Published in final edited form as:

J Public Econ. 2011 February 1; 95(1-2): 177–188. doi:10.1016/j.jpubeco.2010.10.006.

The Effect of Capital Gains Taxation on Home Sales: Evidence from the Taxpayer Relief Act of 1997

Hui Shan^{*}

Hui Shan: hui.shan@frb.gov

^{*} Federal Reserve Board of Governors

Abstract

The Taxpayer Relief Act of 1997 (TRA97) significantly changed the tax treatment of housing capital gains in the United States. Before 1997, homeowners were subject to capital gains taxation when they sold their houses unless they purchased replacement homes of equal or greater value. Since 1997, homeowners can exclude capital gains of \$500,000 (or \$250,000 for single filers) when they sell their houses. Such dramatic changes provide a good opportunity to study the lock-in effect of capital gains taxation on home sales. Using 1982–2008 transaction data on single-family houses in 16 affluent towns within the Boston metropolitan area, I find that TRA97 reversed the lock-in effect of capital gains taxes on houses with low and moderate capital gains. Specifically, the semiannual sales rate of houses with positive gains up to \$500,000 increased by 0.40–0.62 percentage points after TRA97, representing a 19–24 percent increase from the pre-TRA97 baseline sales rate. In contrast, I do not find TRA97 to have a significant effect on houses with gains above \$500,000. Moreover, the short-term effect of TRA97 is much larger than the long-term effect, suggesting that many previously locked-in homeowners took advantage of the exclusions immediately after TRA97. In addition, I exploit the 2001 and 2003 legislative changes in the capital gains tax rate to estimate the tax elasticity of home sales during the post-TRA97 period. The estimation results suggest that a \$10,000 increase in capital gains taxes reduces the semiannual home sales rate by about 0.1–0.2 percentage points, or 6–13 percent from the post-TRA97 average sales rate.

Keywords

Housing; Taxation; Capital gains

1 Introduction

In the United States, capital gains are generally taxed upon realization and appreciated assets are not subject to taxation when transferred by bequest. These features of capital gains taxation may lead individuals to hold their assets for a longer time than they otherwise would. Economists have long recognized such a potential lock-in effect of capital gains taxation in financial markets. However, very few empirical studies have examined the lock-in effect of capital gains taxation in housing markets. The Taxpayers Relief Act of 1997 (TRA97) has generated the largest changes in the tax treatment of housing capital gains

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

since the late 1970s, and therefore, serves as a natural experiment for researchers to study the impact of capital gains taxation on housing markets.

Prior to TRA97, homeowners had to pay capital gains taxes when they sold their homes unless they resorted to the “roll-over rule” or the “age-55 rule.” The roll-over rule allowed a home seller to postpone his capital gains provided that he bought another home of equal or greater value within two years. The age-55 rule allowed home sellers of age 55 or older to claim a one-time exclusion of \$125,000 against their capital gains. The pre-TRA97 capital gains taxation had been criticized for its complexity and potentially large distortions of homeowners’ mobility and housing consumption decisions. For example, Burman, Wallace and Weiner (1996) showed that the pre-TRA97 capital gains taxation discouraged renting and moving to less expensive homes while raising little revenue.

TRA97 abolished both the roll-over rule and the age-55 rule. Instead, homeowners can exclude capital gains of \$500,000 (or \$250,000 for single filers) when they sell their homes after TRA97, and they can potentially claim such an exclusion as often as every two years. Using public survey data, Farnham (2006), Biehl and Hoyt (2007), and Cunningham and Engelhardt (2008) find evidence suggesting that capital gains taxes during the pre-TRA97 period locked in many homeowners and that TRA97 released such lock-in effects. For example, Cunningham and Engelhardt (2008) showed that the mobility rate of under-55 homeowners increased significantly after TRA97.

In this paper, I use housing transaction data to study the effect of capital gains taxation on home sales. More specifically, I construct a panel of single-family houses using the 1982–2008 sales records and ZIP code level house price indices in 16 affluent cities and towns within the Boston metropolitan area. The data set does not have information on individual characteristics such as age, income, and marital status, but it has accurate information on the dates and prices of home sales. To identify the effect of capital gains taxation on home sales, I exploit the cross-sectional variation in accumulated capital gains and the arguably exogenous change in exclusion levels introduced by TRA97. I also exploit legislative changes in capital gains tax rates in 2001 and 2003 to estimate the tax elasticity of home sales during the post-TRA97 period. This paper contributes to the existing literature by using a unique data set and exploiting sources of variation different from previous research.

A number of interesting findings emerge from my analysis. First, among homeowners with capital gains between \$0 and \$500,000, TRA97 increased the semiannual sales rate by 0.40–0.62 percentage points on average after TRA97, or 19–24% from the pre-TRA97 baseline levels. The short-term effect is particularly large, with the sales rate jumping 70–81% in the three years immediately after TRA97. Second, among homeowners with capital gains above \$500,000, I do not find TRA97 to have a significant effect on home sales in the long-run, although the sales rate increased by 51% in the three years immediately after TRA97. Lastly, estimation results on the tax elasticity of home sales during the post-TRA97 period suggest that a \$10,000 increase in capital gains taxes lowers the semiannual home sales rate by 0.1–0.2 percentage points, or 6–13% from the average sales rate in the post-TRA97 sample. Taken together, I find empirical evidence consistent with the theoretical prediction that housing capital gains taxation has a lock-in effect on homeowners.

Even though house prices in the United States have dropped significantly since 2006, the economic impact of housing capital gains taxation is likely to remain important for three reasons. First, capital gains exclusions are defined in nominal terms and many homeowners will eventually find themselves with more than \$500,000 housing capital gains. Second, capital gains tax rates may increase after the Jobs and Growth Tax Relief Reconciliation Act of 2003 expires in 2011, which could potentially affect housing markets nationwide. Third,

tens of millions of baby-boomers are entering retirement age and are considering selling their homes to reduce housing consumption. Capital gains taxes will become relevant to many of them since they tend to have lived in their homes for decades and have accumulated sizable gains.

The rest of this paper proceeds as follows. Section 2 introduces the background on housing capital gains taxation and illustrates how TRA97 may affect home sales. In section 3, I describe the data used in this paper. I then explain my empirical strategy, discuss estimation results, and show robustness checks and extensions in section 4. The last section concludes.

2 Background

2.1 Tax Law

TRA97 greatly simplified the tax treatment of housing capital gains. Before 1997, a home seller was subject to capital gains taxation if the selling price net of selling expenses exceeded the adjusted basis of the home. The adjusted basis is defined as purchase price plus purchase costs (e.g. settlement fees and closing costs) and the cost of improvements and additions.¹ However, if the home seller bought a replacement home of equal or greater value within a four-year window, which started two years before and ended two years after the date of the sale, he would postpone the capital gains taxes until the next time he sells his home. If the replacement home value was between the purchase price and the selling price of the current home, the difference between the replacement home value and the selling price of the current home would result in immediate taxes, and the difference between the replacement home value and the purchase price of the current home would be postponed. The amount of postponed capital gains would be subtracted from the basis of the newly purchased replacement home. This tax provision, unofficially called the “roll-over rule,” had been in the Internal Revenue Code since 1951. Hoyt and Rosenthal (1990,1992) showed that the roll-over rule generated “kinks” in home sellers’ budget sets and encouraged individuals to consume more housing than they otherwise would have.

In addition to the roll-over rule, the Internal Revenue Code also featured preferential tax treatment for older home sellers before TRA97. Beginning in 1964, homeowners aged 65 and over who had lived in their homes for at least five out of the past eight years could claim a once-in-a-lifetime exclusion of up to \$20,000 against taxable capital gains.² The maximum exclusion amount was raised to \$35,000 in 1976. In 1978, the age requirement was lowered to 55, the residence requirement was changed from living in the home for at least five out of previous eight years to three out of previous five years, and the maximum exemption amount was raised to \$100,000.³ Newman and Reschovsky (1987) show that the annual mobility rate of homeowners 55 to 64 years old increased after the 1978 reform. In 1981, the maximum exclusion amount was raised to \$125,000. This “age-55 rule” remained unchanged until TRA97.⁴

TRA97 was signed into law on August 5, 1997. Effective for sales after May 6, 1997, it fundamentally altered the tax treatment of housing capital gains. First, TRA97 eliminated the roll-over rule. Second, it eliminated the age-55 rule. Third, it allowed home sellers to

¹According the IRS rules, the cost of improvements and additions can be added to the adjusted basis, whereas the cost of repairs cannot. IRS publication 523 has more details on the distinction.

²The exclusion amount equaled the total capital gain if the sales price was less or equal to \$20,000. For homes selling for more, the excludable portion was calculated by multiplying the capital gains by the ratio of \$20,000 to the sales price.

³This \$100,000 exclusion did not depend on the sales price.

⁴This one-time exclusion was \$125,000 for both single filers and married joint filers. Married separate filers, however, had a one-time exclusion of only \$62,500. In addition, the exclusion could only be used once in a lifetime and no balance could be carried forward for a future sale.

exclude housing capital gains of \$500,000 (or \$250,000 for single filers) if they have owned and lived in their homes for at least two years of the previous five years. There is no limit on how many times one can claim such exclusions during one's lifetime.⁵ Finally, TRA97 lowered the top tax rates on long-term capital gains (defined as capital gains on assets held at least 12 months) from 28% to 20%.

Capital gains tax rates have been changed many times since 1981. Before the Tax Reform Act of 1986, the top marginal tax rate was 20%. The Tax Reform Act of 1986 raised it to 28%, although effective tax rates exceeded 28% for many high-income taxpayers because of interactions with other tax provisions. TRA97 reduced capital gains tax rates and introduced a separate rate schedule for long-term gains. Beginning May 7, 1997, the top rate on long-term capital gains was 20%. The Economic Growth and Tax Relief Reconciliation Act of 2001 lowered the top rate on assets held for at least five years to 18%. The Jobs and Growth Tax Relief Reconciliation Act of 2003 lowered the top capital gains tax rate further to 15%. Figure 1 summarizes the key changes in housing capital gains taxation from 1981 to 2008.

2.2 Theoretical Predictions

To evaluate the impact of TRA97 on home sales, we need to analyze how homeowners with different levels of capital gains and with different desired replacement homes are affected differently by TRA97. Suppose a homeowner bought his house at time 0 when the per-unit house price was p_0 . Let H denote the amount of housing purchased by this homeowner. At time t , the per-unit housing price is p_t , and the homeowner considers selling his house. In the event that he sells his house at time t , he would like to purchase a replacement home of quantity H' at price p_t . If his replacement home is a rental housing unit, $H' = 0$. To simplify the analysis, I ignore the age-55 rule and only consider married couples for the moment. I will discuss the implications of the age-55 rule and single filers at the end of this section. I also assume away purchase expenses and selling expenses when imputing capital gains.⁶

Given the tax laws described above, this homeowner's tax liability under the pre-TRA97 tax law is

$$Tax_t^{pre} = \begin{cases} \tau_t^{pre}(p_t H - p_0 H) & \text{if } p_t H' \leq p_0 H \\ \tau_t^{pre}(p_t H - p_t H') & \text{if } p_0 H < p_t H' < p_t H \\ 0 & \text{if } p_t H' \geq p_t H \end{cases}$$

where τ_t^{pre} is the capital gains tax rate faced by the homeowner before 1997. Similarly, his tax liability under the post-TRA97 tax law is

$$Tax_t^{post} = \begin{cases} 0 & \text{if } p_t H - p_0 H \leq \$500K \\ \tau_t^{post}(p_t H - p_0 H - 500K) & \text{if } p_t H - p_0 H > \$500K \end{cases}$$

where τ_t^{post} is the capital gains tax rate faced by the homeowner after 1997.

⁵The required two years of ownership and use during the five-year period ending on the date of the sale do not have to be continuous. In fact, one can even claim the capital gains exclusion on a second home, as long as the ownership and use tests are met.

⁶Without imposing these assumptions, the qualitative conclusions drawn in this section remain the same, but the notation would be far more complicated.

Figure 2 illustrates the difference between Tax_t^{pre} and Tax_t^{post} graphically. The bold solid line represents Tax_t^{post} , which does not depend on replacement home values $p_t H'$. Tax_t^{post} is zero before capital gains $(p_t H - p_0 H)$ reach \$500K. As $(p_t H - p_0 H)$ continues to rise above \$500K, Tax_t^{post} begins to increase with a slope of τ_t^{post} . On the other hand, capital gains taxes before TRA97 depend on replacement home value $p_t H'$. If the home seller chooses to purchase a less expensive replacement home ($H' \leq \frac{p_0 H}{p_t}$), Tax_t^{pre} increases in $(p_t H - p_0 H)$ with a slope of τ_t^{pre} . The bold dashed line represents Tax_t^{pre} in this case. As the home seller chooses to buy more replacement housing (H'), tax liabilities (Tax_t^{pre}) shift to the right in Figure 2 and become the thin solid lines. The upward-sloping segment of Tax_t^{pre} is steeper than that of Tax_t^{post} because TRA97 reduced capital gains tax rates (i.e., $\tau_t^{post} < \tau_t^{pre}$).

All else equal, higher capital gains taxes raise moving costs and reduce the probability of home sales. To predict how TRA97 would affect home sales is equivalent to comparing Tax_t^{post} to Tax_t^{pre} . For homeowners with capital gains between \$0 and \$500K,

$$Tax_t^{post} - Tax_t^{pre} = \begin{cases} -\tau_t^{pre}(p_t H - p_0 H) & \text{if } p_t H' \leq p_0 H \\ -\tau_t^{pre}(p_t H - p_t H') & \text{if } p_0 H < p_t H' < p_t H \\ 0 & \text{if } p_t H' \geq p_t H \end{cases}$$

In this case, $Tax_t^{post} - Tax_t^{pre}$ is non-positive, suggesting that these homeowners face weakly lower taxes after TRA97, and therefore, are more likely to sell their homes after TRA97.

For married homeowners with capital gains above \$500K,

$$Tax_t^{post} - Tax_t^{pre} = \begin{cases} (\tau_t^{post} - \tau_t^{pre})(p_t H - p_0 H) - \tau_t^{post} \cdot 500K & \text{if } p_t H' \leq p_0 H \\ (\tau_t^{post} - \tau_t^{pre})p_t H + \tau_t^{pre} p_t H' - \tau_t^{post}(p_0 H + 500K) & \text{if } p_0 H < p_t H' < p_t H \\ \tau_t^{post}(p_t H - p_0 H - 500K) & \text{if } p_t H' \geq p_t H \end{cases}$$

The sign of $Tax_t^{post} - Tax_t^{pre}$ in this case is ambiguous because it depends on the replacement home value $p_t H'$. For example, suppose the homeowner bought a \$1,000,000 house and sold it for \$2,000,000. If he bought another \$2,000,000 house as a replacement home, he would pay less taxes before TRA97 because of the roll-over rule. If his replacement home is worth \$1,000,000, he would pay less taxes after TRA97 because of the exclusion.

In summary, the above analysis predicts that TRA97 would weakly increase home sales rates among homeowners with positive gains up to the maximum exclusion amount. However, the effect of TRA97 on homeowners with capital gains above the maximum exclusion level is ambiguous in theory. Note that the simple model described above ignores the age-55 rule and only considers married homeowners. To the extent that some elderly homeowners were able to take advantage of the age-55 rule before 1997, the average effect of TRA97 on home sales would be smaller for homeowners with capital gains between \$0 and \$125K. For single filers, the maximum exclusion amount available after 1997 is \$250K instead of \$500K. To the extent that some homeowners with capital gains between \$250K and \$500K are singles, the average effect of TRA97 on home sales would also be smaller in this capital gains range. Taken together, TRA97 should have the largest effect on homeowners with capital gains between \$125K and \$250K.

Also, the simple model described above takes a rather “static” view of the roll-over rule. If a home seller purchases a more expensive replacement home to take advantage of the roll-over rule in the pre-TRA97 regime, the model states that he owes zero tax liability. In reality, his capital gains are postponed instead of being eliminated. A more dynamic model should equate the homeowner’s tax liability to the expected present discounted value of future taxes on the deferred gains which depends on factors such as the homeowner’s discount rate and moving probability in subsequent years. The larger this present value of future taxes is, the less valuable the roll-over rule becomes, and the bigger impact TRA97 would generate.

3 Data

The key component of the data analyzed in this paper are the sales records provided by *The Warren Group*, a private company that collects real estate and financial information. These are transactions on single-family houses in 16 cities and towns within the Boston metropolitan area from 1982 to 2008. The 16 cities and towns are Belmont, Brookline, Cambridge, Carlisle, Cohasset, Concord, Dover, Lexington, Lincoln, Needham, Newton, Sherborn, Sudbury, Wellesley, Weston, and Winchester. These are relatively homogeneous cities and towns with high house prices and no active real estate markets for second homes. Homeowners living in these cities and towns are mostly high-income and well-educated individuals. Table 1 shows that 61–83% of individuals 25 years and older who live in these places have at least a Bachelor’s degree, compared to an average of 33% in Massachusetts. The median household income in 1999 in these places was around \$100,000.⁷

The sales data have two pieces. The first contains exhaustive records on single-family house transactions from 1987 to 2008. In other words, if there was a single-family house in the 16 cities and towns that was sold anytime between January 1, 1987 and December 31, 2008, the sales record would appear in the data. Each record has information on parcel ID, parcel address, sales date, sales price, buyer name, seller name, assessment value, and some house characteristics.⁸ The raw data have 85,797 transactions on 50,803 parcels.

Because house prices in the Boston metropolitan area did not appreciate significantly enough from 1987 to 1997 for me to observe many houses with large capital gains to be compared with the post-TRA97 period, I also obtain a data set of 1982–86 sales records. During this period of time, house prices in the studied area experienced steep increases. This second piece of sales data was compiled by a company that was later acquired by *The Warren Group*. As a result, it is very different from the 1987–2008 data set and only has information on parcel address, sales date, sales price, buyer name, and seller name. Moreover, it is unclear whether these sales records are exhaustive and whether the transaction was on a single-family house. The raw data have 36,103 transactions.

Because the 1982–86 sales records have no property type indicator differentiating single-family properties from other types of properties, combining them directly with the 1987–2008 single-family sales records would result in loss of records on single family houses that were sold between 1982–86 but not between 1987–2008. Therefore, I collect the FY2007 assessment data from each of the 16 cities and towns to help identify single-family houses from other types of properties in the 1982–86 sales data, as the assessment records contain both property addresses (used to merge the two data sets) and a property type indicator.⁹

⁷Brookline and Cambridge are different from the rest as they have a large student population with relatively low income who tend to be renters. In a robustness check shown later in the paper, I drop these two cities and the estimates are very similar.

⁸Parcel and parcel ID are terms used in the property tax assessment practice in Massachusetts. In this paper, a parcel refers to a single-family house. A parcel ID is a unique ID that is attached to each single-family house.

⁹The FY2007 assessed value reflects the market price of the house as of January 1, 2006.

The FY2007 assessment data show that the 16 cities and towns have a total of 80,987 parcels, and 14,447 single-family transactions from the 1982–86 data are identified.

After combining the 1982–86 single-family transactions with the 1987–2008 single-family transactions, I perform a series of data cleaning procedures and obtain a data set with a total of 89,731 sales records on 52,109 parcels.¹⁰ On average, 64% of all single-family houses in the 16 cities and towns appear in the final data set. In other words, 36% of the parcels were never sold between 1982 and 2008 and are omitted from my data set. To the extent that these unobserved parcels are less responsive to tax changes, I would find TRA97 has a bigger effect on home sales than if I had observed all parcels, including those that were never sold between 1982 and 2008. Among the parcels that I do observe, 54% of them were sold only once, while the remaining 46% were sold more than once during the sample period. Table 2 displays the mean and median nominal sales prices and the number of sales by year.

To study the lock-in effect of capital gains taxation, I need to know homeowners' accumulated capital gains at each point of time. However, the sales records only contain price information when there is a transaction. To be able to impute the market price of each parcel at each point of time regardless of whether there is a transaction, I use ZIP code level semiannual house price indices provided by *Fiserv Lending Solutions*. These indices, also called *Case-Shiller Home Price Indices*, are repeat sales indices on single-family houses. Figure 3 shows the semiannual house price appreciation rates for the 26 ZIP codes in the 16 cities and towns during the sample period. The vertical bars connect the maximum and minimum appreciation rates, and the circles indicate the appreciation rates averaged across the 26 ZIP codes at each point of time. From 1982 to 2008, the housing markets in the sample areas experienced significant ups and downs. Differences in house price movements over time and across ZIP codes provide useful variations to study the effect of capital gains taxation on home sales.

Using the FY2007 assessment values, which reflects the market values of properties as of January 1, 2006, and the 1982–2008 ZIP code level house price indices, I extrapolate the median single-family house prices of each ZIP code to all years between 1982 and 2008 at half-year intervals. Figure 4 shows these extrapolated median house prices. The vertical bars connect the maximum and minimum median values, and the circles indicate median prices averaged across the 26 ZIP codes. The average median house price increased from about \$100,000 in 1982 to about \$750,000 in 2008 in nominal terms.

Putting together the purchase prices from the sales data and the ZIP code level house price appreciation rates, I impute prices for each parcel at half-year intervals for all subsequent years before the next transaction. For example, if a parcel was sold in the first half of 1990 at price P_{1990}^0 and then was sold again in the second half of 2000 at price $P_{2000.5}^0$, I derive prices $\{P_{1990.5}^1, P_{1991}^1, P_{1991.5}^1, \dots, P_{2000}^1, P_{2000.5}^1\}$ by inflating the sales price P_{1990}^0 with the 1990–2000 house price appreciation rates specific to that ZIP code. Similarly, I derive prices $\{P_{2001}^1, P_{2001.5}^1, P_{2002}^1, \dots, P_{2008}^1, P_{2008.5}^1\}$ by inflating the sales price $P_{2000.5}^0$ with the 2001–08 house price appreciation rates.

In the example given above, it is possible that the price $P_{2000.5}^1$, imputed using the ZIP code level house price indices, is different from the actual sales price $P_{2000.5}^0$. Similarly, the

¹⁰Such procedures include dealing with sales between non-individual parties (e.g. financial institutions, trusts, builders, and developers), multiple sales on the same date or within a short period of time, sales with suspiciously low prices, sales between individuals of the same last name, and other unusual cases.

imputed price P_{2006}^1 may be different from the FY2007 assessment value. In fact, for residence spells that end in another transaction, the correlation between the imputed prices and actual sales prices is 0.86. The correlation between the imputed prices in the first half of 2006 and the FY2007 assessment values is 0.84. Such discrepancies may be caused by a number of factors. For example, different houses within the same ZIP code may experience different appreciation rates. Personal circumstances may induce some home sellers to accept a lower bidding price in order to close the deal sooner. I discuss the impact of these “discrepancies” on my estimation results in more detail later in the paper.

In the end, I create a panel of single-family houses in which each observation is a parcel-time combination and time is measured in the unit of half-year. I define capital gains (CG) as

$$\begin{aligned} CG_t &= \text{Current Price} - \text{Purchase Price} \\ &= p_0 H \prod_{s=1}^t (1 + h_s + \pi_s) - p_0 H \end{aligned} \quad (1)$$

where h_s is the real house price appreciation rate at time s and π_s is the inflation rate at time s .¹¹ Following Biehl and Hoyt (2007) and Cunningham and Engelhardt (2008), I also drop the 1997 observations because the law was signed in the second half of 1997 but applied retrospectively to home sales in the first half of 1997. My final analysis sample has 1.46 million observations.

Equation (1) suggests that the variations in CG_t may come from three sources. The first is the variation in purchase price of the house ($p_0 H$). Suppose two houses were bought at the same time, but one was bought for \$500,000 and the other \$1,000,000. Even if the two houses follow exactly the same house price appreciation rates subsequently, their nominal capital gains at time t are likely to be very different. For example, if house prices double between the time of purchase and time t , the house that was bought for \$500,000 would have \$500,000 capital gains and the house that was bought for \$1,000,000 would have twice as much capital gains. The second is the variation in the time of purchase. Suppose there are two identical houses next to each other with the same market price of \$1,000,000. If one was bought twenty years ago and the other two years ago, the two houses are likely to have different capital gains. The third is the variation in the house price appreciation rates between time 0 and time t (i.e., $h_1, h_2, \dots, h_t$). Suppose two houses were bought at the same time for the same price. One house is located in a ZIP code where house prices rose sharply and the other in a ZIP code where house prices rose moderately. The two houses will have different capital gains at time t . Note that although the third source of variation is at the ZIP code level, the first two contain within-ZIP but cross-parcel variation. This paper uses all three sources of variation to identify the effect of TRA97 on home sales.

Figure 5 shows histograms of capital gains before and after TRA97. Only a very small fraction of observations had capital gains above \$500K before 1997. In contrast, a large number of parcels have accumulated more than \$500K capital gains after TRA97. The housing market boom in the late 1990s to early 2000s contributed to this pattern, and so did the fact that capital gains exclusions are written in nominal terms rather than being indexed

¹¹Technically speaking, taxable capital gains exclude selling expenses such as the 6% brokerage fee usually paid by home sellers. In practice, it is unclear whether homeowners take selling expenses into account when thinking about their housing capital gains. Here, I assume that homeowners perceive their capital gains to be the simple difference between the current and purchase prices. Later in the paper, I define CG differently to allow homeowners to be more sophisticated as a robustness check.

by inflation. When comparing home sales rates of houses with capital gains above \$500K before and after TRA97, we need to be cautious because the houses that had more than \$500K capital gains before 1997 may be very different from their counterparts after 1997.

Table 3 shows summary statistics of some key variables for the pre-TRA97 and post-TRA97 samples separately. The average semiannual home sales rate was 2.2 percent before 1997 and 2.1 percent after 1997. Average nominal capital gains were about \$75,000 before 1997 and \$374,000 after 1997. The average semiannual house price appreciation rate after 1997 was 2% in real terms, which is significantly higher than the 1% before 1997. Consistent with the pattern shown in Figure 5, Table 3 indicates that only 1% of the pre-TRA97 observations have capital gains over \$500K, whereas 26% of the post-TRA97 observations are in that category. Home sales rates of houses in different capital gains categories also changed notably after 1997. In particular, the sales rates of houses with negative capital gains or capital gains above \$500K declined considerably, while the sales rates of houses with capital gains between \$0 and \$500K changed little after TRA97.

4 Empirical Strategy and Estimation Results

4.1 A Difference-in-Differences Framework

To identify the effect of capital gains taxation on home sales, I estimate the following linear probability model:¹²

$$\begin{aligned}
 Sale_{ict} = & \alpha_0 + \alpha_1 \cdot (0 < CG_{ict} \leq 125K) \\
 & + \alpha_2 \cdot (125K < CG_{ict} \leq 250K) \\
 & + \alpha_3 \cdot (250K < CG_{ict} \leq 500K) \\
 & + \alpha_4 \cdot (CG_{ict} > 500K) + \beta_1 \cdot (0 < CG_{ict} \leq 125K) \cdot TRA97 \\
 & + \beta_2 \cdot (125K < CG_{ict} \leq 250K) \cdot TRA97 \\
 & + \beta_3 \cdot (250K < CG_{ict} \leq 500K) \cdot TRA97 \\
 & + \beta_4 \cdot (CG_{ict} > 500K) \cdot TRA97 \\
 & + \gamma_1 \log(lotsize)_{ict} \\
 & + \gamma_2 h_{ict-1} + \delta_c + \theta_t
 \end{aligned} \tag{2}$$

where $Sale_{ict}$ indicates whether homeowner i in city c sells his house at time t . $TRA97$ is an indicator variable that equals one in years after 1997. $\log(lotsize)_{ict}$ measures how large the parcel is. h_{ict-1} is the real house price appreciation rate in the previous time period. δ_c and θ_t are city fixed effects and time fixed effects, respectively.¹³ Because capital gains are imputed using ZIP code level house price indices, I cluster standard errors at the ZIP code level.

The “control” group is houses with non-positive capital gains, as they are not subject to taxes either before or after TRA97. The housing downturns in the early 1990s and after 2005 ensure that many homeowners are in the control group both before and after TRA97. There are four “treatment” groups. The first two treatment groups are houses with capital gains

¹²I choose LPM because it is easy to estimate and interpret, and measurement error problems tend to be worse in non-linear models (see Bound, Brown, and Mathiowetz (2001)). Results shown later in the paper suggest that the estimates are robust to a range of non-linear models.

¹³I control for city fixed effects rather than ZIP code fixed effects in the main regression model because cities and towns are the effective jurisdictions. For example, property taxes are collected at the city/town level. In a robustness check not shown here, I control for ZIP code fixed effects instead and the results are almost identical.

between \$0 and \$125K and houses with capital gains between \$125K and \$250K, respectively. As discussed earlier, TRA97 weakly reduced capital gains taxes on these houses. Therefore, we expect β_1 and β_2 to be positive. Moreover, we expect β_1 to be smaller than β_2 because of the age-55 rule prior to TRA97. The third treatment group is houses with capital gains between \$250K and \$500K. For houses owned by married couples, TRA97 weakly reduced their capital gains taxes and we expect $\beta_3 > 0$. For houses owned by singles, however, the effect of TRA97 is ambiguous. The last treatment group is houses with capital gains over \$500K. Again, the effect of TRA97 on this group is ambiguous because of the tradeoff between the roll-over rule before 1997 and the large exclusion after 1997. Since the effect of TRA97 is ambiguous for homeowners in this group regardless of their marital status but it is only ambiguous for single homeowners with capital gains between \$250K and \$500K, we expect $\beta_3 > \beta_4$.¹⁴

The coefficients β_1 , β_2 , β_3 , and β_4 capture the average effect of TRA97 on houses with positive capital gains. Individual homeowners within the same capital gains category may be affected by TRA97 differently. For example, homeowners age 55 and over and homeowners who plan to move to more expensive homes probably experienced less of a treatment from TRA97 than other homeowners because of the age-55 rule and the roll-over rule before 1997. As a result, the average effect that I estimate in this paper may be smaller than the effect of treatment on the treated. Also, the identification assumption here is that all else equal, sales rates of houses in each treatment group would not have changed relative to the control group in the absence of TRA97. In other words, houses with different levels of capital gains may have different sales probabilities, but without the arguably exogenous changes to the exclusion levels introduced by TRA97, we would not expect home sales rates to change systematically accordingly to these breakpoints (e.g., \$125K, \$250K, and \$500K). Although the identification assumption cannot be directly tested, I perform various falsification and robustness tests later in the paper to support the its plausibility.

Table 4 displays the main estimation results. $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are positive and statistically significant. Furthermore, $\hat{\beta}_1$ appears to be smaller than $\hat{\beta}_2$, consistent with the theoretical prediction that the age-55 rule prior to 1997 should reduce the impact of TRA97 on homeowners with capital gains between \$0 and \$125K. $\hat{\beta}_3$ also appears to be smaller than $\hat{\beta}_2$, consistent with the theoretical prediction that the lower maximum exclusion amount for singles would reduce the effect of TRA97 on homeowners with capital gains between \$250K and \$500K. Overall, the semiannual home sales rate of houses with capital gains between \$0 and \$500K increased by 0.40–0.62 percentage points after TRA97, or 19–41% from the pre-TRA97 baseline levels. In contrast, I find that TRA97 had no significant effect on the sales rate of houses with capital gains over \$500K, suggesting that on average the loss of the roll-over rule balances out the gain of generous exclusions after 1997 for these homeowners.¹⁵

I also control for lot size ($\log(\text{lotsize})$) and house price appreciation rate (h_t) in equation (2). Lot size is time-invariant and captures the variation in home sales rate across different parcels. House price appreciation rates in the previous time period vary across ZIP codes and time. It captures two things. First, if homeowners form their expectations about future house price appreciation based on past house price movements, as suggested by Case and

¹⁴Although the majority of homeowners in the 26 ZIP codes are married couples, there are non-trivial numbers of homeowners who are not married couples. According to the 2000 Census, between 19 and 51% of owner-occupied units are not owned by married couples in these ZIP codes. Therefore, I use \$250K as a breakpoint in constructing treatment groups. In a robustness check not shown here, I combine houses with capital gains between \$125K and \$500K into one treatment group and the estimation results remain qualitatively the same.

¹⁵In a robustness check not shown here, I control for city-time fixed effects to allow for city-time specific shocks to home sales rates. The results are very similar to the main results.

Shiller (1988), then h_t captures the variation in house price expectation across ZIP codes. Second, to the extent that changes in local house prices reflect changes in the desirability of the neighborhood, h_t captures whether the ZIP code is becoming more desirable. Table 4 shows that the estimated coefficient on lot size ($\hat{\gamma}_1$) is negative and statistically significant, suggesting that larger houses are sold less often than smaller houses, possibly due to higher moving costs. The magnitude of $\hat{\gamma}_1$ implies that a house on a 30,000 square-foot lot (75th percentile) is 23% less likely to be sold than a house on a 9,000 square-foot lot (25th percentile). The estimated coefficient on house price appreciation rate $\hat{\gamma}_2$ is negative but statistically insignificant.

4.2 Further Discussion of the Empirical Strategy

Regression results shown in the previous section suggest that TRA97 increased home sales rates of houses with capital gains between \$0 and \$500K. This estimated effect of TRA97 combines two mechanisms. The first is the direct reversal of the lock-in effect of capital gains taxation. Put simply, when the cost of selling a house is reduced, homeowners are more likely to sell their homes. The second is the capitalization of TRA97 into house prices which in turn affect home sales. After TRA97, potential home buyers recognize that future gains are subject to lower taxes and they are willing to bid more for the same house.¹⁶ Separating out the capitalization effect precisely from the reversal of the lock-in effect is difficult and beyond the scope of this paper. However, I can conduct simple simulation exercises to bound the relative importance of the capitalization effect by recognizing that the capitalization effect should only cause a one-time change in house prices at the time of the tax law change.

The capital gains measure used in this paper is constructed using the purchase price of the house and the ZIP code level house price indices. According to these indices, nominal single-family house prices in the sample ZIP codes increased on average by about 10 percent from 1994 to 1996, 24 percent from 1996 to 1998, and 32 percent from 1998 to 2000. Presumably, an unknown fraction of the 24 percent appreciation between 1996 and 1998 was related to TRA97. In the first simulation, I assume half of the 24 percent appreciation reflects TRA97 being capitalized and the other half is unrelated to the tax law change. In the second simulation, I assume all of the appreciation between 1996 and 1998 is due to TRA97 and nominal house prices would have been flat if it were not for TRA97. In both exercises, I first simulate a new set of ZIP code house price indices by purging out the price appreciation assumed to be related to the capitalization effect. The house price movements implied by this new set of indices are presumably exogenous to TRA97. I then use the simulated indices to impute capital gains and estimate the main specification. The estimation results are shown in Table 5. Even when I attribute all house price appreciation between 1996 and 1998 to the capitalization effect of TRA97 and throw out the price changes in this period entirely, I still find estimates similar to the main results, suggesting that the capitalization mechanism plays a relatively small role in the estimated effect of TRA97. Therefore, the estimated coefficients mostly represent the reversal of the lock-in effect of capital gains taxation.

The empirical strategy used in this paper compares houses with different levels of capital gains before and after TRA97. The identification assumption is that in the absence of TRA97, the home sales rate of houses with positive capital gains would not have changed differentially from that of houses with non-positive capital gains, conditional on other control variables. Although we will never know the counterfactual sales rate of houses with different levels of capital gains were TRA97 not passed, the specific sample analyzed in this

¹⁶Poterba (1984, 1991) discusses the link between house prices and the real after-tax cost of homeownership.

paper and various falsification tests shown below appear to lend support to the identification assumption.

The data used in this paper span 1982–2008. During the pre-TRA97 period, the Boston metropolitan area experienced a remarkable boom (mid 1980s) and bust (early 1990s). During the post-TRA97 period, the area also experienced a significant boom (early 2000s) and bust (after 2005). Having a full housing cycle both before and after TRA97 makes the comparison between the pre and post periods less likely to be driven by different housing market trends in the two periods. In addition, houses examined in this paper are all single-family houses in relatively homogeneous areas without active second-home markets. Hence, confounding factors such as single-family and multi-family properties behaving differently during housing cycles and the fraction of investment properties changing with housing market movements are unlikely to drive my results.

To further investigate the plausibility of my identification assumption, I carry out a number of falsification tests using different sample periods and regime cutoff dates. The idea is that if I also find significant results using regime cutoff dates when no tax law change occurs in reality, then the effect of TRA97 that I find in the previous section is likely to be caused by some uncontrolled spurious factors. Note that because not all years have observations with negative capital gains or capital gains above \$500K in my sample, I focus on the \$0–\$500K capital gains range and use houses with capital gains between \$0 and \$25K as the “control” group in the falsification tests.¹⁷

As shown in Figure 1, tax reforms passed in 1986, 1997, 2001, and 2003 affected housing capital gains tax treatment. Hence, 1987–96 and 2004–08 are periods when capital gains tax law did not change. In column (2) of Table 6, I limit the sample to 1988–96 and use July 1, 1992 as the regime cutoff date.¹⁸ In column (3), I limit the sample to 2005–08 and use January 1, 2007 as the regime cutoff date. It is reassuring that the estimated coefficients in both columns are small and insignificant. In contrast, when I limit the sample to 1994–2000 and use the actual TRA97 cutoff date, the estimated coefficients are large, positive, and statistically significant.

In Table 7, I limit the sample to five-year windows and choose the mid-point to be the regime cutoff date. For example, the first column in the top panel uses 1991–95 observations and defines July 1, 1993 to be the cutoff date for a hypothetical regime change. Again, when the year 1997 falls outside the five-year window, the estimated coefficients are mostly small and insignificant. On the other hand, home sales rates of houses with positive capital gains increase the most when 1997 falls inside the five-year window. In summary, although we cannot be absolutely sure that the identification assumption holds, the falsification tests shown in Table 6 and Table 7 suggest that it is very unlikely that the precise patterns of my findings are driven by some coincidental changes taking place in 1997.

4.3 Other Robustness Checks

In the main specification, I use a simple LPM to estimate the effect of capital gains taxation on home sales. Alternatively, I can estimate non-linear models that may be superior in theory when the positive outcome is rare or when the data are residence spells with right censoring. In panel A of Table 8, I show the estimation results for the probit, complementary log-log, and hazard models. The estimated coefficients of these non-linear models are not directly comparable to the LPM coefficients. Therefore, I show the average marginal effects

¹⁷If one is concerned that houses with non-positive capital gains is not a good control group, using the \$0–\$25K capital gains range as the control group also serves as a robustness check.

¹⁸I exclude 1987 to avoid the 1986 tax reform having a lagged effect. Similarly, I exclude 2004 in column (3).

relative to the pre-TRA97 baseline sales rates for the probit and complementary log-log models and the hazard ratios for the hazard model. Panel A of Table 8 shows that the estimated effect of TRA97 are robust to different model specifications.

To test whether the results are driven by the anticipation effect – namely, homeowners knew that TRA97 was going to pass and decided to delay selling their homes until the law took effect, I drop the 1996 observations in panel B of Table 8. The estimated coefficients are little changed, suggesting that the main findings shown in Table 4 are unlikely to be driven by the anticipation effect. In fact, Dai, Maydew, Shackelford and Zhang (2008) show evidence suggesting that the passage of TRA97 was unexpected. To test whether the main results are driven by composition changes where houses first sold before TRA97 are systematically different from houses first sold after TRA97, I limit the sample to houses that were first sold before 1997. The estimates are similar to the main results. As shown in Table 1, Brookline and Cambridge are somewhat different from the other 14 cities and towns because of the large student population. To test whether these two cities are driving the estimates, I drop them in a robustness check and find the results remain the same.

The empirical strategy used in this paper depends on capital gains being accurately measured. Although sales prices in transaction data tend to be more accurate than self-reported house values in public survey data (see Goodman and Ittner (1992) and Kiel and Zabel (1999)), capital gains are likely to be measured with errors in this paper for a number of reasons. First, the roll-over rule prior to TRA97 allowed home sellers to postpone capital gains. Because I do not observe whether homeowners in my sample postponed capital gains before 1997, I may have underestimated their capital gains. Second, the Internal Revenue Code allows home sellers to subtract selling expenses from sales prices to calculate taxable capital gains. To the extent that homeowners are sophisticated enough to take this provision into account when thinking about housing capital gains, I may have overestimated their capital gains. Third, home improvements and additions affect capital gains because they increase house values and the costs of improvements and additions can be subtracted from capital gains for tax purposes. Since I do not observe home improvement and addition activities, I may have overestimated homeowners' capital gains.¹⁹ Lastly, houses within the same ZIP code may experience different appreciation rates. By using the ZIP code level house price indices, I may have measured capital gains with errors.

Even though small measurement errors that do not cause capital gains to be misclassified do not bias my estimates, there is no guarantee that I have not placed observations into the wrong bin because of measurement issues. Since measurement errors in binary variables are not classical, I cannot sign the bias easily and have to take them seriously. In column (2) of Table 9, I assume selling expenses to be 7% of the current house value (i.e., 6% brokerage fee and 1% other expenses) and adjust capital gains accordingly. The estimated effects of TRA97 become smaller, but remain roughly comparable to the main results. To get a direct measure of home improvement costs, I use the 1985–2007 American Housing Survey (AHS) to estimate homeowners' expenditure on home improvements and additions as a percent of house value in the Boston metropolitan area.²⁰ Because houses in my sample tend to be more expensive than the average house in the Boston area, I apply the 75th percentile of the home improvement expenditure as a percent of house value in the AHS data to adjust the capital gains measure. Column (3) shows the estimation results after adjusting for home

¹⁹The size of such measurement errors depends on the costs of home improvements, whether homeowners take the costs into account when thinking about housing capital gains, and whether the Case-Shiller indices used in this paper are truly quality-constant house price indices.

²⁰I use the AHS national sample because the Boston MSA sample was surveyed much less frequently. Before 1985, the AHS only asked about whether the homeowner had done any home improvements but not the costs of these improvements. In addition, the cost measure was not available in the 1995 survey and the questions on home improvements changed significantly in 1997.

improvement and addition costs. The estimates of the key coefficients are similar to the main results. In column (4), I control for both real house price and its linear time trend (i.e., the interaction between real house price and a continuous time measure). If more expensive homes are more likely to undergo major improvements and this pattern is trending up over time, the inclusion of real house price and its linear time trend may implicitly control for home improvement activities. Again, the estimates are similar to the main results.

Recall that because I use ZIP code level house price indices to impute house prices, the imputed prices may be very different from actual sales prices for residence spells that end in another transaction and from assessed values indicated by the FY2007 assessment records. To check how the discrepancies affect my estimates, I drop residence spells with imputed prices very different from actual sales prices or assessed values. In this sample, the correlation between imputed prices and actual sales prices is 0.95 and the correlation between imputed prices and assessed values is 0.93. Column (5) shows the estimates for this restricted sample. It is reassuring that the results are comparable to the full sample results. Because observations near the breakpoints are more likely to be misclassified, I place zero weights on observations with capital gains between $-\$5,000$ and $\$5,000$, between $\$118,750$ and $\$131,250$, between $\$237,500$ and $\$262,500$, and between $\$475,000$ and $\$525,000$. Column (6) shows the estimation results after applying these weights and they are also very close to the main results.

Besides measurement error, one may be concerned about a potential simultaneity problem in my empirical strategy. More specifically, the ZIP code level house price indices, which enter the calculation of capital gains on the right hand side of equation (2), are weighted repeat sales indices based on transaction data. Also, the outcome variable in equation (2) is based on these same transactions. Since prices and sales volume tend to be positively correlated, my estimates may suffer from simultaneity bias.²¹ Although there is no direct way of remedying this problem, I can indirectly test the potential impact of the simultaneity issue on my estimates by using a different measure of capital gains. Instead of using these ZIP code level house price indices, I use the semiannual house price indices for the Cambridge-Newton-Framingham Metropolitan Division to impute house prices and capital gains.²² There are a total of 140 ZIP codes in the Cambridge-Newton-Framingham Metropolitan Division, only 20 of which are in my sample. The potential simultaneity problem should be much less severe if I use the MSA level HPI to construct capital gains measures, although the measurement error problem may be exacerbated. Column (7) of Table 9 shows the estimation results. The estimated coefficients on the key interaction terms are qualitatively similar to the main results.

4.4 Timing of the TRA97 Effect

Because I have a relatively long panel, I am able to investigate both the short-term and the long-term effects of TRA97. If a great number of homeowners were locked-in by capital gains taxation before 1997, we should see a large release effect of TRA97 in the short run. The long-term effect of TRA97, in contrast, may be smaller but should still be positive for homeowners with capital gains below the maximum exclusion level.

I first limit the sample to the 1982–2000 observations to examine the short-term effect of TRA97. The estimation results are shown in column (2) of Table 10. The estimated coefficients on the interaction terms of all capital gains categories are large, positive, and

²¹I thank the anonymous referee for bringing up this point.

²²The HPIs for the Cambridge-Newton-Framingham Metropolitan Division are also provided by *Fiserv Lending Solutions*. I tried to obtain the HPIs for the Boston-Quincy Metropolitan Division as well. However, the data are not available before 1985 for the Boston-Quincy Metropolitan Division.

statistically significant. In the three years immediately after TRA97, the semiannual home sales rates of houses with positive gains up to \$500K increased by 1.66–1.82 percentage points, or 70–81% from the pre-TRA97 baseline levels. The semiannual home sales rate of houses with capital gains above \$500K increased by 1.25 percent points, or 51% from the pre-TRA97 baseline level. These results suggest that there were many homeowners who were locked-in before 1997. As soon as TRA97 was signed into law, these homeowners took advantage of the newly provided capital gains exclusions and sold their homes.

Next, I drop the 1998–2000 observations to investigate the long-term effect of TRA97. Column (3) of Table 10 shows the estimation results. The long-term effects of TRA97 are very similar to the main results shown before. The semiannual home sales rate of houses with positive gains up to \$500K increased by 0.37–0.66 percentage points, or 17–25% from the pre-TRA97 baseline levels. On the other hand, the long-run effect of TRA97 on the sales probability of houses with capital gains above \$500K is negative but statistically insignificant.

In column (4), I use the full sample but interact the four capital gains categories with a 1998–2000 dummy and a 2001–08 dummy separately. The results echo the findings shown in column (2) and column (3). In the short run, TRA97 released the lock-in effect of capital gains taxation for houses with all levels of capital gains. The effect is economically and statistically significant. In the long run, the effect became smaller on houses with low or moderate capital gains, and the effect disappeared on houses with very large capital gains.

The striking difference between the short-term and long-term effect of TRA97 on houses with capital gains above \$500K warrant more discussion. The simple model presented earlier in the paper shows that the effect of TRA97 on houses with capital gains above \$500K depends on the tradeoff between the loss of the roll-over rule and the gain of generous exclusions. For homeowners with capital gains not much above \$500K at the time of TRA97, the gain of the exclusions is likely dominate the loss of the roll-over rule, and therefore, their moves immediately after TRA97 may have contributed to the large short-term effect. In the long-run, the remaining homeowners in this capital gains range are more likely to have gains substantially higher than \$500K and the elimination of the roll-over rule locks them in. Therefore, the empirical pattern observed in the data is consistent with the theoretical prediction of heterogeneous responses among homeowners with over \$500K capital gains.

4.5 Estimating the Tax Elasticity of Home Sales

As mentioned earlier, TRA97 lowered the top tax rate on long-term capital gains from 28% to 20%. In 2001, the top rate was further reduced to 18% for capital gains on assets held for five years or longer. Since 2003, long-term capital gains have been taxed with a maximum rate of 15%. In this section, I exploit the legislative changes in the top capital gains tax rate during the post-TRA97 period to estimate the tax elasticity of home sales.²³

To impute the tax liability, Tax_{ict} , for a homeowner with capital gains above \$500K, I have to make an assumption about the marginal tax rates faced by homeowners and their marital status since I do not have homeowner characteristics and income data. Table 1 shows that homeowners in the 16 cities and towns studied in this paper are mostly high-income individuals. Therefore, I assume that they all face the top capital gains tax rate. I also assume these homeowners are married couples as the 2000 Census data suggest that the majority of

²³Estimating the tax elasticity of home sales for the pre-TRA97 period is very difficult because capital gains taxes depended on age of the seller and value of the replacement home before 1997, neither of which is observed in my data.

them are. These assumptions allow me to calculate the amount of taxes that each homeowner would owe if he were to sell his house at time t :

$$Tax_{ict} = Real(\tau_t \cdot \max(0, CG_{ict} - 500,000)), \begin{cases} \tau_t = 0.20 & \text{for 1998 – 2000} \\ \tau_t = 0.18 & \text{for 2001 – 2002} \\ \tau_t = 0.15 & \text{for 2003 – 2008} \end{cases}$$

where function $Real(\cdot)$ converts nominal dollars into real 2000 dollars. Panel A of Table 11 shows the summary statistics of Tax_{ict} . About 26% of the post-TRA97 sample have nominal capital gains above \$500K and are therefore subject to capital gains taxation if they decide to sell their house. The median homeowner faces a tax liability of approximately \$30,000. To compare observations in my sample with the general population, I do the same calculation for homeowners in the 2007 Survey of Consumer Finances (SCF). About 5% of homeowners in the 2007 SCF have more than \$500K housing capital gains. Among them, the median homeowner also faces a tax liability of around \$30,000.

To obtain the tax elasticity of home sales, I estimate the following linear probability model:

$$Sale_{ict} = \alpha_0 + \lambda Tax_{ict} + \rho RCG_{ict} + \gamma_1 \log(lotsize) + \gamma_2 h_{ict-1} + \delta_c + \theta_t \quad (3)$$

where RCG_{ict} is the real capital gains and the key parameter of interest is λ . The tax elasticity is identified by comparing houses with similar real capital gains in different years, as arguably exogenous legislative changes cause tax rates to differ across years. Panel B of Table 11 shows that the estimated tax elasticity of home sales, $\hat{\lambda}$, is negative and statistically significant. Its magnitude implies that a \$10,000 increase in capital gains taxes reduces the semiannual home sales rate by 0.1 percentage points, or 6% from the average sales rate in the post-TRA97 sample.

I also estimate the tax elasticity of home sales using an alternative measure of capital gains where 7% of current house value is subtracted from the nominal capital gains to obtain the tax measure Tax_{ict} . This tax measure allows homeowners to be more sophisticated and to take selling expenses into account when thinking about taxable capital gains. The estimate $\hat{\lambda}$ is still negative and statistically significant, but the size of the estimate doubles. The size of the estimate suggests that a \$10,000 increase in taxes reduces the semiannual home sales rate by 0.2 percentage point, or 13% from the average sales rate in the post-TRA97 sample.

The reduced-form estimate of the tax elasticity allows us to do policy simulations and to infer the impact of hypothetical changes in housing capital gains taxation on home sales. For example, the estimates shown in Table 11 suggest that if we eliminate capital gains taxes on housing altogether, the semiannual home sales rate would increase by 17–40% for homeowners with capital gains above \$500K. Another interesting scenario is when the Jobs and Growth Tax Relief Reconciliation Act of 2003 expires in 2011 and the top capital gains tax rate increases from the current 15% to 20%. The estimates shown in Table 11 suggest that the semiannual home sales rate would drop by 6–13% for homeowners with capital gains above \$500K.

A few caveats are worth mentioning. First, measurement errors introduced by the assumptions about homeowners' marginal tax rates and marital status may bias my estimates toward zero. Second, if homeowners perceive the 2001 and 2003 tax cuts as transitory rather than permanent, my estimates may overstate the magnitude of the true tax elasticity. In summary, although caution is warranted when making out-of-sample predictions using the estimates shown in this section, the tightly estimated elasticity of home sales with respect to

housing capital gains taxes is useful for many informative back-of-the-envelope calculations.

5 Conclusion

TRA97 introduced the largest change to the tax treatment of housing capital gains in decades. In this paper, I use price and transaction data to study the effect of TRA97 on home sales. I find evidence suggesting that TRA97 reversed the lock-in effect of capital gains taxes for houses with capital gains between \$0 and \$500,000. After 1997, the semi-annual sales rate of these houses increased by 0.40–0.62 percentage points, or 19–24% from the pre-TRA97 baseline levels. I do not find TRA97 to have a significant effect on the sales rate of houses with capital gains above \$500,000. In addition, I show that the short-term effect of TRA97 is very different from the long-term effect. From 1998 to 2000, home sales rates rose sharply across all capital gains categories as previously locked-in homeowners took advantage of the newly available exclusions. From 2001 to 2008, the effect of TRA97 became relatively moderate. I also estimate the tax elasticity of home sales during the post-TRA97 period, using legislative changes in capital gains tax rates. The estimation results suggest that a \$10,000 increase in tax liability reduces the semiannual sales rate by 0.1–0.2 percentage points, or 6–13% from the average sales rate in the post-TRA97 sample.

This paper complements the existing literature on TRA97 and brings new evidence on the lock-in effect of capital gains taxation in housing markets. However, it is worth emphasizing that we need to be cautious in generalizing the findings of this paper. The homeowners analyzed in this paper are not representative of the U.S. population. Therefore, they may have responded to TRA97 differently from homeowners in other places and with different characteristics.

More research is needed to fully understand the welfare impact of TRA97. First, this paper does not consider any general equilibrium effects potentially generated by TRA97. By reducing taxes on housing capital gains, TRA97 reduced the user cost in the housing market, which could have increased housing investment at the expense of non-housing investment. Second, the data set analyzed in this paper is a panel of houses instead of a panel of households. Thus, I do not observe where people moved to once they sold their houses. We need high quality longitudinal data on households to quantify how capital gains taxation before 1997 distorted homeowners' mobility and housing consumption decisions and to understand the extent to which TRA97 corrected these distortions.

Acknowledgments

I am very grateful to Jim Poterba, Bill Wheaton and Jerry Hausman for advice and support. I thank Tonja Bowen Bishop, Amy Finkelstein, Amanda Kowalski, Byron Lutz, Therese McGuire, Raven Molloy, and David Powell for helpful discussions. Valuable comments from anonymous referees significantly improved the paper. I thank Chip Case, Joe Nugent, Karen MacTavish, Knorr Maryanne, Tim Warren Jr, and especially Alan Pasnik, Jim Shaughnessy, and David Stiff for generously providing me with data and patiently answering my many questions. This research was supported by the National Institute on Aging, Grant Number P01-AG05842. The findings and conclusions expressed are solely those of the author and do not represent views of the Board of Governors, the staff of the Federal Reserve System, or the National Institute on Aging.

References

- Biehl, A.; Hoyt, W. Unpublished Manuscript. University of Kentucky; 2007. The Taxpayer Relief Act of 1997 and homeownership: is smaller now better?.
- Bound, J.; Brown, C.; Mathiowetz, N. Measurement error in survey data. In: Heckman, JJ.; Leamer, E., editors. *Handbook of Econometrics*. Vol. 5. Elsevier; 2001. p. 3705-3843.

- Burman, L.; Wallace, S.; Weiner, D. How capital gains taxes distort homeowners' decisions?. Proceedings of the National Tax Association 89th Annual Conference; 1996.
- Case, K.; Shiller, R. New England Economic Review. 1988 November/December. The behavior of home buyers in boom and post boom markets; p. 29-46.
- Cunningham C, Engelhardt G. Housing capital-gains taxation and homeowner mobility: evidence from the Taxpayer Relief Act of 1997. *Journal of Urban Economics* 2008;63(3):803–815.
- Dai Z, Maydew E, Shackelford D, Zhang H. Capital gains taxes and asset prices: capitalization or lock-in? *Journal of Finance* 2008;63(2):709–742.
- Farnham, M. Unpublished Manuscript. University of Michigan; 2006. Residential mobility, the capital gains tax, and the Taxpayer Relief Act of 1997.
- Goodman J, Ittner J. The accuracy of home owners' estimates of house value. *Journal of Housing Economics* 1992;2(4):339–357.
- Hoyt W, Rosenthal S. Capital gains taxation and the demand for owner-occupied housing. *The Review of Economics and Statistics* 1990;72(1):45–54.
- Hoyt W, Rosenthal S. Owner-occupied housing, capital gains, and the Tax Reform Act of 1986. *Journal of Urban Economics* 1992;32(2):119–139.
- Kiel K, Zabel J. The accuracy of owner-provided house values: the 1978–1991 American Housing Survey. *Real Estate Economics* 1999;27(2):263–298.
- Newman S, Reschevsky J. Federal policy and the mobility of older homeowners. *Journal of Policy Analysis and Management* 1987;6(3):402–416.
- Poterba J. Tax subsidies to owner-occupied housing: an asset-market approach. *Quarterly Journal of Economics* 1984;99(4):729–52.
- Poterba J. House price dynamics: the role of tax policy and demography. *Brookings Papers on Economic Activity* 1991;1991(2):143–203.

Figure 1.
Key Changes in Housing Capital Gains Tax Treatment Since 1981

Figure 2.

Tax Liability as a Function of Capital Gains before and after TRA97

Note: For simplicity, I ignore the age-55 rule in the pre-TRA97 tax regime in this graph.

Figure 3.

Semiannual House Price Appreciation Rates in the 26 ZIP Codes, 1982–2008

Note: Each vertical bar connects the maximum and minimum semiannual house price appreciation rates among the 26 ZIP codes at a given time. The circle on each vertical bar indicates the mean of the 26 house price appreciation rates. House price appreciation rates refer to single-family houses and are in nominal terms.

Figure 4.

Median Values of Single-Family Houses in the 26 ZIP Codes, 1982–2008

Note: Each vertical bar connects the maximum and minimum median house values among the 26 ZIP codes at a given time. The circle on each vertical bar indicates the mean of the 26 median house values. House values refer to single-family houses and are in nominal terms.

Figure 5.
Histograms of Nominal Capital Gains before and after TRA97

Table 1

Characteristics of the 16 Cities/Towns

	Bachelor's Degree	Graduate or Professional	Median HH Income 1999	Percent Owners
MA	19.5%	13.7%	50,502	62
Belmont	26.4%	36.7%	80,295	61
Brookline	31.7%	45.3%	66,711	45
Cambridge	26.7%	38.5%	47,979	32
Carlisle	44.3%	39.1%	129,811	94
Cohasset	40.0%	20.7%	84,156	85
Concord	31.4%	34.7%	95,897	81
Dover	43.5%	34.3%	141,818	95
Lexington	26.8%	42.2%	96,825	83
Lincoln	28.5%	40.7%	79,003	61
Needham	31.3%	33.5%	88,079	81
Newton	29.1%	38.9%	86,052	70
Sherborn	39.0%	36.7%	121,693	93
Sudbury	34.4%	37.5%	118,579	92
Wellesley	34.7%	41.2%	113,686	83
Weston	30.1%	45.0%	153,918	86
Winchester	32.2%	32.7%	94,049	81

Note: Data are from MA State Data Center. "Bachelor's Degree" and "Graduate or Professional" refer to educational attainment for the population 25 years and older in 2000. "Percent Owners" refers to owner-occupied housing units as a percentage of all occupied housing units in 2000.

Table 2

Sales Prices and Number of Sales by Year

Year	Mean	Median	N
1982	133,563	120,000	1,782
1983	149,126	132,000	3,098
1984	182,233	160,000	3,031
1985	234,514	210,000	3,048
1986	307,488	268,500	2,688
1987	337,536	288,125	3,349
1988	367,017	309,000	2,982
1989	357,790	303,975	2,751
1990	345,899	292,500	2,409
1991	316,865	270,000	3,378
1992	324,116	279,000	3,706
1993	338,285	292,500	3,672
1994	362,934	315,000	3,726
1995	384,245	330,625	3,292
1996	407,630	350,000	3,614
1997	441,455	375,000	3,796
1998	484,066	410,000	4,019
1999	534,243	450,000	4,026
2000	652,489	535,000	3,582
2001	724,393	589,000	2,941
2002	749,607	625,000	3,405
2003	802,447	668,000	3,241
2004	886,244	725,000	3,701
2005	957,707	783,000	3,302
2006	939,009	766,000	3,036
2007	935,880	750,000	3,283
2008	942,208	759,500	2,622
Total			87,480

Note: Sales prices are in nominal terms.

Table 3

Summary Statistics

	Pre-TRA97 (N=499,271)		Post-TRA97 (N=960,492)		Difference
	Mean	SD	Mean	SD	p-Value
Semiannual Sales Probability	0.022	0.148	0.021	0.142	0.00
Nominal Capital Gains	75,104	96,327	374,352	354,797	0.00
Log(lot size)	9.68	0.95	9.71	0.97	0.00
Real House Price Appreciation	0.010	0.047	0.020	0.040	0.00
Fraction of CG≤0	0.14	0.34	0.03	0.16	0.00
Fraction of 0<CG≤125K	0.64	0.48	0.21	0.41	0.00
Fraction of 125K<CG≤250K	0.17	0.38	0.20	0.40	0.00
Fraction of 250K<CG≤500K	0.05	0.22	0.31	0.46	0.00
Fraction of CG>500K	0.01	0.07	0.26	0.44	0.00
Sales Rate CG≤0	0.022	0.148	0.013	0.113	0.00
Sales Rate 0<CG≤125K	0.022	0.145	0.021	0.145	0.00
Sales Rate 125K<CG≤250K	0.026	0.159	0.027	0.161	0.00
Sales Rate 250K<CG≤500K	0.021	0.145	0.020	0.139	0.00
Sales Rate CG>500K	0.025	0.155	0.017	0.129	0.00

Note: Capital gains (CG) are measured in nominal terms. Lot size is measured in square footage. Real house price appreciation refers to the semiannual appreciation rate. Real dollars are in 2000 dollars.

Table 4

Effect of TRA97 on Home Sales - Main Results

β_1 : (0<CG≤125K) * TRA97	0.45 ** (0.15) [21%]
β_2 : (125K<CG≤250K) * TRA97	0.62 ** (0.18) [24%]
β_3 : (250K<CG≤500K) * TRA97	0.40 * (0.19) [19%]
β_4 : (CG>500K) * TRA97	-0.29 (0.42) [-12%]
Log(Lot Size)	-0.42 ** (0.04)
Real House Price Appreciation Rate	-0.01 (0.01)
Time Fixed Effects	Y
City Fixed Effects	Y
N	1.46M

Note: The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before 1997.

* and ** denote significance at the 5% and 1% levels respectively.

Table 5

Simulations of the Capitalization Effect of TRA97

	Main Results	Half HPA 1996–98	No HPA 1996–98
	(1)	(2)	(3)
(0<CG≤125K) * TRA97	0.45** (0.15) [21%]	0.50** (0.15) [23%]	0.82** (0.16) [38%]
(125K<CG≤250K) * TRA97	0.62** (0.18) [24%]	0.47* (0.17) [18%]	0.60** (0.17) [23%]
(250K<CG≤500K) * TRA97	0.40* (0.19) [19%]	0.34 (0.18) [16%]	0.54** (0.17) [25%]
(CG>500K) * TRA97	−0.29 (0.42) [−12%]	−0.26 (0.42) [−11%]	0.02 (0.42) [1%]
N	1.46M	1.46M	1.46M

Note: The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before 1997.

* and ** denote significance at the 5% and 1% levels respectively.

Table 6

Falsification Tests

	1994–2000	1988–96	2005–08
	1997	1992	2007
	(1)	(2)	(3)
(25K<CG≤125K) * After	1.11 ** (0.21) [43%]	−0.12 (0.14) [−4%]	0.14 (0.25) [6%]
(125K<CG≤250K) * After	1.58 ** (0.26) [67%]	−0.21 (0.11) [−8%]	−0.38 (0.22) [−15%]
(250K<CG≤500K) * After	1.28 ** (0.29) [61%]	−0.25 (0.21) [−12%]	−0.33 (0.22) [−17%]
N	262,647	351,806	227,582

Note: The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before the regime cutoff date.

* and ** denote significance at the 5% and 1% levels respectively.

Table 7

Falsification Tests

	1991–95	1992–96	1993–97
	1993	1994	1995
(25K<CG≤125K) * After	−0.30 (0.15)	0.33 (0.17)	1.01 ** (0.15)
(125K<CG≤250K) * After	−0.09 (0.20)	0.11 (0.20)	0.64 ** (0.21)
(250K<CG≤500K) * After	−0.52 * (0.20)	0.44 (0.23)	0.95 ** (0.25)
N	173,907	214,425	250,128
	1994–98	1995–99	1996–2000
	1996	1997	1998
(25K<CG≤125K) * After	1.10 ** (0.19)	0.93 ** (0.20)	0.58 * (0.21)
(125K<CG≤250K) * After	1.09 ** (0.18)	1.18 ** (0.26)	1.34 ** (0.22)
(250K<CG≤500K) * After	1.11 ** (0.32)	1.07 ** (0.24)	0.96 ** (0.21)
N	275,153	293,215	305,571
	1997–2001	1998–2002	1999–2003
	1999	2000	2001
(25K<CG≤125K) * After	0.14 (0.20)	−0.62 * (0.25)	−0.19 (0.22)
(125K<CG≤250K) * After	0.96 ** (0.20)	−0.25 (0.25)	−0.20 (0.20)
(250K<CG≤500K) * After	0.95 ** (0.21)	−0.42 (0.26)	−0.56 ** (0.20)
N	310,916	309,517	302,097

Note: The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level.

* and ** denote significance at the 5% and 1% levels respectively.

Table 8**Robustness Checks with Alternative Specifications and Samples**

	A. Alternative Specifications			
	LPM	Probit	CLogLog	Hazard
(0<CG≤125K) * TRA97	0.45** (0.15) [21%]	0.12** (0.04) [31%]	0.32** (0.09) [35%]	0.32** (0.09) [1.37]
(125K<CG≤250K) * TRA97	0.62** (0.18) [24%]	0.15** (0.04) [33%]	0.39** (0.10) [36%]	0.38** (0.10) [1.46]
(250K<CG≤500K) * TRA97	0.40* (0.19) [19%]	0.11* (0.04) [28%]	0.28* (0.11) [30%]	0.28* (0.11) [1.32]
(CG>500K) * TRA97	-0.29 (0.42) [-12%]	-0.03 (0.08) [-6%]	-0.05 (0.19) [-4%]	-0.05 (0.19) [0.95]
N	1.46M	1.46M	1.46M	1.46M
	B. Alternative Samples			
	Original Sample	Drop 1996	House Sold before 1997	Drop 2 Cities
(0<CG≤125K) * TRA97	0.45** (0.15) [21%]	0.48** (0.15) [22%]	0.58** (0.19) [27%]	0.50** (0.16) [23%]
(125K<CG≤250K) * TRA97	0.62** (0.18) [24%]	0.58** (0.18) [22%]	0.60** (0.20) [23%]	0.70** (0.18) [27%]
(250K<CG≤500K) * TRA97	0.40* (0.19) [19%]	0.41* (0.20) [19%]	0.35 (0.18) [16%]	0.37 (0.21) [17%]
(CG>500K) * TRA97	-0.29 (0.42) [-12%]	-0.24 (0.39) [-10%]	-0.24 (0.43) [-10%]	-0.23 (0.36) [-9%]
N	1.46M	1.40M	1.24M	1.31M

Note: Other control variables include log(lotsize), real house price appreciation rate, city fixed effects and time fixed effects. The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. In the hazard model column, numbers in brackets show hazard ratios. In all other columns, numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before 1997.

* and ** denote significance at the 5% and 1% levels respectively.

Table 9
Robustness Checks to Address Measurement Error and Potential Simultaneity Problems

	Original Sample (1)	Adjust for Sales Costs (2)	Adjust for Imp/Add (3)	Add Price and Trend (4)	Restricted Sample (5)	Apply Weights (6)	Use MSA Level HPI (7)
(0<CG≤125K) * TRA97	0.45 ^{***} (0.15) [21%]	0.33 ^{***} (0.11) [15%]	0.69 ^{***} (0.13) [32%]	0.51 ^{***} (0.15) [24%]	0.36 [*] (0.14) [17%]	0.45 ^{***} (0.17) [21%]	0.75 ^{***} (0.10) [35%]
(125K<CG≤250K) * TRA97	0.62 ^{***} (0.18) [24%]	0.39 ^{***} (0.10) [15%]	0.59 ^{***} (0.14) [23%]	0.63 ^{***} (0.18) [24%]	0.49 ^{***} (0.18) [19%]	0.61 ^{***} (0.18) [23%]	0.56 ^{***} (0.13) [22%]
(250K<CG≤500K) * TRA97	0.40 [*] (0.19) [19%]	0.00 (0.13) [0%]	0.35 [*] (0.14) [16%]	0.32 (0.19) [15%]	0.02 (0.19) [1%]	0.43 (0.22) [20%]	0.35 [*] (0.15) [16%]
(CG>500K) * TRA97	-0.29 (0.42) [-12%]	-0.74 (0.51) [-30%]	-0.34 (0.47) [-14%]	-0.65 (0.42) [-27%]	-0.62 (0.62) [-25%]	-0.17 (0.42) [-7%]	-0.38 (0.36) [-15%]
N	1.46M	1.46M	1.46M	1.46M	1.31M	1.29M	1.46M

Note: Other control variables include log(lotsize), real house price appreciation rate, city fixed effects and time fixed effects. The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before 1997.

* and ** denote significance at the 5% and 1% levels respectively.

Table 10

Timing of the TRA97 Effect on Home Sales

	Original Sample	After = 1998–2000	After = 2001–08	1998–2000 vs. 2001–08
	(1)	(2)	(3)	(4)
(0<CG≤125K) * TRA97	0.45 ** (0.15) [21%]	1.66 ** (0.21) [77%]	0.48 * (0.15) [22%]	
(125K<CG≤250K) * TRA97	0.62 ** (0.18) [24%]	1.82 ** (0.24) [70%]	0.66 ** (0.17) [25%]	
(250K<CG≤500K) * TRA97	0.40 * (0.19) [19%]	1.74 ** (0.23) [81%]	0.37 (0.19) [17%]	
(CG>500K) * TRA97	−0.29 (0.42) [−12%]	1.25 ** (0.24) [51%]	−0.31 (0.42) [−13%]	
(0<CG≤125K) * (1998–2000)				1.83 ** (0.19)
(0<CG≤125K) * (2001–08)				0.48 ** (0.15)
(125<CG≤250K) * (1998–2000)				1.99 ** (0.22)
(125<CG≤250K) * (2001–08)				0.65 ** (0.17)
(250<CG≤500K) * (1998–2000)				1.92 ** (0.22)
(250<CG≤500K) * (2001–08)				0.37 (0.19)
(CG>500K) * (1998–2000)				1.45 ** (0.41)
(CG>500K) * (2001–08)				−0.31 (0.42)
N	1.46M	0.73M	1.23M	1.46M

Note: Other control variables include log(lotsize), real house price appreciation rate, city fixed effects and time fixed effects. The coefficients are expressed in percentage for ease of exposition. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the baseline sales rate before 1997.

* and ** denote significance at the 5% and 1% levels respectively.

Table 11**The Tax Elasticity of Home Sales using Post-TRA97 Data**

A. Summary Statistics				
	CG Definition 1		CG Definition 2	
	Fraction>500K	Median Tax	Fraction>500K	Median Tax
Analysis Sample	0.26	28,267	0.20	25,871
SCF 2007	0.05	29,975	0.04	28,138
B. Estimation Results				
	CG Definition 1		CG Definition 2	
Tax Liability (in 10,000s)	-0.094 [*] (0.040)		-0.228 ^{**} (0.043)	
Implied Effect of a \$10,000 Tax Increase on Home Sales Probability	[-6%]		[-13%]	
N	245,450		193,470	

Note: Other control variables include real capital gains, log(lotsize), real house price appreciation rate, city fixed effects and time fixed effects. The coefficients are expressed in percentage for ease of exposition. "CG Definition 1" defines capital gains as the difference between the nominal current price and the nominal purchase price. "CG Definition 2" defines capital gains as "CG Definition 1" minus 7% of the current home price. Median tax is in 2000 dollars. Standard errors shown in parentheses are clustered at the ZIP code level. Numbers in brackets show the magnitude of the marginal effects relative to the mean of the dependent variable.

* and ** denote significance at the 5% and 1% levels respectively.