency Index (GTI), published by Transparency International, which measures the level of transparency of public institutions through an evaluation of data and information available on the organisation's website. To make the results using the GTI easier to compare with those using our corruption index, we normalised its values, so that the index ranges from 0 (maximum transparency) to 100 (minimum transparency).[56] GTI covers 110 Spanish municipalities: 1,115 observations in our 'whole sample' overlap with a municipality covered by the GTI. Generally, the GTI includes five sub-indexes: information, relation with citizens, economic transparency, transparency in contracting services and transparency in urban planning and public works. Columns (1)-(2) of Table 7 depict the results using the adjusted GTI as the measure of corruption, while columns 3-4 depict the results using the adjusted GTI sub-index 'transparency in urban planning and public works' (GTI-Urban). This robustness test confirms our results: we observe more fraudulent transactions (both on the extensive and intensive margin) in more corrupt areas, which we interpret as areas with lower law enforcement (audit) and, therefore, where stigma plays a minor role (proposition 2). In particular, an increase in one point of either the GTI or the GTI-Urban index reduces the amount of the value that is undeclared by 300 euros.[57] Colums (5) and (6), instead, are computed using the Quality of Government (QoG) data from the Quality of Government Institute; in particular, we used the corruption variable (data available at the regional level).[58] Results are consistent with the previous ones.

We used, in Table 8, a different indicator of social capital and morality provided by the European Values Study. The study provides insights into the ideas, beliefs, preferences, attitudes, values and opinions of citizens across Europe. Specifically, we exploit the question 'justify cheating on tax' and compiled this information for every Spanish region for both the 1999 and 2008 waves. We use their difference as a proxy for the changes in tax evasion behaviour. A higher index value means that tax evasion is more tolerated. Again, the results are significant and their sign is that predicted by the theoretical model and in accordance with those obtained using different proxies for shame.

[56] We adjusted the GTI by replacing it with (100-GTI). Within our sample, the most transparent municipality has a score of 2.5 (Gijón), while the most opaque has a score of 80 (Vélez Málaga).

[57] Results are robust to transparency and corruption data aggregation at the provincial level.

[58] In this case, we used the inverse of the original index in the estimation.

	GTI		GTI-Urban		QoG	
	(1)	(2)	(3)	(4)	(5)	(6)
	Logit	Tobit	Logit	Tobit	Logit	Tobit
Corruption (s)	0.014***	299.895***	0.014***	266.570***	0.295	13407.220**
	(0.004)	(85.733)	(0.003)	(71.262)	(0.289)	(5974.915)
Overappraisal ($\tilde{V}$)	-0.717***	-32053.534***	-0.746***	-32679.301***	-0.748***	-29526.042***
	(0.274)	(5920.893)	(0.274)	(5914.562)	(0.237)	(4866.638)
Spending (H)	0.000	0.104***	0.000	0.100***	-0.000	0.095***
	(0.000)	(0.022)	(0.000)	(0.021)	(0.000)	(0.019)
Intercept	1.209**	51635.566***	1.344**	52033.244***	0.078	13457.246
	(0.606)	(12969.858)	(0.604)	(12858.207)	(0.489)	(10134.826)
N. obs	1,115		1,115		1,440	
Year F.E.	Yes		Yes		Yes	
* p <0.10. ** p <0.05. *** p <0.01						

Table 7: Estimated model using the GTI corruption and the Quality of Government indices.

	Logit	Tobit
Absence of Social values (n)	0.223**	3134.439*
	(0.087)	(1825.523)
Overappraisal ($\tilde{V}$)	-0.723***	-29512.898***
	(0.237)	(4874.073)
Spending (H)	0.000	0.092***
	(0.000)	(0.019)
Intercept	0.115	19335.386**
	(0.467)	(9621.313)
N. obs	1,445	
Year F.E.	Yes	
* p <0.10. ** p <0.05. *** p <0.01		

Table 8: Estimated model using the European Social Values index.

Educational level	
Secondary	4776.386*
	(2581.457)
Tertiary	13439.668***
	(2594.288)
Number of mortgagors	857.281
	(840.248)
Mortgage (monthly payment)	4.155***
	(1.395)
Net wealth	0.001***
	(0.000)
Intercept	212.753
	(3164.502)
N. obs	1,231
References. Education: Primary	
* p <0.10. ** p <0.05. *** p <0.01	

Table 9: Determinants of liquid assets in the Spanish Survey of Household Finance

	Logit	Tobit
Spread	-0.308	-9739.158
	(0.283)	(6272.138)
Educational level		
Secondary	-0.457*	-12413.662**
	(0.258)	(5856.108)
Tertiary	-0.654*	-14127.292*
	(0.366)	(8406.455)
Number of mortgagors		
Two	0.134	7790.224
	(0.238)	(5504.489)
Three or more	1.623**	14468.533
	(0.690)	(10879.369)
Labour situation		
Private sector	-0.512	-22004.804**
	(0.468)	(10666.177)
Public sector	-0.184	-15306.126
	(0.541)	(12437.782)
Self-employed	0.375	-2617.273
	(0.650)	(13787.316)
Intercept	-13.477	13736.956
	(680.939)	(19794.326)
N. Obs.	430	
Year F.E.	Yes	
Local F.E.	Yes (Region)	
References. Education: Primary; Mortgagors: One; Labour: Non-Occupied		
* p <0.10. ** p <0.05. *** p <0.01		

Table 10: Estimated model: sample with individual characteristics without overappraisal

References

Akin, O., Montalvo, J. G., Villar, J. G., Peydró, J.-L. & Raya, J. M. (2014), 'The real estate and credit bubble: evidence from Spain', *SERIEs: Journal of the Spanish Economic Association* **5**, 223–243.

Alesina, A. & La Ferrara, E. (2000), 'Participation in heterogeneous communities', *Quarterly Journal of Economics* **115**, 847–904.

Alesina, A. & La Ferrara, E. (2002), 'Who trusts others?', *Journal of Public Economics* **85**, 207 – 234.

Allingham, M. G. & Sandmo, A. (1972), 'Income tax evasion: a theoretical analysis', *Journal of public economics* **1**, 323–338.

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

Alm, J., Martinez-Vazquez, J. & Schneider, F. (2004), "Sizing' the problem of the hard-to-tax', *Contributions to Economic Analysis* **268**, 11–75.

Alm, J. & Torgler, B. (2006), 'Culture differences and tax morale in the United States and in Europe', *Journal of economic psychology* **27**, 224–246.

Amromin, G., Huang, J., Sialm, C. & Zhong, E. (2018), 'Complex Mortgages*', *Review of Finance* **22**(6), 1975–2007.

Artavanis, N., Morse, A. & Tsoutsoura, M. (2016), ' Measuring Income Tax Evasion Using Bank Credit: Evidence from Greece *', *The Quarterly Journal of Economics* **131**(2), 739–798.

Ben-David, I. (2011), 'Financial constraints and inflated home prices during the real estate boom', *American Economic Journal: Applied Economics* **3**, 55–87.

Benito, B., Guillamón, M.-D. & Bastida, F. (2015), 'Determinants of urban political corruption in local governments', *Crime, Law and Social Change* **63**(3-4), 191–210.

Best, M. C. & Kleven, H. J. (2018), 'Housing market responses to transaction taxes: evidence from notches and stimulus in the U.K.', *The Review of Economic Studies* **85**(1), 157–193.

Boffa, F., Piolatto, A. & Ponzetto, G. (2016), 'Political centralization and government accountability', *Quarterly Journal of Economics* **131**, 381–422.

Bradley, S. (2018), 'Assessment limits and timing of real estate transactions', *Regional Science and Urban Economics* **70**, 360 – 372.

Bruhin, A., Fehr-Duda, H. & Epper, T. (2010), 'Risk and rationality: uncovering heterogeneity in probability distortion', *Econometrica* **78**, 1375–1412.

Burguet, R., Ganuza, J. J. & Montalvo, J. G. (2016), The microeconomics of corruption. a review of thirty years of research, *in* L. C. Corchón & M. A. e. Marini, eds, 'Handbook of Game Theory and Industrial Organization', Vol. II, Edward Elgar Publishing, chapter 16.

Carozzi, F. (In press), 'Credit constraints and the composition of housing sales. Farewell to first-time buyers?', *Journal of the European Economic Association* (doi.org/10.1093/jeea/jvz017).

Chander, P. & Wilde, L. L. (1998), 'A general characterization of optimal income tax enforcement', *Review of Economic Studies* **65**, 165–183.

Cho, M. & Megbolugbe, I. F. (1996), 'An empirical analysis of property appraisal and mortgage redlining', *The Journal of Real Estate Finance and Economics* **13**, 45–55.

Cloyne, J., Huber, K., Ilzetzki, E. & Kleven, H. (2019), 'The effect of house prices on household borrowing: a new approach', *American Economic Review* **109**(6), 2104–2136.

Commission, E. (2014), 'Special eurobarometer 397: Corruption'.

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A. & Yao, V. (2017), 'Interest rate pass-through: Mortgage rates, household consumption, and voluntary deleveraging', *American Economic Review* **107**(11), 3550–88.

Di Porto, E., Persico, N. & Sahuguet, N. (2013), 'Decentralized deterrence, with an application to labor tax auditing', *American Economic Journal: Microeconomics* **5**(1), 35–62.

Dwenger, N., Kleven, H., Rasul, I. & Rincke, J. (2016), 'Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany', *American Economic Journal: Economic Policy* **8**(3), 203–32.

Engström, P., Nordblom, K., Ohlsson, H. & Persson, A. (2015), 'Tax compliance and loss aversion', *American Economic Journal: Economic Policy* **7**(4), 132–64.

Esteller-Moré, A., Piolatto, A. & Rablen, M. D. (2018), Taxing high-income earners: tax avoidance and mobility, *in* N. Hashimzade & Y. Epifantseva,

eds, 'The Routledge Companion to Tax Avoidance Research', Taylor & Francis Group Ltd, chapter 20.

Falkinger, J. (1995), 'Tax evasion, consumption of public goods and fairness', *Journal of Economic Psychology* **16**, 63–72.

Fernández-Vázquez, P., Barberá, P. & Rivero, G. (2016), 'Rooting out corruption or rooting for corruption? The heterogeneous electoral consequences of scandals', *Political Science Research and Methods* **4**, 379–397.

Fortin, B., Lacroix, G. & Montmarquette, C. (2000), 'Are Underground Workers More Likely to be Underground Consumers?', *Economic Journal* **110**, 838–60.

Freybote, J., Ziobrowski, A. & Gallimore, P. (2014), 'Residential real estate appraisal bias in the absence of client feedback', *Journal of Housing Research* **23**, 127–142.

Fritzsche, C. & Vandrei, L. (2019), 'The german real estate transfer tax: Evidence for single-family home transactions', *Regional Science and Urban Economics* **74**, 131 – 143.

Ganong, P. & Noel, P. (2018), 'Liquidity vs. wealth in household debt obligations: Evidence from housing policy in the great recession', NBER WP 24964.

Ganong, P. & Noel, P. (2019), 'Consumer spending during unemployment: Positive and normative implications', *American Economic Review* **109**(7), 2383–2424.

Gete, P. & Reher, M. (2016), 'Two extensive margins of credit and loan-to-value policies', *Journal of Money, Credit and Banking* **48**, 1397–1438.

Gordon, J. P. (1989), 'Individual morality and reputation costs as deterrents to tax evasion', *European economic review* **33**, 797–805.

Hashimzade, N., Myles, G. D. & Tran-Nam, B. (2013), 'Applications of behavioural economics to tax evasion', *Journal of Economic Surveys* **27**, 941–977.

Kim, Y. (2003), 'Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion', *Journal of Public Economics* **87**, 1591–1616.

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

Kopczuk, W. & Munroe, D. (2015), 'Mansion tax: The effect of transfer taxes on the residential real estate market', *American Economic Journal: Economic Policy* **7**(2), 214–57.

LaCour-Little, M. & Malpezzi, S. (2003), 'Appraisal quality and residential mortgage default: evidence from Alaska', *The Journal of Real Estate Finance and Economics* **27**, 211–233.

Lang, W. W. & Nakamura, L. I. (1993), 'A model of redlining', *Journal of Urban Economics* **33**, 223–234.

Loebs, T. (2005), 'Systemic risks in residential property valuations: Perceptions and reality', Working Paper: Collateral Assessment and Technologies Committee.

Macho-Stadler, I. & Pérez-Castrillo, J. D. (1997), 'Optimal auditing with heterogeneous income sources', *International Economic Review* **38**(4), 951–968.

Medina, L. & Scheneider, F. (2017), 'Shadow economies: what did we learned the last 20 years', *IMF Working Paper* **WP/18/17**, 1–59.

Montalvo, J. G. & Raya, J. M. (2012), 'What's the right price of Spanish residential real estate?', *Spanish Economic and Financial Outlook* **1**, 22–29.

Montalvo, J. G. & Raya, J. M. (2018), 'Constraints on LTV as a macroprudential tool: a precautionary tale', *Oxford Economic Papers* **70**, 821—-845.

Nakamura, L. I. et al. (2010), 'How much is that home really worth? Appraisal bias and house-price uncertainty', *Business Review* pp. 11–22.

Piolatto, A. & Rablen, M. (2017), 'Prospect theory and tax evasion: A reconsideration of the Yitzhaki puzzle', *Theory and decision* **82**, 543–565.

Piolatto, A. & Trotin, G. (2016), 'Optimal income tax enforcement under prospect theory', *Journal of Public Economic Theory* **18**, 29–41.

Reinganum, J. F. & Wilde, L. L. (1985), 'Income tax compliance in a principal-agent framework', *Journal of Public Economics* **26**, 1–18.

Sardá, J. (2014), 'La economía sumergida pasa factura', *El avance del fraude en España durante la crisis* .

Scheneider, F. (2005), 'Shadow economies around the world: what do we really know', *European Journal of Political Economy* **21 (3)**, 548–642.

Scheneider, F. & Enste, D. (2000), 'Shadow economies: size, causes and consequences', *The Journal of Economic Literature* **38 (1)**, 77–114.

Schildberg-Hörisch, H. & Strassmair, C. (2012), 'An experimental test of the deterrence hypothesis', *Journal of Law, Economics, & Organization* **28**(3), 447–459.

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

Slemrod, J. & Yitzhaki, S. (2002), Tax avoidance, evasion, and administration, *in* A. J. Auerbach & M. Feldstein, eds, 'Handbook of Public Economics', Vol. 3 of *Handbook of Public Economics*, Elsevier, chapter 22, pp. 1423–1470.

Traxler, C. (2010), 'Social norms and conditional cooperative taxpayers', *European Journal of Political Economy* **26**, 89–103.

Treisman, D. (2000), 'The causes of corruption: a cross-national study', *Journal of public economics* **76**, 399–457.

Weck-Hanneman, H. & Frey, B. S. (1985), 'Measuring the shadow economy: the case of Switzerland', *The Economics of the Shadow Economy, Berlin: Springer-Verlag* pp. 142–65.

Yitzhaki, S. (1974), 'A note on income taxation: a theoretical analysis', *Journal of Public Economics* **3**, 201–202.