Tax evasion and British real estate: an estimation of the offshore wealth held through British real estate

Jeanne Bomare

Under the supervision of Thomas Piketty (PSE) and Gabriel Zucman (UC Berkeley)

Abstract

This work proposes a region-by-region estimation of the offshore wealth held through London real estate. It does so by exploiting information leaked in the Panama Papers, allowing to identify some of the companies buying in London. The analysis in focused on shell companies, and machine learning techniques are used to recover the nationality of their beneficial owner(s). I estimate that £75 billion are stored by shell companies in British real estate. Most of this amount is attributed to British citizens and Middle East countries. These new estimates shed a new light on the preferred assets to hold offshore wealth for different nationalities. They also give a new insight in the use of shell companies around the world.

Keywords: tax evasion, offshore wealth, housing market, safe haven effect.

JEL Codes: H26, H87, E21, G12

Contents

1 Introduction

A new surcharge on stamp duty for foreigners, a tax levied on property purchases, is being considered by the United Kingdom Treasury[1]. It follows an announcement made by Theresa May in October 2018 at the Conservative party conference and would apply to individuals as well as companies. The measure echoes a dramatic increase in the flows of foreign money directed at the UK's real estate since the beginning of the century. For example, overseas companies had bought a little less than 2,000 properties in London in 1999. This number had about doubled in 2018 (1).

The implications of the surge in capital flooding the UK's real estate market are not clear and this issue has attracted a lot of attention over the last few years. One of the most debated question lies in the effects of foreign investment on nationals and their access to the property market[2]. On the one hand, it has been argued that the attractiveness of the country prompted the construction of properties and dwellings to meet the residential or investment needs of this new class of buyers (Scanlon et al. (2017)). This development would in return participate to the accession to home-ownership of some categories of British people who are currently excluded from the market.

If the housing stock is relatively inelastic however, the globalization of the British real estate market could only result in an increase of competition for a fixed number of properties, preventing low income buyers from getting on to the property ladder. This second scenario appears to be the most plausible. Sà (2016) shows that foreign investment does not have a significant impact on the number of dwellings constructed in the UK. Rather, it seems to have a positive effect on house prices across the country, even at the lowest levels of the distribution of prices. She links this increase in prices to the finding that foreign investment appears to reduce the rate of home ownership, indicating that some groups are pushed out of the real estate market.

Another question raised by the surge of sales to overseas buyers lies in the origins of the funds used by them. Since 2004, almost £200 million have been investigated under suspicions of criminal activities (Transparency_International (2012)). The wide use of anonymous shell companies in those transactions raises the suspicion that they are used in money-laundering operations. This issue appears to be of particularly large-scale in London (Reumer (2019)) where the purchases made by firms incorporated in tax havens are very frequent. The banking and estate agency intermediaries are required to gather evidence of the legality of the resources

[1] https://www.gov.uk/government/news/new-stamp-duty-land-tax-surcharge-for-non-uk-resident-homebuyers-to-be-introduced

[2] See for example "How basement-loving billionaires are forcing everyone else out", *The Guardian*, March 2918 or "A tale of Two London", *Vanity Fair*, April 2013

involved in the transactions but a recent report of the Treasury shows that this rule is largely circumvented in practice (HM Treasury (2017)).

In this work, I hope to make the first step towards tackling this issue. Using publicly available data published by the Land Registry and recording every purchase of real estate in London by a company incorporated overseas, I propose to predict the region of origin of all the people buying through a shell company. To do that, I build on an insight of Badarinza and Ramadorai (2018). They argue that foreign nationals have specific preferred habitats in London according to their nationality. They also show that London real estate operate as a safe haven asset: when risk in foreign countries intensify, the demand for London real estate increases, probably because of its high price ans its high liquidity. In other words, safe haven assets function as a store of value in times of economic or political distress. These two features suggest that the timing and the location of the transaction are strong predictors of the nationality of the buyer. To test this hypothesis, I begin by finding the preferred habitat of a group of nationalities who have completed transactions in London. To do this, I exploit several files that have been leaked to the press since 2013: The Bahamas Leaks, the Offshore Leaks, the Panama Papers and the Paradise Papers. The leaks provide information on the ultimate beneficial owners of more than 785,000 companies, constituting a very valuable source of information in the study of shell companies. With this insight on the location of the buyers, I am able to draw a link between the evolutions of risk in the home countries and the amount of capital invested in the preferred habitats of these same countries. I show that some regions of the world like the Middle East and Eastern Europe, but also Western Europe, exhibit investing behaviors compatible with the existence of safe haven assets: they seem to invest more in London real estate at times of economic uncertainty. On the contrary, some regions like South Asia and East Asia are associated with more traditional motives of investment, like capital gains or rental yields, investing more in London real estate when the risk in their country is low.

This first analysis shows that according to their region of origin, the buyer will not invest at the same time nor in the same location. I use these patterns to predict the region of origin of all the beneficial owners of the companies buying in London through shell companies listed in the Land Registry dataset. The predictions are carried using machine learning techniques, more specifically random forests and some derivated techniques. Following this, I propose a back-of-the-envelope calculation to estimate the actual amount of offshore wealth stored in London's real estate, region-by-region. As most prices are not available in the data, I make some conservative hypotheses to propose a lower-bound on this amount.

I propose two sets of estimation. First, I look at the total offshore wealth stocked in London

real estate by every group of countries since 1999. This estimation can include properties that have since been sold or have been demolished. However, it is informative as it gives a picture of the identity of the main investors in London real estate, potentially providing an insight into the motivations presiding at those transactions. Following on, I attempt to estimate the value owned by every region in London today, using the current version of the Land Registry record.
This work is related to several strands of the literature. First, it contributed to the literature on offshore wealth and tax evasion. Several estimations of the global amount of offshore wealth have been proposed. The most convincing is Zucman (2013), who makes the hypothesis that anomalies in portfolio investment positions of the countries indicate the existence of capital stored in tax havens. He finds that around 8% of the world's financial wealth, corresponding to about 10% of the world's GDP, is hidden in tax havens. But it is likely to be a conservative estimate, as only a part of the financial wealth is taken into account. Studies conducted by the Boston Consulting Group conclude to about 14% of the world's GDP being held in tax haven, while Henry (2012) argue that more than the fourth of the world's GDP corresponds to offshore wealth. Alstadsæter et al. (2019) provide a more granular estimation, calculating a country-by-country decomposition of the wealth held in tax havens. Some works have studied the implications of tax evasion. Alstadsæter et al. (2018) produce new estimates of income distributions and show that taking into account offshore wealth greatly increase the measure of inequality.
This work is also related to a strand of the literature studying the creation and use of shell companies. Omartian (2017) show that tax information treaties increase the number of shell companies created as recorded by the Panama Papers. These findings suggest that they are used as a vehicle of tax evasion and not for legal investment or financial purposes. They are supported by Caruana-Galiza and Caruana-Galizia (2016) who show that policies trying to fight tax evasion prompt the use of sophisticated concealment techniques such as shell companies. Bayer et al. (2018) demonstrate that the fear of expropriation and property confiscations drives the creation of shell entities up. The time patterns of incorporation recovered by the authors indicate that the shell vehicles are used by individuals trying to hide assets obtained illegally or to evade tax, not by people securing their capital from arbitrary political regimes.
Ultimately, my thesis is linked to the research on the effects of the increasing amount of foreign investment in global cities and in London in particular. Sà (2016) shows that it led to an increase in UK's housing prices, at every point of the price distribution. However, it is not the case for every global city: Sotura (2011) demonstrates that foreigners' impact on prices on Paris' real estate is negligible. Deverteuil (2017) studies the social impacts of foreign in-

vestment on the demography of London and on the city itself. He shows that overseas buyers locate mostly in prime areas and that mostly have no attachment to their particular location, changing the way the city is inhabited.

My contribution in this work is twofold. On the theoretical side, it focuses on non-financial offshore wealth, more precisely on offshore wealth held through real estate. Prior to this, most of the literature had studied exclusively financial offshore wealth, captured by comparing global macroeconomic imbalances. No comprehensive real estate register is available at the international level, making it almost impossible to estimate the global amount of offshore wealth held through housing. However, it is likely that London is one of the few cities attracting the biggest amounts of capital, with others like New York for example. Indeed, being a global city particularly open to foreign investment and offering a large stock of luxury properties, it is a destination of choice for wealthy foreigners. Moreover, my thesis directly tackles the issue of shell companies. Shell entities are usually a problem when it comes to estimating offshore wealth, because they cannot be attributed to a country in particular without making strong hypotheses. Alstadsæter et al. (2019) stop their country-by-country analysis in 2007, before the large rise in the number of shell companies of the financial sector to avoid this issue. My method to recover the nationality of the beneficial owners of sham corporations could be expanded to other situations, when enough information is available. The notion of preferred habitat used by Badarinza and Ramadorai (2018) is actually taken from the financial field, where investors can have different preferred assets to invest in.

On the methodological side, this work proposes a way in which it is possible to exploit the files leaked from banks and assets management companies. The Panama Papers and other leaked files have been analyzed extensively in the press and by the non-governemental sector. However, the files have only been exploited in a few academic papers (Badarinza and Ramadorai (2018), Omartian (2017), Caruana-Galizia and Caruana-Galizia (2016)). This is probably due to the fact that shell companies are usually not comprehensively listed on a register like the Land Registry record.

The rest of this essay is structured as follows. Section 2 presents the data used in the analysis. Section 3 recovers the links between the variations of risk in a country and the level of investment in London real estate. Section 4 proposes a prediction of the region of origin of the people buying London real estate through shell companies. Section 5 gives an estimation of the amount of wealth held in London real estate by citizens of each region of the world.

2 Data

2.1 Investments of foreign buyers

The British Land Registry records all the real estate purchases made in the United Kingdom by foreign companies in the Overseas Companies Ownership Dataset (OCOD). The register does not list the transactions made by companies incorporated in the UK or by private individuals. The OCOD provides the name, country of incorporation and address of the buying company, address and date of the purchase, tenure (Freehold or Leasehold [3]) and the price paid, when available. It is exhaustive as companies are required to report their purchase. The register is updated every month, excluding properties that have been sold and adding newly bought titles. The last few months of the record are publicly available. I obtained the entire record through a Freedom Of Information (FOI) request, allowing me to make use of all the data available since the beginning of the registry, including sold properties. It amounts to 439,893 transactions, the first title being recorded in 1959. When keeping only the transactions attributed to companies incorporated in tax havens[4], the record counts 217,153 transactions. The analysis therefore focuses on transactions made through shell companies, meaning a company which has no actual activity but has probably been created to be a vehicle of investment. It is important to note that the use of a shell company itself is not illegal, even if a large part of these entities serve criminal purposes. Relatively speaking, The OCOD is likely to list many purchases that are at the top of the real estate distribution in terms of prices. This is partly due to the fact that using the vehicle of shell companies is a complex method of investment.

The analysis is restricted to the transactions taking place in London. Firstly, real estate in London could present some specific characteristics compared to the rest of England. House prices in London tend to be higher and follow a different trend compared to the UK as a whole: London's prices are significantly higher than the national average, and throughout the crisis they experienced a lesser fall than the rest of the country. Different persons could then buy in the capital for distinct motives: the high returns ensured by real estate in London make them more likely to be bought for investment and capital gains purposes (Fernandez et al. (2016)). Secondly, restricting the sample to London allows for a more robust analysis as it gathers about 30% of all the observations. Indeed, it is a destination of choice for wealthy foreigners, being a global city particularly open to foreign investment flows and with a high stock of luxury

[3]Freehold properties are held for an infinite duration while Leasehold properties have a fixed or maximum duration.

[4]To determine which countries are tax havens, I follow the list established in Alstadsæter et al. (2019) (see the file panama_explore.do on http://gabriel-zucman.eu/offshore/)

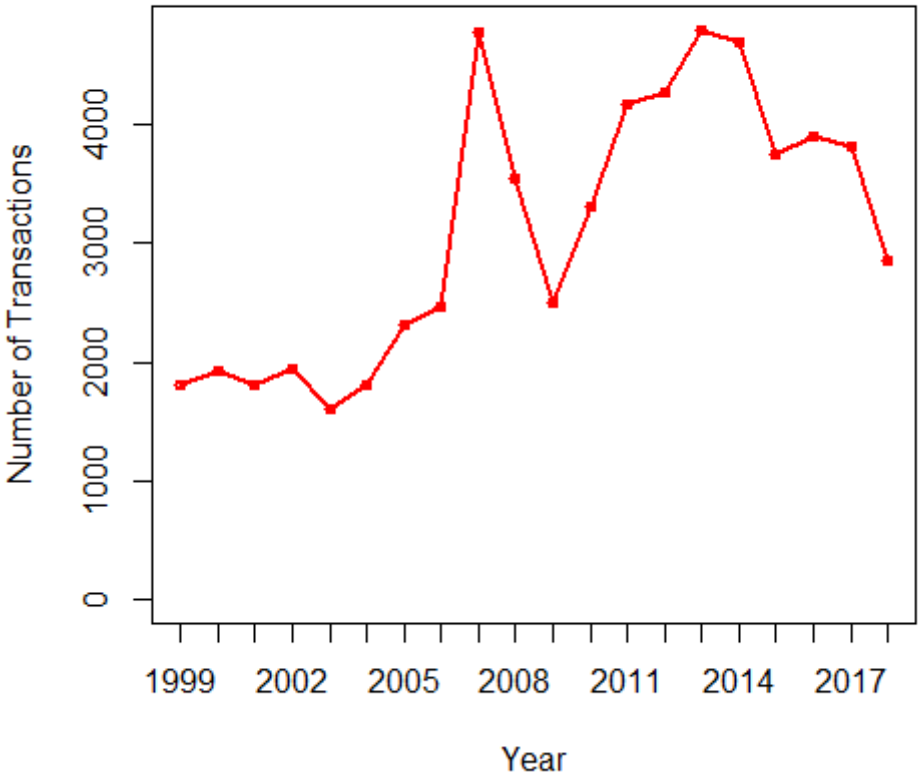

FIGURE 1: Number of transactions per year
Source: Land Registy OCOD

properties. Taking into account only the purchases made in the capital, the OCOD is reduced to 63,586 transactions, made by 29,592 unique companies.

The first transaction registered in the OCOD dates back to 1959, but most of the purchases take place after 1999. Figure 1 shows the number of transactions in the record from 1999 to 2018. The flows of capital increase continuously from 2003 to 2007, with a peak of more than 4,000 transactions in 2007. The number of purchases rapidly declines in 2008, most likely as a result of the financial crisis. Flows of capital recover quickly and steeply increase until 2013. This rise can be explained by the evolution of the institutional and financial environment. First, after the crisis, G20 countries made fiscal evasion a "political priority" (Johannesen and Zucman (2014)), pressuring tax havens to sign bilateral treaties. As a result, numerous treaties have been signed in the period 2009-2010, making information exchange easier in the case of a suspicion of tax evasion. The beginning of what appeared to be a crackdown in tax havens could

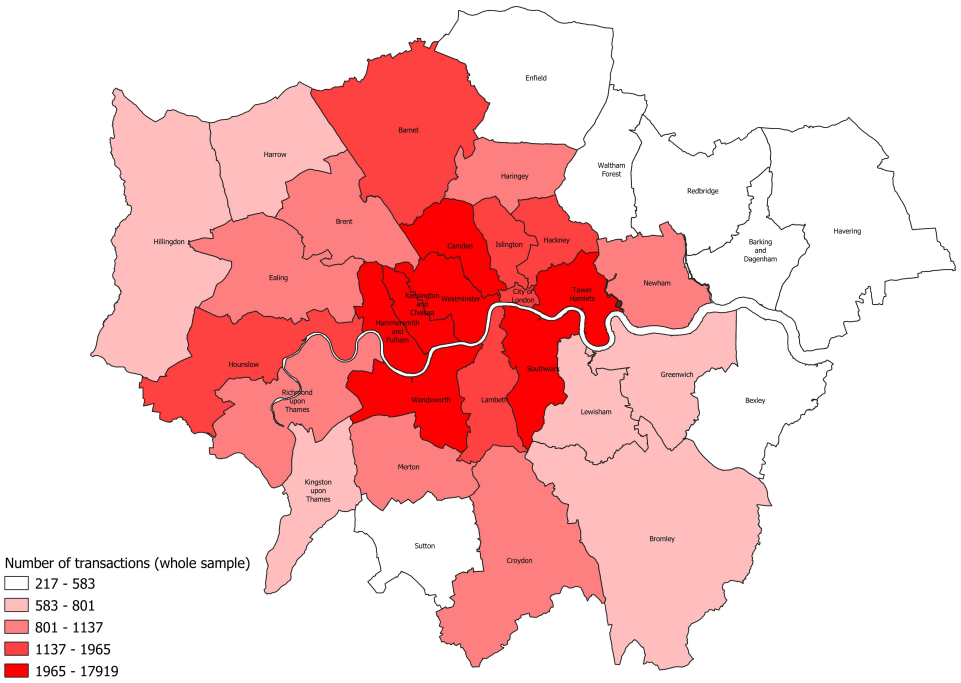

FIGURE 2: Geographical Distribution of the transactions in the OCOD
Source: Land Registry OCOD

have prompted the use of concealment technology making detection less likely, one technique of which would be operating through shell companies (Omartian (2017)). Moreover, the weakness of the pound during the period of 2008 – 2011 made transactions in euros or dollars relatively more attractive, which potentially led to an increase of purchases by foreign nationals. The number of transactions reaches another peak in 2013, before diminishing almost continuously. This contraction could reflect a deprecated political and economic environment in the UK, with an uncertainty fueled by the start of the Brexit campaign and the results of the referendum.

Figure 2 shows the location of the transactions across London boroughs. They are concentrated in the Center and in the West of the City, particularly in the boroughs of Westminster and Kensington and Chelsea. 3 shows the mean income by borough[5] for the year 2017. The two figures indicate that the transactions tend to locate in relatively wealthy areas of London, including the North-West quarter of the city. In particular, Westminster and Kensington and Chelsea correspond to the two richest boroughs of London. This distribution seems to confirm that the register is likely to capture properties located at the top of the distribution of prices. Table 1 presents the decomposition of the unique companies appearing in the OCOD by country incorporated. Most of the companies are incorporated in the British Virgin Islands, Jersey,

[5]The income statistics are recovered from the Office for National Statistics.

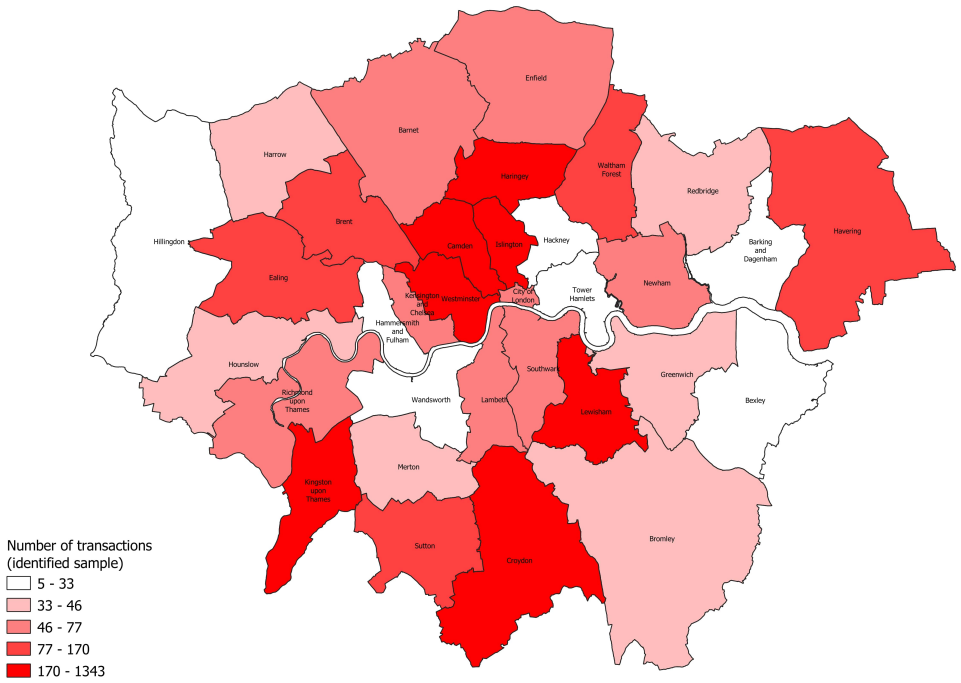

FIGURE 3: Mean income by borough as of 2017
Source: Office for National Statistics

Guernsey and the Isle of Man. One would expect that shell companies would be held in these regions as opposed to other more prominent tax havens (such as Switzerland and Hong Kong), due to the economic, geographic and cultural proximity between the UK and these islands. The high percentage of companies registered in the British Virgin Islands can be explained by the haven's regulations which allows to create shell entities easily.

	Proportion of incorporated companies
Bahamas	2,1%
BVI	40,9%
Guernsey	6,7%
Isle of Man	7,9%
Jersey	16,1%
Panama	4,2%
Seychelles	2,2%
Mean number of properties/company	2,14

TABLE 1: Proportion of companies incorporated for each country, OCOD sample

2.2 Leaks data

To be able to recover patterns linking the characteristics of the transaction and the nationality of the owners, it is essential to identify some of the investors appearing in the OCOD. To do that, I exploit several files that have been leaked to the press since 2013. The Offshore Leaks, the Bahamas Leaks, the Paradise Papers and the Panama Papers are documents from law firms and corporate service providers revealing information on the beneficial owners of thousands of shell companies they created or managed for their clients. Taken together, they provide an insight in the structure of more than 785,000 entities. The files have been gathered by the International Consortium of Investigative Journalists[6] and have resulted in the publication of many large-scale analyses in national newspapers. However, they have not been used extensively in the academic world. To identify the nationalities of foreign investors in the UK, I compare the companies listed in the Land Registry record and the ones appearing in the leaked files[7]. 1,618 companies are identified in the process, 84,9% of them thanks to the Panama Papers. Their beneficial owners are from a panel of 70 nationalities. Table 10 shows the proportion of each nationality in the sample. The most represented country is the United Kingdom, with more than 1 out of 4 beneficial owners being identified as British. Buying through a shell company for a British citizen provides many financial benefits, such as increasing the ease with which to avoid inheritance tax. But tax evasion is not the only motive to use shell companies. As a matter of fact, more than 22% of the proprietors identified in my sample are from the United Arab Emirates, a country which does not have an income tax, indicating that they did not buy real estate in the UK through such a vehicle to hide the money from the fiscal administration. Alternatively, shells are used to gain access to better financial services that are unavailable in the home country(Johannesen and Zucman (2014)). They can also be used to add another layer of secrecy on the origins or the real owners of the funds managed.

However, my identification procedure could suffer from a selection bias if I am able to identify the owners associated to transactions of specific nationalities more easily. To explore this hypothesis, I first compare the identified sample to the OCOD. As shown in table 2, the geographical dispersion of the identified transactions is very similar to the one of the whole Land Registry dataset. Most of the transactions are located in the borough of the City of Westminster, followed by Kensington and Chelsea. The mean number of proprieties owned by

[6] All of the files have been made public and free to access on https://offshoreleaks.icij.org/
An online database has also been created, which allows users to search directly for a name or a country.

[7] To identify the companies present both in the leaked files and in the OCOD, I first standardized the names of all the companies. I realized a first matching by names. On those companies therefore identified, I only kept the ones who where incorporated in the same country.

	Total sample	Identified sample
Barnet	3%	1%
Camden	5%	7%
City of London	2%	1%
City of Westminster	27%	31%
Hammersmith and Fulham	3%	5%
Hounslow	3%	3%
Islington	3%	1%
Kensington and Chelsea	14%	15%
Lambeth	4%	2%
Southwark	3%	2%
Tower Hamlets	5%	3%
Wandsworth	5%	7%
Mean n of properties	2,1	2,0

TABLE 2: Percentage of transactions located in each ward
Source: Land Registry OCOD

proprietor is also similar.

A significant difference emerges, however, when comparing the country of incorporation of the companies: while only 40% of the companies are registered to the British Virgin Islands in the OCOD, this number reaches 76% in the identified sample. This could be due to the fact that the leaked files allow me to identify relatively more companies incorporated in the BVI because firms involved in the leaks created a lot of companies in this location.
To test this hypothesis, I turn to the analysis of the Panama Papers as they are the files from which most of the sample has been identified. Table 1 in the appendix shows the distribution of the jurisdictions in which the companies created by Mossack-Fonsecca are incorporated. The British Virgin Islands indeed appear to be a country of preference for incorporation: more than half of the companies created by the firm are located there. If the percentage of companies incorporated in the BVI is distributed evenly across nationalities of the legal owner, the identification process should not be affected. On the other hand, if nationals from particular countries maintain a special relationship with the BVI and I am able to identify them more easily, they could be over-represented in my identified sample.

	Proportion of investors from the BVI
India	85,9%
Israel	53,0%
Jordan	95,6%
Saudi Arabia	88,3%
South Africa	89,7%
UAE	69,3%
UK	83,7%
US	83,0%

TABLE 3: Proportion of companies incorporated in the BVI listed in the Panama Papers, by country of the beneficial owner
Source: Author's calcuations from the Panama Papers

Some countries seem indeed to maintain a closer relationship with the British Virgin Islands. Table 3 presents the percentage of legal owners of a company incorporated in the BVI listed in the Panama Papers, for the countries most represented in the identified sample. Countries like the United Kingdom, and Saudi Arabia seem to own BVI companies relatively frequently. On the contrary, countries like the United Arab Emirates own BVI companies relatively less frequently.

However, it does not appear, that I identify in priority these countries strongly associated to the BVI. In particular, the UK, exhibiting a high percentage of companies incorporated in the BVI according to the Panama Papers, present a less-than-average percentage of companies associated to this country in my identified sample (Table 12, in appendix). Similarly, the UAE, less strongly associated to the BVI in the leaks, shows a high percentage of companies incorporated in the BVI in my sample. The inverse pattern would be expected if I was identifying nationalities that prioritize incorporation in the BVI.

2.3 Risk data

To measure the economic risk in the home countries, I use the index developed by Ahir, Bloom and Furceri (2018), the World Uncertainty Index. It is based on the number of times the word "uncertainty" and its derivatives are used in the Economist Intelligence Unit country reports, reviewing political and economic risks in 143 countries. This source is likely to focus on events

affecting trade and business and to correctly capture the risk perceived by the investors[8]. The index is weighted by the number of words of each report to ensure comparability among countries. It is available quarterly since 1996.

3 Mechanisms of purchasing

3.1 Modelling the behavior of buyers

The aim of this first analysis is to gain a deeper understanding of the mechanisms driving the purchase of real estate in the UK by foreign nationals. There could be a number of different explanations as to why these transactions are taking place. Surveys conducted among wealth advisors and bankers show that reasons to purchase real estate in London essentially include traditional investment purposes like expected capital gains and portfolio diversification, but also safe-haven purposes (Knight_Franck (2016)). Safe-haven effects occur when the perceived risk in the home country increases, which provokes a surge in capital being invested in assets that are considered to be safe and highly liquid, like London real estate. Purchases being made primarily to gain a new residence seem to constitute only a minor part of the transactions. Recovering the motivations driving each group of foreign buyers could help give accurate predictions of the nationality of the beneficial owners of the OCOD.

The timing of the transaction allows me to partly distinguish between those purchases driven by traditional investment purposes and those reflecting safe haven effects. Indeed, if the transaction is made at a time when the perceived risk in the country of the buyer is on the rise, it is likely to be motivated by safe haven considerations. On the contrary, if it is made at a time when the perceived risk in the country of the buyer is decreasing, it is likely to be motivated by more traditional investment considerations, as a safe economic and political environment can dynamize investment.

To recover the likely reasons foreign buyers have to purchase real estate in London, the following specification is estimated:

$$\Delta Vol_{i,t} = \alpha_i + \delta_t + \beta \sum_{c=1}^{n} h_{c,i} z_{c,t-1} + \Delta Vol_{i,t-1} + u_{i,t} \tag{3.1}$$

[8]The website of the Economist Intelligence Unit states that the reports they make "help business leaders prepare for opportunity".

With $Vol_{i,t}$ being the number of transactions made by foreign companies in the area i at a period t. $\Delta Vol_{i,t}$ corresponds to the evolution in the number of purchases between t–1 and t. The coefficient of interest is β, capturing the effects of evolutions of risk z in a country c on the volume of transactions in London. If there are safe haven effects, the coefficient β is expected to be positive. If the number of transactions is driven by more traditional investment motives, the coefficient is expected to be negative. The risk variable is lagged to take into account the time necessary to find and acquire real estate.

The risk variable is weighted by an indicator $h_{c,i}$ of the attractiveness of the area i for investors from a country c. To construct it, I build on Badarinza and Ramadorai (2018). They argue that foreign buyers have preferred habitats in London according to their nationality, meaning that they will prefer to buy properties in certain areas of the city rather than in others. Badarinza and Ramadorai (2018) use the idea that foreign investors exhibit "home bias abroad": they congregate in areas where there is a relatively high population density of their own nationality. This could be due to the cultural and linguistic affinity with the neighborhood, which drives the location choice in a city. It could also be due to the existence of personal or professional networks which makes the purchase easier in certain areas: relatives already located in London or real estate agents and legal firms specialized in the provision of services to specific nationalities. As a result, they use the share of residents of an area i born in country c as a proxy for the attractiveness of i for investors from c. My data allows for a more precise identification of the preferred habitat for each nationality. The identification of a sample of buyers lets me precisely observe the location choices of people from 70 countries. I draw on Cvijanovic and Spaenjers (2015) and use the number of purchases made by investors from a country c in a area i as an indicator of the attractiveness of the area i for the country c. The higher this number, the more popular the neighbourhood is for people from country c[9]. This count is a robust indicator if it provides a non biased picture of the diversity of the locational choices of the foreign investors. It is likely to be the case as the distribution of transactions across London in my identified sample is the same than the one of the full registry. It should be noted that such a variable also captures the effects of the size of the country : if a country is small, it is likely that nationals will buy a lower number of properties in London.

The sum $\sum_{c=1}^{n} h_{c,i} z_{c,t}$ then provides a weighted index of how each area should be impacted by evolutions of risk in foreign countries according to its appeal to buyers from different nationalities. As a control, I also include a lagged dependent variable to allow for persistency

[9]Badarinza and Ramadorai (2018) proxy is likely to be less effective/efficient when used with my dataset. As my analysis focuses on transactions made through shell companies, it is likely to capture a lot of relatively expensive transactions. It is probable that wealthy individuals have different location preferences to others, due to the fact that they can afford to buy in a broader range of areas.

in the number of transactions made in each neighbourhood. Time fixed effects are included to capture the evolutionary trend of the number of purchases made in London. They allow me to make sure that surges in the transactions in the London real estate market are indeed cause by changes in risk and not by idiosyncratic shocks that would affect global investment. Neighborhood fixed effects are also added, to allow for unobserved heterogeneity among areas.

3.2 Results of the estimation

The equation is estimated on the full OCOD, including identified and unidentified observations. The analysis concerns only the period 1999-2018, as most of the transactions are concentrated after 1999 and 2018 is the last year for which complete data is available. Moreover, the country of incorporation of the buying company started to be systematically recorded only in 1999.
It is difficult to disentangle specific country-risk effects on the flows of investment in London, as people coming from countries similar in language and culture tend to buy in the same areas (Badarinza and Ramadorai (2018)). Moreover, the identified sample does not provide enough information on the preferred locations of certain nationalities. As a result, I divide countries into 11 groups according to cultural and linguistic proximity and identify foreign demand effects emanating from these regions rather than specific countries. The groups are the UK, Sub-Saharan Africa, the Middle East, North America and anglophone countries, South East Asia, South Asia, North Africa, Eastern Europe, Southern Europe, Israel and Western Europe. The description the composition of the groups is described in the appendix[10].

Firstly, the equation is estimated without distinguishing specific group effects. The results are presented in table 4. All standard errors are clustered by area to account for serial correlation (Bertrand et al. (2004))[11]. Columns (1) and (2) present the results of the equation taking boroughs as reference areas, while columns (3) and (4) take electoral wards. London is composed of 33 boroughs[12], equivalent to 654 electoral wards[13], which are smaller administra-

[10]Other decompositions according to different criteria could have been relevant. The geographical and cultural division is only one solution among many

[11]Bertrand et al. (2004) show that standard errors are likely to be under-estimated in difference-in-differences analysis, if serial correlation is not taken into account. One of their proposed solutions is to compute the standard errors allowing for a correlation pattern within an area over time. This correction is showed to work well when the number of clusters is high enough. It is likely to be the case in my empirical application, as I have at least 33 clusters.

[12]It technically counts 32 boroughs, plus the City of London. For simplicity, I consider the City to be a borough as well.

[13]The division in wards is current as of 2015. Small changes can have occurred since then.

	General effects			
	(1)	(2)	(3)	(4)
Weighted Risk	0,170***	0,092***	0,101**	0,095**
	(0,015)	(0,015)	(0,040)	(0,039)
$\Delta Volume_{t-1}$		0,281***		0,039
		(0,044)		(0,068)
Time FE	Yes	Yes	Yes	Yes
Neighborhood FE	Yes	Yes	Yes	Yes
N of clusters	33	33	654	654
R	58,33%	70,06%	17,0%	19,9%

TABLE 4: Gross results, not decomposed by group
Note: the standard errors are presented under the coefficients. * corresponds to a coefficient significant at 10%, ** at 5% and *** at 1%
Source: Land Registry OCOD and observations identified from the leaked files.

tive divisions. Dividing the city in boroughs allow to make sure to have information on the attractiveness of each area, as the identified sample counts purchases in almost every borough for each group of countries. Nevertheless, using the wards could provide a more granular identification of preferred habitats if identified transactions cover a large enough geographical span for each group of buyers.

The coefficient of the risk cannot be interpreted directly because it has been weighted by a proxy of the desirability of the neighborhood. However, the weighted risk index is always positive and significant, whether I add the lag of the dependent variable as a control or not and whether I estimate the equation on London boroughs or electoral wards. These results indicate that London real estate is globally considered as a safe-haven asset, which increases in demand even during times of economic and political uncertainty. Its high value and high liquidity allow it to act as a "safe deposit box" (Fernandez et al. (2016)). The results are in line with those of Badarinza and Ramadorai (2018) and Sà (2016) who find that increases in risk in foreign countries affect positively the flows of capitals coming to London real estate.

It can be noted that the lagged dependent variable is not significant when the equation is estimated taking electoral wards as areas of reference. It could be due to the fact that OCOD records no transactions in a year for a significant number of wards, as they are fairly small areas. It could also explain why the equation estimated on wards appear to have less explanatory power than when estimated on boroughs.

Following this, I estimate specific group effects, to allow buyers to be driven by different motivations when purchasing real estate in London. Firstly, I separately fit one equation for each one

of the eleven groups of countries. The risk coefficients thus estimated are presented in column (1) and (3), respectively for boroughs and wards of table 5. But economic and political risk is likely to be correlated among the different regions of the world. To take that into account, I estimate a variant of equation (3.1) which allow for the risk coefficients to be different across regions:

$$\Delta Vol_{i,t} = \alpha_i + \delta_t + \sum_{c=1}^{n} \beta_c h_{c,i} z_{c,t-1} + \Delta Vol_{i,t-1} + u_{i,t} \tag{3.2}$$

The results for equation (3.2) are presented in column (2) of table 5. The findings are similar for the two specifications. Controlling for the country-correlation of uncertainty causes more coefficients to be insignificant in equation (3.2). However, the signs of the effects estimated are all similar, except for the North America and anglophone countries group.

Taking a closer look at the results of equation (3.2), it appears that only a part of the countries exhibits safe-haven effects: the UK, the Middle-East, North Africa, Eastern Europe and Western Europe. On the contrary, for South-East Asia, South-Asia, South America, Southern Europe and Israel, an increase in risk is linked to a decrease in the flow of capital going to London real estate.

These findings can be compared to the typology of overseas buyers of UK's real estate established in McCarvill et al. (2012). They distinguish three categories:

- European and North American buying to gain a residence or for investment purposes
- East Asian buying to rent and for capital gains
- Non-OECD countries nationals buying for economic and political security purposes, particularly the UAE and Eastern Europe (Heywood (2012))

The first two categories correspond to purchases made for traditional investment purposes, while the third corresponds to safe haven motivations. The signs obtained in the two group-by-group specifications are coherent with this decomposition: the positive coefficients obtained for North America, South-Asia, South-East Asia and Israel are consistent with the idea that they buy in priority for capital gains and to rent. The negative coefficients obtained for Sub-Saharan Africa, the Middle-East, North Africa and Eastern Europe are consistent with wealthy individuals in highly risky countries storing their capital in London real estate for safety. The negative coefficient of the UK could be explained by the fact that a rise in uncertainty could lead to the more frequent use of shell companies to buy properties. The positive coefficient found for Western Europe, indicating the existence of safe haven effects, is surprising in regard with this

typology. As Western Europe is a relatively safe region not suffering from high uncertainty, a negative effect of risk on transactions would be expected. One explanation for this could be because most of the traditional investment transactions do not take place through shell companies, so it's possible that there could be a criminal element to these types of activities. Another reason could be that Western Europe contains some offshore centres, such as parts of Belgium and Austria which are included in my analysis. If the increase in risk in non OECD countries increases the number of shell companies incorporated in them, we could also see a rise in the number of transactions registered to them in the UK.

	(1)	(2)	(3)	(4)
Home Country	0,889***	0,826	0,637**	0,349**
	(0,254)	(0,509)	(0,300)	(0,193)
Sub-Saharan Africa	0,237	0,556	0,109**	0,035
	(0,173)	(0,656)	(0,55)	(0,049)
Middle-East	0,259***	0,627**	0,262***	0,161***
	(0,035)	(0,261)	(0,072)	(0,060)
North America	0,027	-1,124**	0,069***	-0,019
	(0,067)	(0,460)	(0,021)	(0,017)
South-East Asia	-4,843**	-5,481**	-0,712*	
	(2,168)	(2,655)	(0,397)	(0,411)
South Asia	-9,648*	-11,536*	-0,983	-1,509
	(4,868)	(6,755)	(1,156)	(1,519)
North Africa	3,760***	1,181	2,283	1,773
	(0,593)	(1,276)	(1,864)	(1,609)
Eastern Europe	3,683***	(2,538)	1,935***	0,853
	(0,568)	(1,669)	(0,603)	(0,652)
Southern Europe	-0,880	-8,694	2,889**	0,643
	(1,755)	(6,331)	(1,279)	(0,674)
Israel	-5,048	-4,639	-1,235*	-1,041*
	(3,705)	(3,362)	(0,737)	(0,567)
Western Europe	2,789*	(3,700)	4,450***	3,130***
	(1,606)	(3,706)	(1,709)	(1,211)

TABLE 5: Results, decomposed by group
Note: the standard errors are presented under the coefficients. * corresponds to a coefficient significant at 10%, ** at 5% and *** at 1%
Source: Land Registry OCOD and observations identified from the leaked files

3.3 Extension

To test whether the overall safe haven effects estimated to take place in London is consistent across the whole country, I estimate equation (3.1) on all the UK's transactions registered in the OCOD. The results are presented in 14 in the appendix. Column (1) and column (2) presents

the coefficients when districts and counties are taken as areas of reference respectively, counties corresponding to a large administrative division. The weighted risk coefficient is negative and insignificant in three of the specifications tested. It is positive and significant when the equation is estimated on counties, but the result is not robust to the addition of proper controls. These findings indicate that real estate in the rest of the country does not act as a safe-haven asset. They are different than Sà (2016)'s, who observes an increase in transaction in the UK as a whole when foreign risk increases. However, my results appear coherent: London is a global city, which attracts a lot of capital. There is no reason why the real estate in the rest of the country, which does not exhibit the same characteristics of high prices and high liquidity, should be a safe haven asset as well.

4 Predictions of the nationality of the beneficial owners

4.1 Technical specifications

The mechanisms presiding at the purchase of real estate in London by foreign investors are different according to the nationality of the buyer. While some acquisition behaviors are coherent with traditional investment purposes, some suggest that London real estate can be used as a safe haven asset. This finding can be used to recover the nationality of the beneficial owners that could not be identified from the leaked documents. They show that the timing of the transaction is crucial to determine who is buying the property. Location of the transaction is also strongly associated with the nationality of the buyer: nationals from a certain country have preferred habitats in London, partially identified with the Panama Papers and the other leaked files. With this insight, I propose to predict the group of nationality of all the buyers in London's OCOD. The geographic and cultural group is predicted rather than the nationality of the beneficial owners. Indeed, the identified sample does not provide enough information to precisely recover specific nationality patterns.Even if the Panama Papers had permitted to identify more companies, it is likely that nationality effects would have overlap in some way. Indeed, uncertainty is correlated among countries having strong political and economic risks, and nationals from these countries are likely to locate near one another in London (Badarinza and Ramadorai (2018)). Trying to disentangling national effects would likely not yield robust results. Moreover, the set of potentially predicted countries would be restricted to the 70 ones identified by the Panama Papers. Classifying the observations in groups allows the coverage of all the regions of the world, even if the process is less precise.

Predictions are made with machine learning techniques. They have been showed to provide

good classification accuracy in large samples, by making use of all the patterns recovered between the different variables. However, most often they do not allow to recover exactly the mechanisms allowing to carry such predictions (Witten *et al.* (2013)). As the aim of this second analysis is only to get the best picture of the nationality of the buyers, it does not appear to be a problem.

Tree-based methods are used in this analysis. A classification tree groups observations in non-overlapping regions at terminal nodes, according to the value of independent variables which allows to construct the several internal nodes of the tree. Among one terminal region, every observation is then predicted to be of the same class. Decision trees are particularly flexible and handle well multiclass predictions. Moreover, combining several classification trees can drastically improve prediction accuracy.

Three tree-based methods are tested. Firstly, I predict the nationality of my unlabelled sample using a random forest. A random forest corresponds to an ensemble of N trees constructed from N bootstrapped samples of the original data. The trees are decorrelated by only allowing to use a reduced number of randomly drawn predictors to split the predictor space. The prediction for an observation then corresponds to the most recurrent prediction among the trees. An interesting feature of random forests is that they provide an estimation of the classification error of the predictor. Indeed, each tree is built from a bootstrapped sample which is shown to correspond to about a third of the original sample (Witten *et al.* (2013)). The classification tree can be tested on those observations not used in its construction, allowing the computation of the out-of-bag error estimate (OOB) measuring the number of times an observation is wrongly predicted. The OOB gives an indication of the overall classification error.

However, random forests only use the identified sample to carry predictions. As it amounts to about 3% of London's OCOD, it is likely to be insufficient to yield satisfactory predictions for the whole sample. Moreover, the unlabelled sample is a source of additional information that needs to be exploited. It provides more details about the links between date and location of the transaction, and nationality of the buyer.

As a result, I turn to semi-supervised techniques for the rest of the analysis. While supervised predictors are constructed strictly from an identified sample, semi-supervised machine learning makes use of the pattern and information present in the unlabelled sample as well. These additional elements can help to build more efficient classifiers and to get a better prediction accuracy for the unidentified proprietors.

The first semi-supervised technique I use is a version of a self-training random forest. Self-training is a method consisting of adding to the labelled sample the observations of the uniden-

tified sample which labels are confidently predicted. Practically, I start by constructing a current random forest of 500 trees on the labelled sample. The classifier is then applied to the unidentified sample and if 2/3 of the trees grown agree on a label for an observation, this observation and its label are added to the identified sample. A new current classifier is then constructed from this new sample, built from the original labelled observations and the confidently predicted one. If this classifier has a lower OOB error rate when it is estimated on the originally identified sample, it becomes the best classifier. This feature ensures that the self-training method improves prediction accuracy. These steps are repeated until no more observations can be confidently predicted or until the unlabelled sample is empty. One problem with this self-training method is that the threshold determining confidently predicted observations appears to be quite arbitrary: 2/3 is chosen in this work but one could also opt for a simple majority rule or use a higher level.

To overcome this issue, I use a modified version of a self-training random forest algorithm developed by Chau and Phung (2016), minSemi_RF. The algorithm is based on the predictions of a random forest of three trees. A first set of predictions is made using only the certain information contained in the labelled sample. Then, if all three trees agree on a label for an observation of the unidentified sample, this observation is added to the identified sample and considered a to be true. A new random forest is then ran making use of those newly identified observations. Again, the predictor thus obtained is becoming the current best predictor only if it yields a lower out-of-bag error estimation than the previous one when estimated on the original identified sample. This operation is repeated until all the observations have been identified or until no more observations can be added to the identified sample. The details of the algorithm are described in Chau and Phung (2016). Compared to the classical self-training method, minSemi_RF has the advantage to be parameter-free: there is no need to set a threshold to determine what corresponds to a confidently predicted label.

4.2 Prediction Results

In addition to the timing and the location of the transaction, several other pieces of information are exploited to recover the nationality of the buyers. Firstly, I add a dummy variable capturing whether the company buying is incorporated in the British Virgin Islands, as it concerns about 40% of the whole sample and 76% of the identified one. Then, I add a variable taking 1 if the price paid for the property is available in the OCOD, 0 otherwise. Indeed, some nationalities could be more strongly associated with the probability to provide the price when registering

the transaction.

Table 6 presents the results for the three methods described, respectively in column (2), (3) and (4). It shows the percentage of buyers estimated to belong to each group. Column (1) shows the percentages of buyers of each group identified in the leaked files, for comparison. For Random Forest, the number of variables tested at each split is chosen by selecting the number yielding the lowest OOB error rate when estimated on the original sample. The results of these comparisons are presented in table 15 in the appendix. For minSemi_rf, I follow Chau and Phung (2016) and randomly try $log(p) + 1$ variables at each node, p being equal to the number of independent variables used in the estimation.

As it is showed by Chau and Phung (2016) to give better results than simple random forests and it is parameter free, my preferred specification is minSemi_rf. It is though useful to note that for all three specifications, the United Kingdom and the Middle-East are identified to be the most frequent nationalities among beneficial owners. The proportions predicted are always superior to the ones obtained in the identified sample, suggesting that the identification procedure led to an underestimation of the percentage of buyers from those countries. This decomposition is coherent with the one obtained by Alstadsæter et al. (2019). They find that the country owning the highest level of financial offshore wealth compared with its GDP is the UAE, the country most represented in the Middle-East sample. The UK comes in 14th position, but as I am focusing on British real estate the over-representation of beneficial owners from this country appears coherent.

	Predictions of the region of origin of the buyers			
UK	26,2%	46,1%	34,8%	32,9%
Sub-Saharan Africa	7,5%	3,6%	3,4%	4,3%
Middle-East	37,2%	37,3%	48,5%	43,6%
North America	11,8%	3,8%	3,1%	6,9%
South-East Asia	3,0%	0,9%	1,1%	2,0%
South Asia	4,6%	1,0%	3,5%	3,5%
North Africa	1,9%	0,8%	0,4%	0,8%
Eastern Europe	1,3%	1,7%	0,9%	1,0%
Southern Europe	1,4%	0,8%	1,9%	0,7%
Israel	4,0%	1,7%	1,9%	3,9%
Western Europe	1,2%	1,0%	0,4%	0,6%

TABLE 6: Prediction of the group of nationality of buyers of London real estate through shell companies
Note: The first column reports the results of the group distribution in the identified sample. The second column reports the results of the random forest specification. Columns (3) and (4) respectively reports the results of self-training and minSemialgorithms. The sum of the columns may not be one due to rounding.
Source: own calculations from the OCOD

The other classes are predicted less frequently, with a classification percentage never being superior to 10%. MinSem_RF gives more weight to them compared to the other classification. While with the Random Forest and the Self Training none of these groups are predicted to compose more than 5% of the beneficial owners of the OCOD sample, buyers from North America represent 6,9% of the observations when estimated through minSemi_RF.
A surprising result is the small number of transactions attributed to Western and Northern European countries. Indeed, geographical proximity is an important factor to explain the amount of offshore wealth held in a country (Alstadsæter et al. (2019)). It could mean that citizens from these countries have no interest in London real estate, because they already have an easy access to the UK. However, the cumulated OCOD reports that about 5% of the traditional companies[14] are incorporated in those countries, a lot more than my estimations, showing that they invest a lot in British real estate lot of capitals in British real estate. These findings could indicate the existence of a substitution effect between purchases through the traditional channels and the purchases through shell companies. This result is noteworthy, as it implies that

[14]Traditional means not incorporated in a tax haven here.

the group of people using special purpose vehicles and the group of people using traditional channels overlap in some way.

To finish, it must be noted that the machine learning techniques used can predict different outcomes for the same company. Indeed, one company can own multiple properties, and therefore can appear in more than one transaction. As about 14% of the companies identified with the Panama Papers have beneficial owners of different nationalities, it can be justified. Nevertheless, to check the robustness of the results, I run a minSemi_RF specification, constraining the outcome for each company to be the most recurrent group prediction for that particular company. The results are presented in table 16 in the appendix, for the two self-training algorithms. They are similar to the ones obtained without applying any constraint to the process. They seem to give more weight to the UK and the Middle-East, however. As I am interested in capturing the diversity of the beneficial owners of London real estate, the non-constrained method appears more appropriate.

5 Estimation of offshore wealth

The identification of the nationality groups of the beneficial owners of the companies listed in the OCOD do not reflect perfectly the amount of capital each country owns in London real estate. Indeed, the dataset records very different transactions, from parking spaces to entire buildings, which are made in areas exhibiting very distinct property prices. This last analysis is aimed at proposing an estimation of the prices of all the purchases, to allow for a gross estimation of the offshore wealth owned through British real estate. Again, I conduct two sets of transactions, one for the whole sample from 1999 to March 2019, and the other for the March 2019 version of the dataset.

To get a first partial picture of the value of the transactions concluded, I study the prices paid for each property as reported in the OCOD. It is not mandatory to state the price paid for a transaction though, so the information is available for only 39,5% of the observations in the cumulative dataset, and for 36,6% of the current dataset[15].

Following this, I try to compute the missing prices based on the location and the date of the transaction. I use the mean housing price of the ward the year the transaction was concluded

[15]The investigation magazine Private Eye has published an interactive map where the readers can give information on the prices paid for each property owned by offshore companies. It allows to complete the sample of prices a bit. However, I do not ue it in the analysis as it does not correspond to official data.
https://www.private-eye.co.uk/registry

as an indicator of the price paid for the property. The mean price by ward data is recovered from the Office for National Statistics, who constructs the measure based on comprehensive administrative datasets of real estate transactions. It is likely to be an underestimate of the price actually paid, because the overseas buyers operating through shell companies tend to locate at the high-end of the property market. Two issues are raised by this imputation. First, the OCOD records undistinctively commercial and residential properties while the mean price by ward is computed using only residential properties. Residential real estate is usually more expensive than commercial one, which could bias the estimation. Since 2013 however, the British government allows the transformation of offices in residential housing, increasing the prices in both markets and blurring the frontier between them (Fernandez et al. (2016)). The price gap between commercial and residential real estate is then likely to be less pronounced after 2013. The second issue posed by this estimation is that the OCOD does not record completely harmonized transactions. Some observations correspond to only a flat or even a parking space, while some others act the purchase of entire buildings. Moreover, if the company buy several properties at a time, the purchase might be gathered in only one transaction. As a result, the actual prices paid for what is recorded as one sale can differ widely, which will not be captured in the imputated price.
The results for the whole sample are reported in table 7. Column (1) corresponds to the first method while column (2) includes the imputed prices. The results for the current sample are reported in table 8. The computed values mirror the results find in the precedent section, with the UK and the Middle East owning the bulk of the offshore wealth stored in British real estate. In total, when considering only the current sample, I find that almost £75 billion are held by shell companies through London real estate nowadays. About a third of this wealth is actually owned by British citizens, probably using the vehicle of shell companies for tax evasion purposes. More than 40% are owned by countries from the Middle-East, principally the United Arab Emirates.

It is interesting to compare these results with the country-by-country estimation of financial wealth proposed by Alstadsæter et al. (2019). To do that, I compute 11 groups composed of the 70 countries I used in my estimation. I report the results obtained in 9. The amounts themselves are not comparable, as the estimations of Alstadsæter et al. (2019) are global whereas mine only concern one country. However, in both estimations, the Middle-East and the UK own a significant part of offshore wealth. The percentage is much larger when considering real estate though. As I said in section 3, it is coherent for the UK. It could reflect a particular

	Present Prices	Imputed Prices
UK	29,962	37,542
Sub-Saharan Africa	2,613	3,624
Middle-East	38,853	56,497
North America	6,524	7,729
South-East Asia	2,478	2,851
South-Asia	4,807	5,437
North Africa	580	1,172
Eastern Europe	3,285	3,663
Southern Europe	486	714
Israel	4,728	5,483
Western Europe	784	1,066
Total	**95,100**	**125,778**

TABLE 7: Estimation of the wealth held through British real estate, whole sample, in million of £

taste of wealthiest Middle-Eastern people for real estate however, and in particular for London real estate. It could also mean that they are more prone to use shell companies, thus being under-represented in the estimations of financial offshore wealth. My data do not allow to disentangle these two effects. Three groups appear notably more represented in the estimations of financial offshore wealth than in my computations: North America, Southern Europe and Western Europe. Similarly, it could indicate that people from these countries prefer to store offshore wealth in financial assets rather than in real estate, or that they do not use extensively shell companies. These comparisons are then indicative of the preferred assets to store offshore wealth for each region of the world, and of which countries use shell companies more instensively.

	Value estimation, group-by-group	
	Present Prices	Imputed Prices
Home Country	17,313	22,775
Sub-Saharan Africa	1,857	2,602
Middle-East	21,882	34,310
North America	4,393	5,255
South-East Asia	1,183	1,455
South-Asia	1,699	2,157
North Africa	378	822
Eastern Europe	1,122	1,380
Southern Europe	192	339
Israel	2,401	2,937
Western Europe	440	667
Total	**52,860**	**74,699**

TABLE 8: Estimation of the wealth held through British real estate as of March 2019 sample, in million of £
Source: Calcuations from the Land Registry OCOD

Group	Financial wealth (in billion of £)
Home Country	393,9
Sub-Saharan Africa	41,175
Middle-East	524,775
North America	930,825
South-East Asia	167,85
South Asia	42,975
North Africa	53,55
Eastern Europe	150,75
Southern Europe	391,575
Israel	59,85
Western Europe	864,45
Total	3621,675

TABLE 9: Alstadsæter et al. (2019) estimates of financial offshore wealth, group-by-group, in bilions of pounds
Source: Alstadsæter et al. (2019) calculations
Note: the estimates have been converted in £, adopting the echange rate 1$USD = £0,75

6 Conclusion

I demonstrated that the location of a real estate transaction in London and the timing of this transaction can be used to predict the nationality of the buyer, even if she is obscuring her identity through a shell company. These predictions are used to estimate the amount of offshore wealth stored in British real estate. I find that about £75 billion of pounds can be attributed to shell companies, which is to be considered as a lower-bound as I made conservative hypotheses. The group-by-group decomposition shows that about a third of this wealth is actually owned by British citizens, while 40% is attributed to Middle-East countries. The comparison of these findings with the estimations of financial wealth gives an insight into the preferred assets of each population group to own offshore wealth. Further research would be needed to disentangle whether a region owns relatively more in British real estate than in financial offshore wealth because of the nature of the assets or because its citizens use shell companies more intensively.

These findings can inform certain policy debates. I showed that the motives to buy British real estate are diverse and seem to diverge among the groups of investors. Digging deeper into these motives could shed light on the best ways to regulate foreign investment flooding the capital. Moreover, it is interesting to not that the Land Registry data combined with my estimations allows to construct time series of the offshore wealth held in London real estate, region by region. Studying these evolutions could shed light on the institutional, economic and political environment favoring the development of offshore wealth.

References

Alstadsæter, A., Johannesen, N. and Zucman, G. (2018), 'Who owns the wealth in tax havens? macro evidence and implications for global inequality', *Journal of Public Economics* **162**, 89–100.

Alstadsæter, A., Johannesen, N. and Zucman, G. (2019), 'Tax evasion and inequality', *American Economic Review* **109**(6), 2073–2103.

Badarinza, C. and Ramadorai, T. (2018), 'Home away from home? foreign demand and london house prices', *Journal of Financial Economics* **30**(3), 532–555.

Bayer, R., Hodler, R., Raschky, P. and Strittmatter, A. (2018), 'Expropriations, property confiscations and new offshore entities: Evidence from the panama papers', *Working Paper* .

Bertrand, M., Duflo, E. and Mullainathan, E. (2004), 'How much should we trust differences in differences?', *Quarterly Journal of Economics* **119**, 249–75.

Caruana-Galizia, P. and Caruana-Galizia, M. (n.d.), 'Offshore financial activity and tax policy: evidence from a leaked data set', *Journal of Public Policy* **36**(03), 457–488.

Chau, V. and Phung, N. (2016), 'A random forest-based self-training algorithm for study status prediction at the program level: minsemi-rf', *CFM discussion paper series (CFM-DP2016-39).* **10053**, 219–230.

Cvijanovic, D. and Spaenjers, C. (2015), 'Real estate as a luxury good: Non-resident demand and property prices in paris'.

D., W. and T., T. R. H. (2013), 'An introduction to statistical learning', *Transparency International Report* .

Fernandez, R., Hofman, . and Aalbers, M. (2016), 'London and new york as a safe deposit box for the transnational wealth elite', *Environment and Planning* **48**(12).

Henry, J. (2012), 'The price of offshore revisited: New estimates for missing globalprivate wealth, income, inequality, and lost taxes', *Tax Justice Network Working Paper* .

Heywood, A. (2012), 'London for sale? an assessment of the private housing market in london and the impact of growing overseas investment'.

Johannesen, N. (2014), 'Tax evasion and swiss bank deposits', *Journal of Public Economics* **111**, 46–62.

Johannesen, N. and Zucman, G. (2014), 'The end of bank secrecy? an evaluation of the g20 tax haven crackdown', *American Economic Review* **6**(1), 65–91.

Knight_Franck (2016), 'The wealth report 2016'.

McCarvill, P., Gaffney, D. and Griffith, M. (2012), 'Addordable capital? housing in london', *Institute for Public Policy Research Report* .

Omartian, J. (2017), 'Tax information exchange and offshore entities: Evidence from the panama papers', *Working Paper* .

Reumer, C. (2019), 'Understanding money laundering through real estate transactions', *European Parliament Briefings* .

Scanlon, K., Whitehead, C. and Blanc, F.and Moreno-Tabarez, U. (2017), 'The role of overseas investors in the london new-build residential market', *LSE Report* .

Sotura, A. (2011), 'Les étrangers font-ils monter les prix de l'immobilier? estimation à partir de la base de la chambre des notaires de paris, 1993-2008.', *Mémoire de master, PSE* .

Sà, F. (2016), 'The effect of foreign investors on local housing markets: evidence from the uk'.

Transparency_International (2012), 'Corruption on your doorstep:: how corrupt capital is used to buy property in the uk', *Transparency International Report* .

Treasury, H. (2017), 'Economic crime - anti-money laundering supervision and sanctions implementation', *Publications of the House of Commons, Treasury Committee* .

Zucman, G. (2013), 'The missing wealth of nations, are europe and the u.s. net debtors or net creditors?', *Quarterly Journal of Economics* **128**(3), 1321–1364.

7 Appendix

7.1 Decomposition of the groups of countries

1 - Home countries: United Kingdom, Ireland

2 - Sub Saharan Africa: Botswana, Cameroon, DR Congo, Gabon, Gambia, Ghana, Kenya, Malawi, Mali, Namibia, Nigeria, Sierra Leone, Tanzania, Uganda, Zambia, Zimbabwe

3 - Middle East: Azerbaidjan, Iran, Iraq, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, Yemen

4 -North America and English-speaking world: Australie, Canada, Jamaica, New Zealand, South Africa, US

5 - South East Asia: Brunei, China, Indonesia, Taiwan, Thailand, Philippines

6 - South Asia: India, Pakistan

7 - North Africa: Egypty, Libya, Morocco

8 - Eastern Europe: Albania, Bulgaria, Czech Republic, Estonia, Hungary, Kazakhstan, Lithuania, Poland, Russia, Ukraine

9 - Southern Europe and South America: Argentina, Brazil, Greece, Italy, Portugal, Spain

10 - Israel: Israel

11 - Western Europe: Austria, Belgium, France, Germany, Ireland, Netherlands,Sweden

7.2 Tables and Graphs

Country	%	Country	%
Albania	0,1%	Kuwait	1,6%
Argentina	0,04%	Libya	0,1%
Australia	1,0%	Lithuania	0,02%
Austria	0,04%	Malawi	1,3%
Azerbaijan	0,08%	Mali	0,04%
Belgium	0,1%	Morocco	0,8%
Botswana	0,5%	Namibia	0,2%
Brazil	0,2%	Netherlands	0,1%
Brunei	1,2%	New Zealand	0,2%
Bulgaria	0,06%	Nigeria	1,6%
Cameroon	0,1%	Oman	0,7%
Canada	0,5%	Pakistan	1,1%
China	0,4%	Philippines	0,02%
Czech Republic	0,02%	Poland	0,06%
DR Congo	0,04%	Portugal	0,4%
Egypt	0,9%	Qatar	0,6%
Estonia	,04%	Russia	0,%
France	0,6%	Saudi Arabia	6,8%
Gabon	0,02%	Sierra Leone	0,2%
Gambia	0,04%	South Africa	7,6%
Germany	0,1%	Spain	0,1%
Ghana	0,04%	Sweden	0,1%
Greece	0,3%	Syria	0,02%
Hungary	0,2%	Taiwan	0,2%
India	3,5%	Tanzania	0,2%
Indonesia	0,3%	Thailand	0,9%
Iran	0,2%	Turkey	0,1%
Iraq	0,1%	Uganda	0,06%
Ireland	0,5%	Ukraine	0,04%
Israel	4,0%	UAE	22,6%
Italy	0,3%	United Kingdom	25,6%
Jamaica	0,08%	United States	2,4%
Jordan	4,2%	Yemen	0,08%
Kazakhstan	0,02%	Zambia	0,1%
Kenya	2,2%	Zimbabwe	0,9%

TABLE 10: Proportion of proprietors by nationality in the identified sample

	Proportion of incorporated companies
Bahamas	7,5%
Belize	<1%
British Anguilla	1,5%
BVI	53,2%
Costa Rica	<1%
Cyprus	<1%
Hong Kong	<1%
Isle of Man	<1%
Jersey	<1%
Malta	<1%
Nevada	<1%
New Zealand	<1%
Niue	4,4%
Panama	22,6%
Ras Al Khaimah	<1%
Samoa	2,5%
Seychelles	7,1%
Singapore	<1%
UK	<1%
Uruguay	<1%
Wyoming	<1%
Mean number of properties/company	2,14

TABLE 11: Proportion of companies incorporated for each country, Panama papers

Country	% of BVI transactions		
Argentina	100%	Libya	33%
Australia	88%	Lithuania	100%
Albania	100%	Kuwait	90%
Austria	100%	Malawi	6%
Azerbaijan	100%	Mali	100%
Belgium	43%	Morocco	100%
Botswana	0%	Namibia	100%
Brazil	100%	Netherlands	60%
Brunei	100%	New Zealand	100%
Bulgaria	100%	Nigeria	71%
Cameroon	100%	Oman	100%
Canada	76%	Pakistan	55%
China	95%	Philippines	100%
Czech Republic	100%	Poland	100%
DR Congo	100%	Portugal	95%
Egypt	80%	Qatar	93%
Estonia	0%	Russia	89%
France	79%	Saudi Arabia	88%
Gabon	100%	Sierra Leone	100%
Gambia	100%	South Africa	87%
Germany	100%	Spain	33%
Ghana	100%	Sweden	0%
Greece	93%	Syria	100%
Hungary	9%	Taiwan	90%
India	89%	Tanzania	100%
Indonesia	100%	Thailand	100%
Iran	8%	Turkey	57%
Iraq	67%	Uganda	100%
Ireland	44%	Ukraine	100%
Israel	62%	UAE	85%
Italy	47%	United Kingdom	58%
Jamaica	100%	United States	66%
Jordan	99%	Yemen	100%
Kazakhstan	100%	Zambia	0%
Kenya	94%	Zimbabwe	100%

TABLE 12: Percentage of companies incorporated in the BVI

Table 13

	Results (1)			
	(1)	(2)	(3)	(4)
Weighted Risk	-0,004	- 0,05	0,007***	-0,008
	(0,007)	(0,007)	(0,000)	(0,009)
$\Delta Volume_{t-1}$		0,089		0,426*
		(0,059)		(0,244)
Time FE	Yes	Yes	Yes	Yes
Neighborhood FE	Yes	Yes	Yes	Yes
N of clusters	349	349	119	119
R	4,3%	0,6%		61,9%

Table 14: Results of the equation estimated on the UK

N of variables	OOB error rate
1	26,85%
2	21,26%
3	21,50%
4	21,42%
5	21,55%
6	21,55%
7	21,65%
8	21,55%
9	21,28%
10	21,32%
11	21,36%
12	21,24%
13	20,86%
14	21,17%
15	21,24%
16	21,07%
17	20,97%
18	21,05%
19	21,03%
20	21,15%
21	21,11%
22	21,01%
23	20,95%
24	20,88%
25	21,07%
26	21,05%
27	20,89%
28	21,13%
29	20,9%
30	21,17%
31	20,87%
32	21,01%
33	20,93%
34	20,89%
35	21,07%
36	20,99%

TABLE 15: Estimated OOB error rate for different number of variables tried at each node

	Predictions of the region of origin of the buyers, constrained method	
	Self-T.	minSemi_RF
UK	39,4%	42,7%
Sub-Saharan Africa	2,9%	6,1%
Middle-East	47,5%	38,6%
North America	2,3%	4,00%
South-East Asia	0,9%	1,8%
South Asia	3,1%	1,7%
North Africa	0,4%	0,9%
Eastern Europe	0,6%	0,9%
Southern Europe	1,3%	1,1%
Israel	1,2%	1,7%
Western Europe	0,2%	0,6%

TABLE 16: Prediction of the group of nationality of buyers of London real estate through shell companies
Note: the sum of the columns may not be one dur to rounding.
Source: own calculations from the OCOD