Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

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

Hui Shan

2008-53

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

Hui Shan *

September 28, 2008

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 $500,000 of capital gains when they sell their houses. Such drastic changes provide a good opportunity to study the lock-in effect of capital gains taxation on home sales. Using ZIP-code level housing price indexes and sales on single-family houses data from 1982 to 2006 in 16 affluent towns within the Boston metropolitan area, this paper finds that TRA97 reversed the lock-in effect of capital gains taxes on houses with low and moderate capital gains. However, TRA97 may have generated an unintended lock-in effect on houses with capital gains over the maximum exclusion amount. In addition, this paper exploits legislative changes in capital gains tax rate to estimate the tax elasticity of home sales during the post-TRA97 period.

Keywords: Housing, Taxation, Capital Gains
JEL classification: H24, H31, R21

*Division of Research and Statistics, Board of Governors of the Federal Reserve System, Washington, D.C. 20551. hui.shan@frb.gov. 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, and David Powell for helpful discussions. I thank Chip Case, Joe Nugent, Karen MacTavish, Knorr Maryanne, David Stiff, Tim Warren Jr, and especially Alan Pasnik and Jim Shaughnessy for generously providing me with data and patiently answering my questions. This research was supported by the National Institute on Aging, Grant Number P01-AG05842, and the Lincoln Institute of Land Policy. 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, the National Institute of Aging, or the Lincoln Institute.

1 Introduction

In the United States, capital gains are generally taxed upon realization and appreciated assets enjoy basis step-up when transferred by bequest. Economists have long recognized the 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 most drastic changes in the tax treatment of housing capital gains since the late 1970s, and therefore, serves as a natural experiment for researchers to study the impact of capital gains taxation on housing markets. For example, Bier, Maric and Weizer (2000), Farnham (2006), Cunningham and Engelhardt (2008), and Biehl and Hoyt (2007) use TRA97 as a policy instrument to examine the effect of housing capital gains taxation on homeowners' moving decisions.

Prior to TRA97, homeowners had to pay capital gains taxes when they sold their houses 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 taxes 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. TRA97 abolished both the roll-over rule and the age-55 rule. Instead, homeowners can exclude $500,000 (or $250,000 for single filers) capital gains when they sell their houses after TRA97, and they can potentially claim such an exclusion as often as every two years.

Existing studies on TRA97, including Farnham (2006), Cunningham and Engelhardt (2008), and Biehl and Hoyt (2007), have found that capital gains taxes during the pre-TRA97 period locked in many homeowners and that TRA97 released such lock-in effects. These studies, however, have two major limitations. First, because survey datasets that are publicly available often do not have sufficient information on house values for researchers

to infer accumulated capital gains, most existing studies on TRA97 have relied on the age-55 rule for identification. Second, even when it is possible to impute accumulated capital gains, nationally representative surveys usually do not capture enough high-end houses for researchers to study homeowners with capital gains above $500,000. For example, Farnham (2006) uses 1989-2003 American Housing Survey (AHS) data where the median house value is only $101,257 and the median capital gains are only $34,856 in 2000 dollars.

Due to these data limitations, several important aspects of TRA97 remain unaddressed or understudied by the existing literature, including whether TRA97 differentially affected homeowners with different levels of accumulated capital gains, how the repeal of the roll-over rule affected homeowners with capital gains over $500,000, and what the tax elasticity of home sales has been after 1997. These unanswered questions have important economic and policy implications for a number of reasons. First, capital gains exclusions are defined in nominal terms and a growing number of homeowners start to find themselves with more than $500,000 housing capital gains, especially during the 2000-2005 housing market boom. Second, capital gains tax rates may increase after the Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA) expires in 2011, which can potentially affect housing markets nationwide. Third, tens of millions of baby-boomers are entering retirement age and considering selling their big houses to reduce housing consumptions. 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 capital gains.

In this paper, I construct a panel of single-family houses using ZIP-code level semi-annual housing price indexes and 1982-2006 sales records in 16 affluent towns within the Boston metropolitan area. Because the sales data originally come from local registries of deeds, they are not subject to top-coding, and they are more accurate than self-reported housing values found in most survey datasets. 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 is the first study to look at how TRA97 affects houses with capital gains exceeding $500,000. It is also the first study to estimate the tax elasticity of home sales using post-TRA97 data.

A number of interesting findings emerge from this paper. First, the semiannual home sale rate increased after TRA97 among homeowners with capital gains between $0 and $500,000, suggesting that these homeowners were locked in by housing capital gains taxes before 1997 and TRA97 reversed such a lock-in effect. Second, for houses with capital gains over $500,000, the semiannual home sale rate declined after TRA97. This finding suggests that TRA97 may have unintentionally locked in homeowners with capital gains exceeding the maximum exclusion level. Furthermore, the releasing effect of TRA97 on homeowners with capital gains between $0 and $500,000 appears to be short-lived, whereas the locking-in effect of TRA97 on homeowners with capital gains above $500,000 appears to be long-lasting. 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 semiannual home sale rate by 0.16-0.25 percentage points.

The rest of this paper proceeds as follows. Section 2 introduces the background on housing capital gains taxation and TRA97, illustrates how TRA97 may affect home sales, and gives an overview of the existing literature. In section 3, I describe the data used in this paper. I then explain my empirical strategies, discuss estimation results, and show robustness checks and extensions to the main model 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 sale, he would postpone the capital gains taxes.[2] 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.

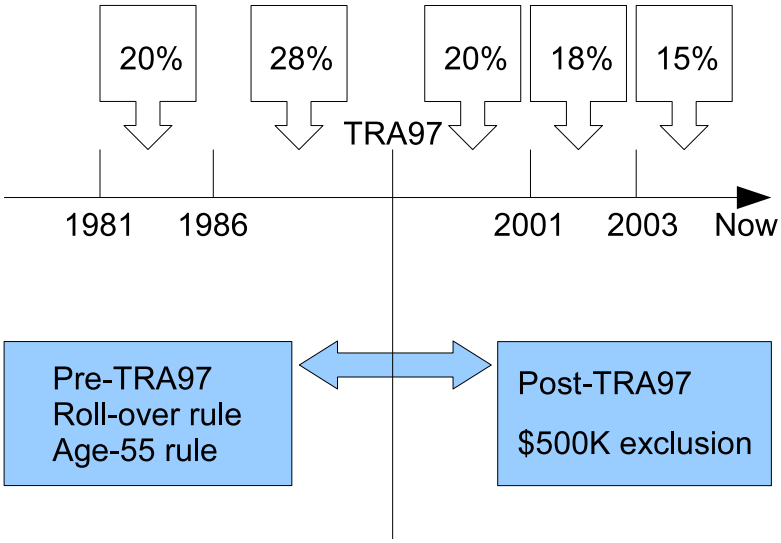
In addition to the roll-over rule, which provided preferential tax treatment for home sellers who bought more expensive replacement homes, the Internal Revenue Code also featured preferential tax treatment for older home sellers before TRA97. Beginning in 1964, homeowners aged 65 and over who had lived in their homes for at least five years during the past eight years could claim a once-in-a-lifetime exclusion of up to $20,000 against taxable capital gains.[3] The maximum exclusion amount was raised to $35,000 in 1976. In 1978, the

[1] According 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.

[2] In fact, IRS Publication 523 explicitly says that "Generally, you *must* postpone tax on the gain on the sale of your main home if you buy and live in a new main home within the replacement period and it costs at least as much as the adjusted sales price of the old home."

[3] The exclusion amount equaled the total capital gain if the sale price was less or equal to $20,000. For

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



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.[4] Finally, in 1981, the maximum exclusion amount was raised to $125,000. This "age-55 rule" remained unchanged until TRA97.[5]

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. After 1997, the tax treatment of housing capital gains no longer depended on whether a home seller bought a replacement home or the value of the replacement home. Second, it eliminated the age-55 rule. Older home sellers now face the same tax treatment as their younger counterparts. Third, it allowed home sellers to exclude $500,000 (or $250,000

homes selling for more, the excludable portion was calculated by multiplying the capital gain by the ratio of $20,000 to sale price.

[4]This $100,000 exclusion did not depend on the sale price.

[5]This 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.

for single filers) housing capital gains if they have owned and lived in their homes for at least two years during the past five. There is no limit on how many times one can claim such exclusions during one's lifetime, and the 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 gain exclusion on a second home, as long as the ownership and use tests are met. Finally, TRA97 lowered the long-term capital gains tax rates from 15% and 28% to 10% and 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%. Beginning in 2001, the top rate on assets held for at least five years was 18%. The Jobs and Growth Tax Relief Reconciliation Act of 2003 lowered the top capital gains tax rate to 15%. Figure 1 summarizes the key changes in housing capital gains taxation between 1981 and now.

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 by TRA97 differently. Suppose a homeowner bought his house at time 0 when the per-unit housing 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 actually a rental housing unit, $H' = 0$. For ease of exposition, I make two simplification assumptions. First, I assume this homeowner

is younger than 55 or he has used the one-time capital gains exclusion under the pre-TRA97 tax regime if he is 55 or older. Under this assumption, we can ignore the age-55 rule for the moment. Second, I assume away purchase expenses and selling expenses when imputing capital gains. Without imposing these assumptions, the qualitative conclusions drawn in this section remain the same, but the notation would have been far more complicated.

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

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

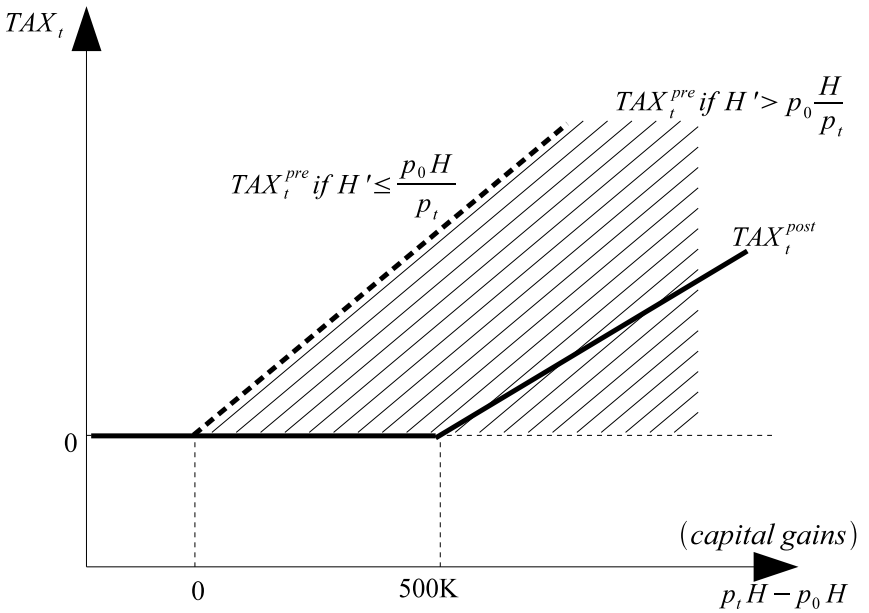
where τ_t^{pre} is the capital gains tax rate faced by the homeowner under the pre-TRA97 tax law. Similarly, his tax liability under the post-TRA97 tax regime is

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

where τ_t^{post} is the capital gains tax rate faced by the homeowner under the post-TRA97 tax law.

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 \$500,000. As $(p_t H - p_0 H)$ continues to rise above \$500,000, Tax_t^{post} increases linearly in $(p_t H - p_0 H - 500K)$ with a slope of τ_t^{post}. On the other hand, capital gains taxes depend on replacement home value $p_t H'$ before TRA97. When the homeowner with positive capital gains chooses not to purchase a sufficiently expensive replacement home, $H' \leq \frac{p_0 H}{p_t}$ and Tax_t^{pre} is a linear function of

Figure 2: Tax Liability as a Function of Capital Gains Before and After TRA97



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

$(p_t H - p_0 H)$ with a slope of τ_t^{pre}. The dotted line represents Tax_t^{pre} in this case. As H' increases, the value of his replacement home, $p_t H'$, increases accordingly, which leads to a parallel shift to the right in Tax_t^{pre} in Figure 2. The upward-sloping segment of Tax_t^{pre} is steeper than that of Tax_t^{post} because TRA97 reduced capital gains tax rates and hence, $\tau_t^{post} < \tau_t^{pre}$.

All else equal, higher capital gains tax liabilities 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} with Tax_t^{pre} for homeowners with different capital gains and different desired replacement homes. If $Tax_t^{post} - Tax_t^{pre}$ is positive, it means that TRA97 raised tax burdens and home sale rates should decline after 1997. On the other hand, if $Tax_t^{post} - Tax_t^{pre}$ is negative, it suggests that TRA97 reduces tax burdens and home sale rates should increase after 1997.

For homeowners with capital gains between \$0 and \$500,000,

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

In this case, $Tax_t^{post} - Tax_t^{pre}$ is non-positive, suggesting that homeowners with capital gains in this range have weakly lower tax liabilities after TRA97, and therefore, are more likely to sell their homes after TRA97. For homeowners with capital gains above \$500,000,

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

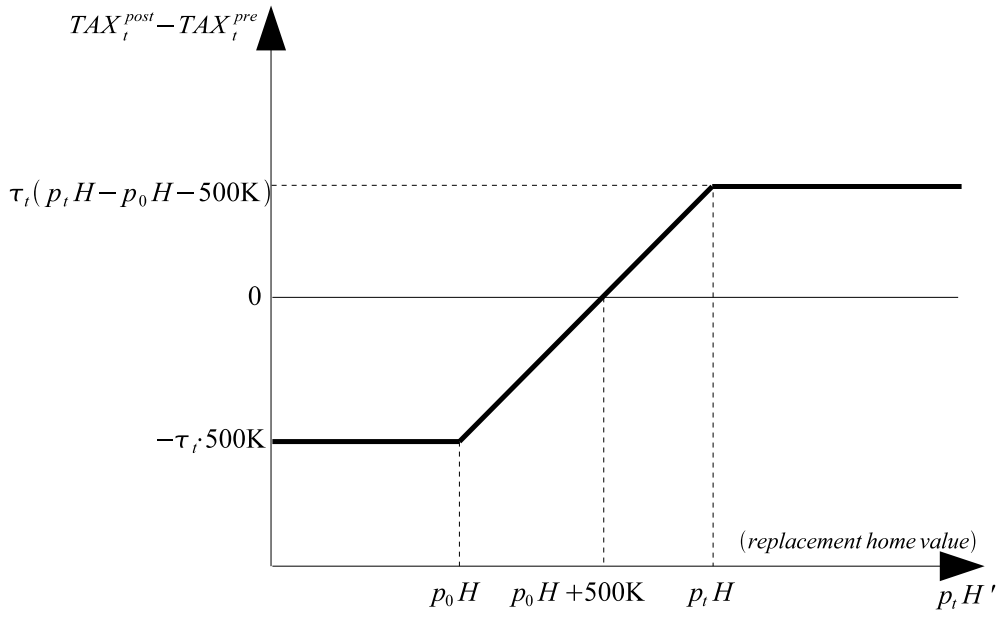
To simplify the discussion, I assume $\tau_t^{post} = \tau_t^{pre} = \tau_t$ and rewrite the above equation:

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

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, if the replacement home value is sufficiently low, then $Tax_t^{post} - Tax_t^{pre}$ is negative and home sale rates would be higher after TRA97 among these homeowners. However, if the replacement home value is sufficiently high, $Tax_t^{post} - Tax_t^{pre}$ becomes positive and home sale rates would decline after TRA97.

Figure 3 depicts the relationship between $Tax_t^{post} - Tax_t^{pre}$ and the replacement home value $p_t H'$ graphically. It shows that, *a priori*, we cannot predict how TRA97 would

Figure 3: Relationship between Tax Differences and Replacement Home Value for Homeowners with Capital Gains over $500K



Note: For simplicity, I ignore the age-55 rule of the pre-TRA97 tax regime and the difference between the pre- and post-TRA97 marginal tax rates in this graph.

affect home sale rate for homeowners with capital gains above $500,000. If most of these homeowners move to significantly less expensive replacement homes after selling their houses, then their tax burdens are lower under the post-TRA97 regime and home sale rates would increase after TRA97. In contrast, if most homeowners with large capital gains prefer living in relatively expensive replacement homes, then their tax burdens are actually higher under the new tax regime, and TRA97 would cause home sale rates to decline. The intuition behind this observation comes from the tradeoff between the elimination of the roll-over rule and the enactment of a more generous exclusion provision. On one hand, the $500,000 exclusion would reduce the amount of capital gains taxes owed, and hence, would reverse the lock-in effect of capital gains taxes. On the other hand, the elimination of the roll-over rule implies that a home seller could always avoid paying any tax by purchasing an equally or

more expensive house before 1997, but after 1997 he must pay tax on the portion of capital gains exceeding $500,000 no matter what. If the latter effect overcomes the former effect, TRA97 might unintentionally lock in homeowners with capital gains over $500,000, and we might observe a drop in home sale rates after 1997 among these people.

In summary, the above analysis predicts that TRA97 would increase home sale rates among homeowners with capital gains between $0 and $500,000. The effect of TRA97 on home sales among homeowners with capital gains above $500,000 is ambiguous in theory, and thus, only empirical investigation can determine the sign and magnitude of the effect of TRA97 on these homeowners. Note that the above analysis refers to married homeowners. For single homeowner, the exclusion level is $250,000, and theoretical predictions differ for single homeowners with capital gains between $0 and $250,000 from those with capital gains over $250,000.

2.3 Previous Studies

The pre-TRA97 capital gains taxation had been criticized for its complexity and potentially large distortions of homeowners' mobility and housing consumption decisions. Using 1970-1981 Panel Study of Income Dynamics (PSID) data, Newman and Reschovsky (1987) showed that that the annual mobility rate of homeowners 55 to 64 years old increased after the 1978 reform, which raised the exclusion level from $35,000 to $100,000 and lowered the age requirement from 65 to 55. Hoyt and Rosenthal (1990) first recognized that the roll-over rule generated "kinks" in home sellers' budget sets and encouraged them to consume more housing than they otherwise would have. Such kinks were ignored by previous studies on housing demand such as King (1980) and Rosen (1979). Using 1981 AHS data, Hoyt and Rosenthal (1990) estimated the price elasticity of housing demand with non-linear budget sets. Hoyt and Rosenthal (1992) performed policy simulations using the estimation results

of their previous paper. Their simulations suggest that by increasing capital gains tax rates, the Tax Reform Act of 1986 enhanced the importance of the capital gains kinks in home sellers' budget sets and therefore produced a larger efficiency loss. Using 1993 IRS Statistics of Income tax return data, Burman, Wallace and Weiner (1996) showed that the tax raised little revenue: a total of $50.5 billion housing capital gains were reported on Form 2119 in 1993, but $18 billion was not taxable because of the age-55 rule and $30 billion was not taxable because of the roll-over rule. The authors estimated a sequential-choice model and found that capital gains taxes had a large and significant impact on the decision to buy or to rent. They also found that capital gains taxes were a large but only marginally significant deterrent to moving down (i.e. moving to less expensive houses), but they did not find any effect of taxes on the amount of housing demanded by home sellers who chose to move down. Furthermore, the authors argued that the housing capital gains taxation at the time was regressive because it treated more favorably home sellers who could afford expensive replacement houses, and the compliance cost was high because of record-keeping and complex rules. Sinai (1998) applied a competing-risk duration model on 1970-1992 PSID data to estimate the effect of capital gains taxes on homeowners' mobility decisions. Unlike Burman, Wallace and Weiner (1996), he found that capital gains taxes had a statistically significant but small impact on the likelihood of moving, the choice of owning versus renting, and the choice of moving up versus down.

To my knowledge, only four papers have studied the impact of TRA97 on residential mobility. Using deed transfer data from nine years before to 17 months after TRA97 in four Ohio metropolitan areas, Bier, Maric and Weizer (2000) found no evidence that the probability of moving down increased after 1997. Instead, they found that moving up dominated all four areas. Farnham (2006) used 1989-2003 AHS data to study how the elimination of the age-55 rule affected residential mobility. He found evidence suggesting that homeowners under age 55 were locked in before TRA97 and the passage of TRA97 boosted residential

mobility among the previously locked-in households. Cunningham and Engelhardt (2008) used 1996 and 1998 Current Population Survey (CPS) data to compare the mobility rates of homeowners just above age 55 (i.e. 56-58) and homeowners just below (i.e. 52-54). They found that the repeal of the age-55 rule raised residential mobility of the 52-54 year olds by 22-31 percent from the mean annual mobility rate of 4 percentage points. Biehl and Hoyt (2007) used 1995-1996, 1998-1999, and 2002-2004 AHS data to conduct a similar analysis. They found that TRA97 increased mobility rates of homeowners aged 50-54 relative to homeowners aged 55-65, but such a release effect dissipated a few years after the passage of TRA97.

Overall, most existing studies on TRA97 have focused exclusively on the elimination of the age-55 rule. It is understandable that they have done so because of data limitations. For example, the CPS does not have information on house values. The AHS top codes house values at $300,000, which makes it impossible for researchers to investigate the impact of TRA97 on houses with capital gains over $300,000, let alone $500,000. Moreover, housing values in survey datasets are self-reported. Goodman and Ittner (1992) and Kiel and Zabel (1999) show that individual homeowners do not report their housing values accurately in survey datasets. Even if such measurement errors in self-reported values are random and do not systematically correlate with mobility outcomes, they may still contribute to attenuation bias in empirical analysis. As a result, there has been no research examining the mobility response to TRA97 of homeowners with capital gains over $500,000. Nor have there been any studies that estimate the tax elasticity of home sales using post-TRA97 data.

This paper fills these gaps in the literature. Instead of using public surveys where the unit of observation is typically a homeowner, this paper uses a unique panel of housing units constructed from ZIP-code level semiannual housing price indexes and 1982-2006 sales records on single-family houses in 16 affluent towns within the Boston metropolitan area.

These sales records are originally taken from local registries of deeds, and therefore, are not subject to top-coding and are presumably more accurate than self-reported values. This paper exploits the variation in exclusion levels before and after TRA97 - $125,000, $250,000, and $500,000 - to identify the effect of capital gains taxation on home sale rates. Both nonparametric and regression analysis are used to study such an effect. It also exploits legislative changes in capital gains tax rates introduced in 2001 and 2003 to estimate the tax elasticity of home sales in the post-TRA97 period. Because of the innovations in both data and identification strategies, this paper complements the existing studies on TRA97 and sheds new light on the lock-in effect of capital gains taxation in housing markets.

3 Data

The backbone of the data analyzed in this paper is the sales records provided by *The Warren Group*. These are single-family house sales in 16 cities and towns within the Boston metropolitan area from 1982 to 2006. The 16 cities and towns are Belmont, Brookline, Cambridge, Carlisle, Cohasset, Concord, Dover, Lexington, Lincoln, Needham, Newton, Sherborn, Sudbury, Wellesley, Weston, and Winchester. I selected these places because their 2006 median single-family house sale prices were over $625,000, they do not have active real estate markets for second homes, and they are all located in the Middlesex and Norfolk counties. 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 with the average of 33% in Massachusetts. The 1999 median household income of homeowners in these places was around $100,000.[6] Another important feature of these 16 cities and towns is that the number of single-family

[6]Brookline and Cambridge had much lower median household income because of the large renting population in these two cities.

houses has been roughly constant in each jurisdiction. Data from the Massachusetts Department of Revenue show no significant changes in the stock of single-family houses during the past 15 years in these 16 cities and towns.

The sales data have two components. The first contains exhaustive records on single-family house sales between 1987 and 2006.[7] In other words, if there was a single-family house in any of the 16 cities and towns that was sold anytime between 1987 and 2006, the sale record would appear in the data. The raw data have a total of 78,599 sales of 48,240 single family houses. Each record has information on parcel ID, parcel location, sale date, sale price, buyer name, seller name, current assessment value, house characteristics such as lot size, living area, year built, total number of rooms, number of bedrooms, number of bathrooms, and building style.[8] The second component contains sales records from 1982 to 1986. This dataset was compiled initially by a company that was later acquired by *The Warren Group*. As a result, it only has information on parcel location, sale date, sale price, buyer name, and seller name. Moreover, it is unclear whether the sales records are exhaustive and whether a sale was on a single-family house. The raw data have a total of 36,103 sales recorded.

To separate sales of single-family houses from sales of other properties in the 1982-1986 dataset, I requested local assessment data from all 16 cities and towns. Based on the assessment data provided by local assessors' offices, I constructed a "universe" of single-family houses for the 16 cities and towns. This constructed universe has a total of 80,978 parcels, and it contains information on parcel ID, parcel location, parcel ZIP code, current assessment value, and lot size.[9] Merging this universe dataset with the 1982-1986 sales

[7]Because the Newton data were extracted in 2006 and the rest were extracted in 2007, I only have sales between January 1, 1987 and June 23, 2006 for the city of Newton. For the other 15 cities and towns, I have all sales between January 1, 1987 and December 31, 2006.

[8]Parcel and parcel ID are terms used in assessment practice. In this paper, a parcel means a single-family house. A parcel ID is a unique ID that is attached to a single-family house. The current assessment value refers to the FY2006 assessment value for the city of Newton and FY2007 for the other 15 cities and towns.

[9]I tried to obtain additional housing characteristics such as living area, year built, total number of rooms, number of bedrooms, number of bathrooms, and building style. But because each of these variables was

records by parcel location, I identified 11,458 sales of single-family houses. Then I combined the 1982-1986 data with the 1987-2006 data. After a series of data cleaning procedures, I obtained a sales dataset with a total of 82,884 sales records on 50,369 parcels.[10] Around 57% of these parcels were sold only once between 1982 and 2006, while the remaining 43% of them were sold more than once during the sample period. On average, 62% of all single-family houses in the 16 cities and towns are in the sales data. Table 2 displays the mean and median sale prices and the number of sales by year.[11] The relatively high prices in the post-1997 period will allow me to examine the effect of TRA97 on homeowners with capital gains above $500,000.

Once I had sales data with precise purchase prices, I used 1982-2006 ZIP-code level semiannual housing price indexes, which were provided by *Fiserv Lending Solutions*, to impute nominal capital gains for these houses in subsequent years at half-year intervals. Such indexes, also called *Case-Shiller Home Price Indexes*, are constructed using repeated sales data on single-family houses.[12] Figure 4 shows the semiannual housing 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 appreciation rates averaged across the 26 ZIP codes at any given time. From 1982 to 2006, the cities and towns under study experienced significant ups and downs in the single-family housing market. Moreover, there appears to be substantial heterogeneity in housing appreciation rates across

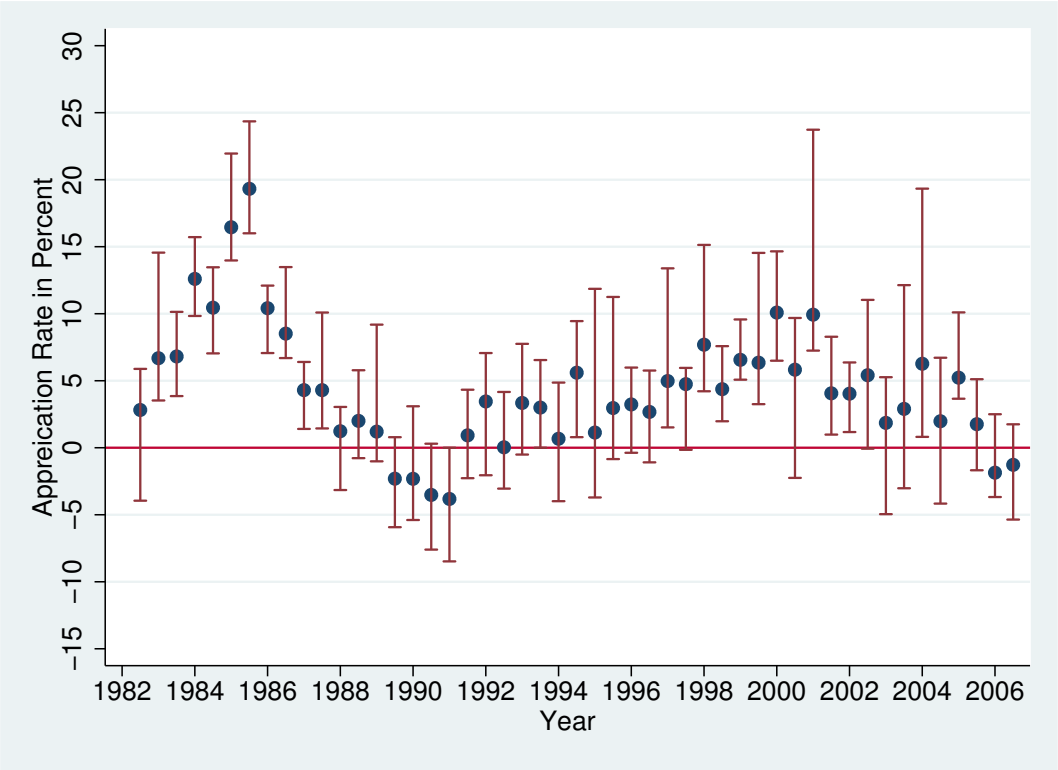
missing in at least one town's assessment data, I had to exclude them from the single-family house universe dataset.

[10]Such procedures include dealing with sales between non-individual parties (e.g. financial institutions, trusts, builders, and developers), multiple sales on the same date, sales with suspiciously low prices, and other unusual cases.

[11]I do not convert the prices into real dollars because capital gains tax exclusions are in nominal terms.

[12]Case-Shiller HPI and the OFHEO (Office of Federal Housing Enterprise Oversight) HPI are the two major sets of housing price indexes in the United States. The methodology behind these two indexes are very similar, but they rely on different underlying data: Case-Shiller uses purchase prices from county records and OFHEO uses conforming mortgage data from Fannie Mae and Freddie Mac. Because houses in my dataset have relatively high values and OFHEO does not include sales in which jumbo mortgages are used, Case-Shiller HPI is more appropriate in this application.

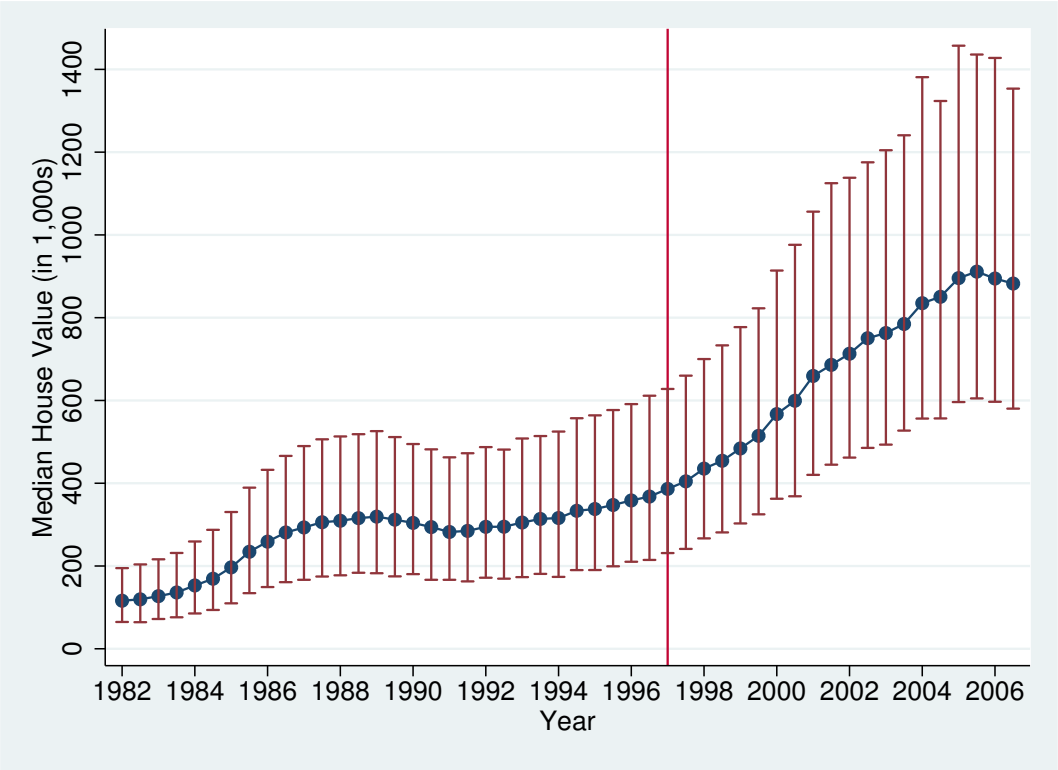
Figure 4: Semiannual ZIP-Code Level Price Appreciation Rates of Single-Family Houses in the 16 Cities between 1982 and 2006



Note: Each vertical bar connects the maximum and minimum semiannual single-family nominal housing appreciation rates of the 26 ZIP codes at a certain time. The circle on each vertical bar indicates the mean of the 26 housing appreciation rates.

ZIP codes at any given time. Combined with the policy shocks brought forward by TRA97, such housing market movements provide useful variations for me to study the effect of capital gains taxation on home sales. Using median assessment values of FY2007 and 1982-2006 ZIP-code level housing value appreciation rates, I extrapolated median assessment values for all years between 1982 and 2006 by ZIP code at half-year intervals. Figure 5 displays these extrapolated median house values. The vertical bars connect the maximum and minimum median values, and the circles indicate median values averaged across the 26 ZIP codes at any given time. The average median house value increased from $116,000 in 1982 to $883,000 in 2006 in nominal terms. There also appears to be substantial heterogeneity in median house values across ZIP codes at any given time.

Figure 5: Semiannual ZIP-Code Level Median Values of Single-Family Houses in the 16 Cities between 1982 and 2006.



Note: Each vertical bar connects the maximum and minimum median single-family nominal house values of the 26 ZIP codes at a certain time. The circle on each vertical bar indicates the mean of the 26 median house values.

Putting together purchase prices from the sales data and semiannual appreciation rates of the corresponding ZIP code, I imputed current prices for each parcel at half-year intervals for all subsequent years before the next sale. For example, if a parcel was sold in the first half-year of 1990 at price P^0_{1990} and then was sold again in the second half-year of 2000 at price $P^0_{2000.5}$, I would derive current prices $\{P^1_{1990.5}, P^1_{1991}, P^1_{1991.5}, ..., P^1_{2000}, P^1_{2000.5}\}$ by applying 1990-2000 semiannual appreciation rates to P^0_{1990}. Similarly, I would derive current prices $\{P^1_{2001}, P^1_{2001.5}, P^1_{2002}, ..., P^1_{2006}, P^1_{2006.5}\}$ by applying 2000-2006 semiannual appreciation rates to P^0_{2000}. For the second half-year of 2000, I have both the actual sale price, $P^0_{2000.5}$, and the predicted current price, $P^1_{2000.5}$. At the end of the sample period, I have both the FY2007 assessment value, P^{av}_{FY2007}, and the predicted current price, $P^1_{2006.5}$. The correlation

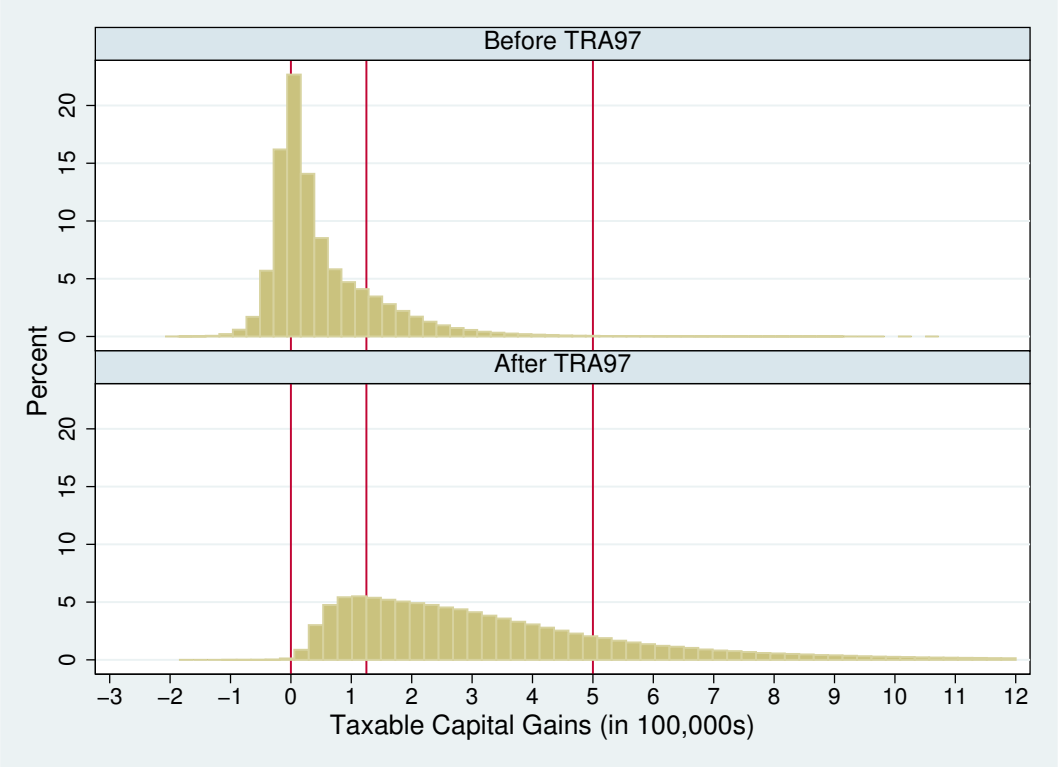
between the actual values (i.e. the actual sale prices and FY2007 assessment values) and the predicted values is 0.80. In the end, I created a panel of single-family houses where each observation is a parcel-time combination, each observation has information on purchase price and current price, and time is measured in the unit of half-year. By law, selling expenses can be subtracted from selling prices when calculating taxable capital gains. Because home sellers usually pay 6% of selling prices to realtors as commission fees, I define taxable capital gains (TCG) as

$$TCG = CurrentPrice - PurchasePrice - 0.06 \times (CurrentPrice).$$

Note that TCG is defined in nominal terms because tax rules are written in nominal terms. In addition, because TRA97 required homeowners to have owned and lived in the house for two years out of the previous five to qualify for capital gains exclusions, I dropped observations that are within two years of the most recent sale for the post-TRA97 records. To prevent extreme cases from driving the estimation results, I also dropped observations where house prices are over $5 million or the difference between the current price and the purchase price is below -$100,000 or above $3 million. The final analysis sample has 1.16 million observations on 46,403 unique parcels.

Figure 6 shows histograms of TCG before and after TRA97. There are two key differences between these two histograms. First, due to the housing market boom between the late 1990s and mid 2000s, only a tiny fraction of parcels had negative capital gains after 1997. Second, only a small fraction of parcels had capital gains over $125,000 before 1997. Even fewer parcels had capital gains above $500,000 before TRA97. In contrast, a large number of parcels have accumulated more than $500,000 capital gains after TRA97. Both the housing market boom and the fact that the $500,000 capital gains exclusion level is written in nominal terms rather than being indexed by inflation contribute to this pattern.

Figure 6: Histograms of Nominal Taxable Capital Gains Before and After TRA97



When comparing home sale rates of houses with capital gains above $500,000 before and after TRA97, we need to be cautious because the houses that had more than $500,000 capital gains before 1997 may be very different from their counterparts during the post-TRA97 period. In the regression analysis shown later in this paper, I control for as many factors as the dataset allows and control for them as flexibly as possible. But the lack of observations with huge capital gains before 1997 poses a serious challenge to any empirical strategy estimating the impact of TRA97 on houses with capital gains above $500,000.

Table 3 shows summary statistics of some key variables for the pre-TRA97 and post-TRA97 periods separately. The average semiannual home sale rate in the sample is 2.4 percentage points during the pre-TRA97 period, and it is the same as the sale rate during the post-TRA97 period. The number of observations is approximately evenly split between the pre-TRA97 period and the post-TRA97 period. Consistent with the pattern described in

Figure 6, Table 3 indicates that only 0.2% of the pre-TRA97 observations are in the category with capital gains over $500,000, whereas 20.5% of the post-TRA97 observations are in that same category. A mere 0.1% of the post-TRA97 observations have negative capital gains. There is no significant change in the average lot size before and after TRA97. The average semiannual housing appreciation rate after 1997 is 3.1% in real terms, which is significantly higher than the average semiannual housing appreciation rate of 1.0% before 1997.

4 Empirical Strategy and Estimation Results

In this section, I first employ non-parametric approaches to illustrate the impacts of TRA97 on home sale rates for homeowners with different levels of accumulated capital gains. Then I use a simple difference-in-differences regression framework to estimate the magnitude of such impacts. I also perform robustness checks and study the dynamic aspect of the TRA97 effect. Lastly, I exploit legislative changes in the top capital gains tax rate to estimate the tax elasticity of home sales in the post-TRA97 period.

4.1 Non-Parametric Approaches

As discussed earlier in this paper, TRA97 eliminated capital gains taxation on home sales for homeowners with capital gains below $500,000. Thus, we expect home sale rates of these homes to increase after 1997. The effect of TRA97 on homes with capital gains above $500,000 is ambiguous. On one hand, TRA97 legislated a very generous exclusion of $500,000. On the other hand, it took away the roll-over rule that enables home sellers to avoid paying any tax at the time of a sale. Before formulating a rigorous regression model, it is instructive to use non-parametric approaches to compare home sale rates for houses with different levels of capital gains before and after TRA97. The idea is to let the data speak

for themselves without imposing any functional form assumptions.

As shown in Figure 2, tax liabilities are continuous functions of housing capital gains in both the pre-TRA97 and the post-TRA97 periods, even though the first derivative of tax liabilities with respect to capital gains is discrete. Because home selling is a binary decision, non-parametric smoothing techniques such as a local polynomial regression are useful to illustrate the relationship between home sale probabilities and taxable capital gains. A local polynomial regression is similar to a kernel regression. While a kernel regression estimates the weighted mean locally, a local polynomial regression estimates a weighted polynomial function locally. In a kernel regression, we minimize

$$\sum_i K\left(\frac{x_i - x_0}{h}\right)(y_i - m_0)^2$$

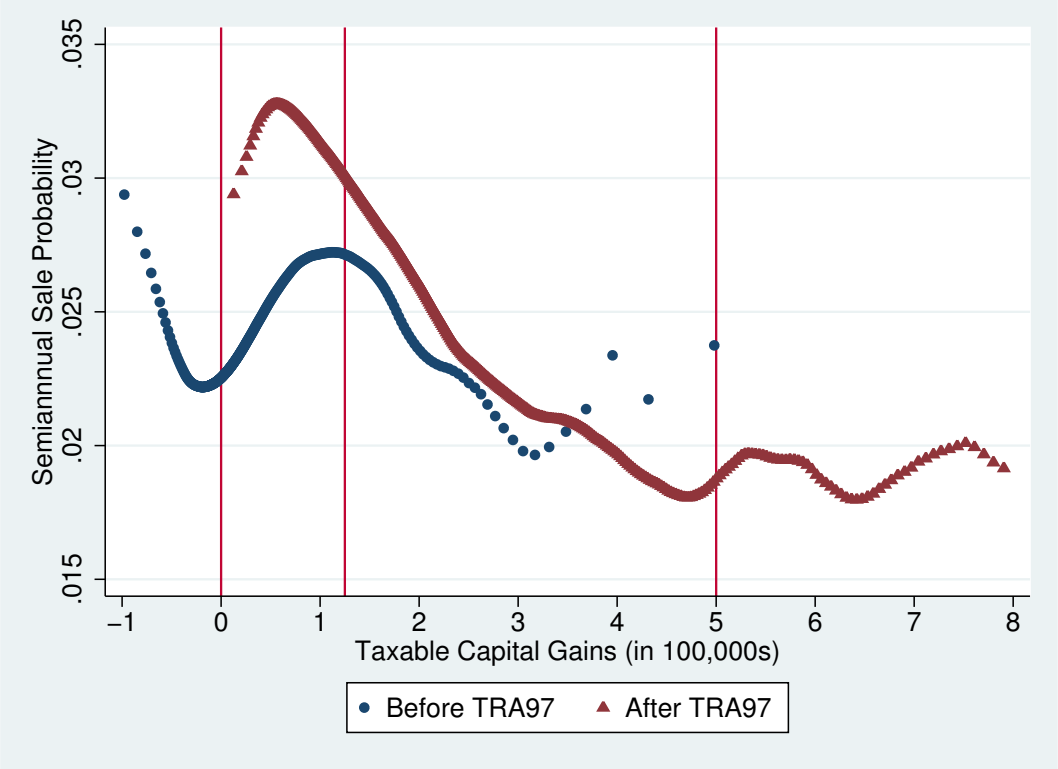
with respect to m_0, where $K(\cdot)$ is a kernel weighting function and h is the bandwidth. In a local polynomial regression, we minimize

$$\sum_i K\left(\frac{x_i - x_0}{h}\right)\left(y_i - a_0 - a_1(x_i - x_0) - ... - a_p\frac{(x_i - x_0)^p}{p!}\right)^2$$

with respect to $(a_0, a_1, ..., a_p)$, where p is preferably an odd number. Fan and Gijbels (1996) list many attractions of local polynomial regressions, including that they have better bias properties than kernel regressions.

I fit a local cubic polynomial model on the pre-TRA97 data and the post-TRA97 data separately, using the Epanechnikov kernel function and a bandwidth of $50,000. Figure 7 shows the estimation results. The scattered circles represent the smoothed home sale probabilities at various capital gains levels before TRA97, and the scattered triangles represent those after TRA97. Because only a tiny fraction of parcels had capital gains over $500,000 before TRA97 or had negative capital gains after TRA97, I can compare only parcels with

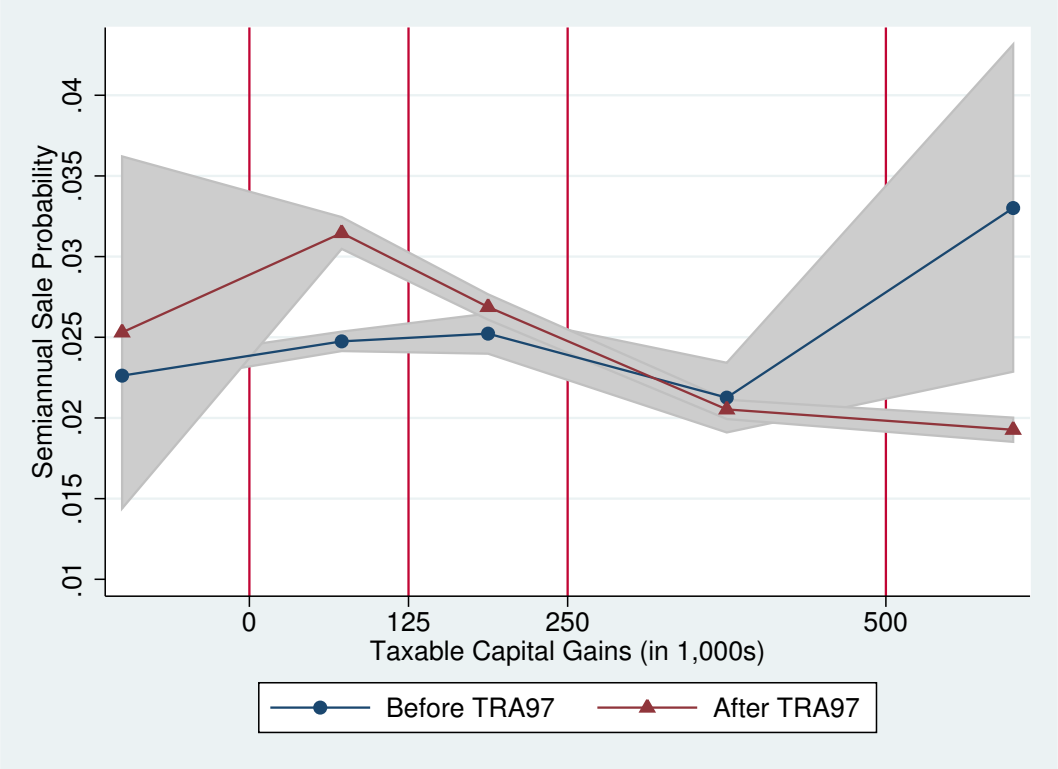
Figure 7: Local Polynomial Regression Results



Note: The bandwidth is 0.5 (or $50,000), and the Epanechnikov kernel function is used.

capital gains between $0 and $500,000 in the local polynomial framework. Figure 7 presents several interesting patterns. First, home sale rates and taxable capital gains have an inverse U-shape relationship within the $0 to $500,000 capital-gain range. When a homeowner has lived in the house only for a short period of time, his accumulated capital gains tend to be small and he is also unlikely to move since he has adjusted his housing consumption recently. When the homeowner has stayed in the house for a long period of time, his accumulated capital gains are usually large. But to the extent that there is heterogeneity among homeowners in their moving propensities, living in the same house for an extended period of time may indicate that this homeowner has a distaste for moving. Thus, the home sale rate of this homeowner is low, as suggested by the "mover-stayer" model in the literature. Second, the local polynomial regression result suggests that home sale rates are higher after

Figure 8: Sale Probability by Capital Gains Categories Before and After TRA97



TRA97 for homes with relatively low capital gains, but the order reverses right before capital gains reach $400,000. This pattern is roughly consistent with the theoretical prediction that TRA97 increases the home sale rate of houses with capital gains between $0 and $500,000. Third, it shows that home sale rates increase slightly near the $500,000 exclusion level during the post-TRA97 period. Such a pattern may imply the presence of the "churning" behavior - namely, homeowners sell their homes to reset tax basis when they accumulate $500,000 taxable capital gains after 1997.

To compare home sale rates before and after TRA97 for parcels with capital gains below $0 or above $500,000, I also impute simple means of sale rates by capital gains categories for the pre-TRA97 and post-TRA97 period respectively. Because the maximum exclusion amount was $125,000 before 1997 and $500,000 (for married couples) or $250,000 (for singles) after 1997, I focus on five capital gains categories: less than $0, $0 to $125K,

$125K to $250K, $250K to $500K, and over $500K. Figure 8 displays the results, where circles represent the pre-TRA97 data and triangles represent the post-TRA97 data. Shaded areas cover the range between two standard errors above and two standard errors below the point estimates. Thus, overlapping gray areas imply that the point estimate using the pre-TRA97 data is not statistically different from the point estimate using the post-TRA97 data for that particular capital gains category. Note that the standard errors of estimated semiannual home sale rates for houses with capital gains over $500,000 before 1997 and for houses with negative capital gains after 1997 are very large due to the lack of many observations in these categories. In fact, there are only 830 observations that have negative capital gains during the post-TRA97 period, compared with 158,598 such observations during the pre-TRA97 period. Similarly, there are only 1,242 observations with over $500,000 capital gains before 1997, whereas there are 134,463 such observations after 1997.

The results shown in Figure 8 suggest that home sale rates of parcels with capital gains between $0 and $125,000 increased after 1997 and the difference is statistically significant, which is consistent with our prediction. Interestingly, for parcels with capital gains above $500,000, sale rates declined after 1997 and the difference is also statistically significant. This finding suggests that for homeowners with extraordinary capital gains, the effect of eliminating the roll-over rule may have outweighed the effect of providing the $500,000 exclusion. Lastly, for parcels with negative capital gains or with capital gains between $125,000 and $500,000, average sale rates after TRA97 are indistinguishable from those before TRA97 in this simple non-parametric framework.

4.2 A Difference-in-Differences Framework

Even though the non-parametric results seem to suggest that TRA97 had an important impact on home sales and the impact varied with capital gain levels, they do not take into

account other factors that may have driven the difference in home sale rates between before TRA97 and after TRA97. To control for these possible confounding factors and isolate the TRA97 effect, I use a simple difference-in-differences regression model.

Because housing is both a consumption good and an investment good, I model the home sale probability as

$$\text{Prob}(Sale_{it}) = F\left(\frac{U_{it}^* - U_{it}}{U_{it}^*}, MC_{it}, r_t, h_t\right) \tag{1}$$

where $Sale_{it}$ is the binary outcome variable that equals one if homeowner i sells his home between time t and $t+1$. Because housing is lumpy and costly to adjust, a homeowner may not be able to change his housing consumption constantly. Thus, his utility from consuming the present house, which presumably was bought years ago, may be well below the utility from consuming a house that he would choose to live in if housing consumption could be adjusted in a costless manner. The first term within the $F(\cdot)$ function measures how the homeowner's utility from consuming the present home, U_{it}, compares with his utility from consuming his ideal home, U_{it}^*. The larger the difference between U_{it} and U_{it}^*, the more likely he will sell his home between time t and $t+1$ and change his housing consumption bundle (i.e. $F_1 > 0$). The second term within the $F(\cdot)$ function measures the moving cost associated with changing one's housing consumption. The higher the moving cost, the less likely the homeowner will sell the house (i.e. $F_2 < 0$). Because housing is also an investment good, both the expected housing appreciation rate, h_t, and the expected return to alternative investment opportunities, r_t, will influence the homeowner's selling decision. In this paper, I assume adaptive expectation and h_t is measured by the real housing appreciation between time $t-1$ and time t.

Capital gains taxes affect home selling decisions by increasing the moving cost, MC_{it}.

I specify MC_{it} as

$$MC_{it} = G(X_{it}, Z_{it}, Tax_{it}) \tag{2}$$

where X_{it} are homeowner i's characteristics and Z_{it} are characteristics of the present home. For example, moving costs may be higher for bigger houses or larger households. Tax_{it} is the amount of capital gains taxes homeowner i will have to pay if he sells his home between time t and $t+1$. Thus, the higher Tax_{it} is, the higher the moving cost becomes, and the lower the sale probability will be.

$$\frac{\partial \text{Prob}(Sale_{it})}{\partial Tax_{it}} = F_2 \cdot \frac{\partial MC_{it}}{\partial Tax_{it}} < 0 \tag{3}$$

TRA97 introduced arguably exogenous changes to tax treatment of housing capital gains, and such changes were different for houses with different levels of capital gains. Therefore, I can use these changes to identify the effect of capital gains taxes on home sales. Specifically, I estimate the following Probit model[13]

$$\begin{aligned}
\text{Prob}(Sale_{ict}) = \Phi \Bigg(& \alpha_0 + \sum_{k=1}^{23} \alpha_k \cdot \mathbf{1}(TCG_{ict} \in C_k) + \beta_1 \cdot \mathbf{1}(TCG_{ict} \leq 0) \cdot After_t \\
& + \beta_2 \cdot \mathbf{1}(0 < TCG_{ict} \leq 125K) \cdot After_t + \beta_3 \cdot \mathbf{1}(125K < TCG_{ict} \leq 250K) \cdot After_t \\
& + \beta_4 \cdot \mathbf{1}(250K < TCG_{ict} \leq 500K) \cdot After_t + \beta_5 \cdot \mathbf{1}(TCG_{ict} > 500K) \cdot After_t \\
& + \gamma_1 h_{ict} + \gamma_2 log(lotsize)_{ict} + \delta_c + \theta_t \\
& + \sum_{j=1}^{4} \rho_{1j} (RPP_{ict})^j + \sum_{j=1}^{4} \rho_{2j} (RCP_{ict})^j + \sum_{j=1}^{4} \rho_{3j} (T_{ict})^j \Bigg)
\end{aligned} \tag{4}$$

where $Sale_{ict}$ indicates whether homeowner i in city c sells his house between time t and $t+1$. Function $\mathbf{1}(\cdot)$ returns one if the condition expressed in the parenthesis is true and zero otherwise. $After_t$ indicates whether the observation is in the post-TRA97 period. Because

[13] I estimate a Probit model instead of a linear probability model because the average semiannual home sale rate is only 0.024 and very far from 0.5. A linear probability model may be biased in this case.

I do not observe homeowners' characteristics in my dataset, I use "time since purchase" - T_{ict} - to control for the difference between the utility from consuming the desired house and the utility from consuming the present house (i.e. $\frac{U^*_{it}-U_{it}}{U^*_{it}}$). The intuition is that the longer one stays in one's home, the more likely that he has experienced shocks that affect housing demand (e.g. change in family size).[14] I use the real purchase price of the present home, RPP_{ict}, as a proxy for permanent household income. The rationale is that permanent household income is a key determinant of housing demand - households with higher incomes can afford more expensive homes. I also use the real current price, RCP_{ict}, as a proxy for wealth. To control for these three variables as flexibly as possible, I put their polynomials to the fourth order in the regression model as explanatory variables. In addition, $log(lotsize)_{ict}$ measures how large the parcel is. Because larger parcels imply higher moving cost, we expect γ_1 to be negative. h_{ict} is the real housing appreciation from time $t-1$ to time t. δ_c and θ_t stand for city fixed effects and year fixed effects, respectively.

As described before, home sellers with positive capital gains were subject to taxes before 1997. The age-55 rule provided a one-time exclusion of $125,000 to home sellers of age 55 or older. After 1997, home sellers who are single filers may claim a capital gain exclusion of $250,000, and married joint filers may claim a $500,000 exclusion. Therefore, I want to compare home sale rates before and after TRA97 for five capital gains categories: $TCG_{ict} \leq 0$, $0 < TCG_{ict} \leq 125K$, $125K < TCG_{ict} \leq 250K$, $250K < TCG_{ict} \leq 500K$, and $TCG_{ict} > 500K$. Instead of controlling just for these five capital gains categories, I assign capital gains into 23 categories: less than -$50K, -$50K-$0, $0-$25K, $25K-$50K, $50K-$75K, $75K-$100K, $100K-$125K, $125K-$150K, $150K-$175K, $175K-$200K, $200K-$250K, $250K-$300K, $300K-$350K, $350K-$400K, $400K-$450K, $450K-$500K, $500K-

[14]Sinai (1998) employs a (s,S) model to describe housing consumption patterns. Because adjusting housing consumption is costly, home buyers make decisions in response to expected future needs rather than current needs. Therefore, the home is not necessarily ideal on the date of purchase, and $\frac{U^*_{it}-U_{it}}{U^*_{it}}$ may be a non-linear function of "time since purchase".

$600K, $600K-$700K, $700K-$800K, $800K-$900K, $900K-$1M, $1M-$1.5M, and more than $1.5M. In this way, I allow for more flexibility in estimating the baseline effect of capital gains on home sales in the pre-TRA97 period. These 23 categories are expressed as $\{C_k, k = 1, 2, ..., 23\}$ in equation (4). As discussed in previous sections, TRA97 should have no impact on houses with negative capital gains since no taxes were due for these houses throughout the sample period. TRA97 unambiguously reduced tax liabilities and should increase home sale rates if capital gains were between $0 and $500,000. The effect of TRA97 is *a priori* ambiguous on houses with capital gains above $500,000. In summary, we expect $\beta_1 = 0$, $\beta_2 > 0$, $\beta_3 > 0$, $\beta_4 > 0$, and $\beta_5 \lesseqgtr 0$. To the extent that there are many single homeowners in the sample, β_4 may be smaller in magnitude than β_2 and β_3 because single homeowners can only exclude $250,000 capital gains.

Table 4 displays the estimation results with standard errors clustered at parcel level. The numbers shown in the table are the marginal effects of corresponding explanatory variables evaluated at means. For ease of exposition, I show these marginal effects in percentage terms. In column (1), I include only year dummies, the 23 TCG category dummies, and interactions of TRA97 with the five TCG category dummies as explanatory variables. As predicted, the estimated coefficient $\hat{\beta}_1$ is statistically indistinguishable from zero at conventional confidence level, whereas $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ are all positive and statistically significant. These results suggest that TRA97 raised sale rates of parcels with capital gains between $0 and $500,000, but did not change sale rates of parcels with negative capital gains. On the other hand, the estimated coefficient on the interaction between TRA97 and capital gains exceeding $500,000, $\hat{\beta}_5$, is negative and statistically significant. It implies that for parcels with capital gains over $500,000, the elimination of the roll-over rule reduced sale rates by more than the increase in sale rates induced by the large exclusion level. Therefore, we observe a net decrease in sale rates among these parcels after TRA97.

In column (2), I add more controls to the regression model, including real semiannual housing appreciation rate, log of lot size, polynomials of real purchase price, polynomials of real current price, polynomials of time since purchase, and city dummies. The estimated coefficient on log of lot size is negative and statistically significant, suggesting that larger houses have lower sale rates, possibly due to higher moving costs. The estimated coefficient on housing appreciation rate is also negative and statistically significant, which is consistent with the notion that housing is partly an investment good. For example, during a housing market boom, adaptive expectations imply that homeowners extrapolate the housing price movement and expect the boom to continue. Therefore, they often hold on to their homes and defer sales. Similar to the estimation results shown in column (1), $\hat{\beta}_1$ is statistically indistinguishable from zero, $\hat{\beta}_2$ and $\hat{\beta}_3$ are positive and statistically significant, and $\hat{\beta}_5$ is negative and statistically significant. $\hat{\beta}_4$ remains positive but is no longer statistically significant.

Column (3) has the same specification as column (2) except that I control for city×year fixed effects to allow for city-year specific shocks to the housing market. All estimated coefficients remain almost identical to the results obtained using the column (2) specification. The estimated marginal effects of the interaction terms suggest that TRA97 increased semiannual sale rates by 0.33-0.54 percentage points for parcels with capital gains between \$0 and \$500,000, representing a 13-22 percent increase from the average sale rate during the pre-TRA97 period. In contrast, TRA97 reduced sale rates of parcels with more than \$500,000 capital gains by 0.79 percentage points, representing a 24 percent decline from the average sale rate during the pre-TRA97 period. Taken together, the evidence shown in this section suggests that TRA97 reversed the lock-in effect of capital gains taxes for houses with low or moderate capital gains. However, it may have generated an unintended lock-in effect for houses with very large capital gains due to the elimination of the roll-over rule.

4.3 Robustness Checks and Extensions

In the previous section, I find that TRA97 appears to have an important effect on home sales. In particular, home sale rates of parcels with capital gains between \$0 and \$500,000 increased and home sale rates of parcels with capital gains above \$500,000 decreased after TRA97. In this section, I present a set of robustness checks and examine how the effect of TRA97 evolved over time.

4.3.1 Potential Measurement Errors in Taxable Capital Gains

The empirical strategies used in this paper depend on taxable capital gains being accurately measured. Compared with most public surveys, the data used in this paper have much better measures of taxable capital gains because they are imputed using actual sale records and ZIP-code level housing price indexes. Nevertheless, to the extent that home improvements and renovations affect taxable capital gains, the data used in this paper may still contain measurement errors because I do not observe these improvement and renovation activities and because the Case-Shiller price indexes assume constant house quality. To ascertain that the estimates shown above are not driven by such measure errors in taxable capital gains, I carry out several robustness checks.

First, I drop houses that are likely to have had significant improvements and renovations during the sample period. Recall that the dataset is a panel of single-family houses where I observe actual sale prices if they are sold and I also observe assessment values at the end of the sample period. Suppose a parcel was first purchased in the first half-year of 1990 at price P^0_{1990} and then was sold in the second half-year of 2000 at price $P^0_{2000.5}$. I impute the values of this parcel at each point of time since 1990 using ZIP-code level semiannual housing appreciation rates. In other words, I use P^0_{1990} as the base to obtain house values

$\{P^1_{1990.5}, P^1_{1991}, P^1_{1991.5}, ..., P^1_{2000}, P^1_{2000.5}\}$. Similarly, I use $P^0_{2000.5}$ as the base to obtain house values $\{P^1_{2001}, P^1_{2001.5}, P^1_{2002}, ..., P^1_{2006}, P^1_{2006.5}\}$. If the actual sale price in the second half-year of 2000, $P^0_{2000.5}$, is very different from the imputed house value at that time, $P^1_{2000.5}$, it is likely that the parcel experienced significant modifications between 1990 and 2000. To reduce biases introduced by such measurement errors, I drop all observations on this parcel for years between 1990 and 2000. In addition, if the FY2007 assessment value of this parcel, P^{av}_{FY2007}, is very different from the imputed house value at the end of the sample period, $P^1_{2006.5}$, it is likely that the parcel experienced major changes between 2000 and 2006. In this case, I drop observations on this parcel for all years between 2000 and 2006.

Second, I change the cut-off point for the \$500,000 capital gains exclusion. Suppose that a homeowner purchased the house at P^0_t, renovated the house during the years when he lived in it, and then sold the house to someone else at price $P^0_{t'}$. According to IRS rules, the cost of renovation can be subtracted from the sale price $P^0_{t'}$ when calculating the taxable capital gains. But the added value to the house due to renovation also appreciated between the time of renovation and t', which contributes to actual taxable capital gains. Because I do not observe the timing of the renovation and the amount of money this homeowner spent on renovating the house, the predicted taxable capital gains of this house at time t' are lower than the actual taxable capital gains. In other words, the homeowner may have accumulated \$500,000 capital gains before t' even though my calculation suggests that his accumulated capital gains reached \$500,000 at time t'.[15] To correct such a discrepancy, I change the cut-off point from \$500,000 to \$450,000 as a robustness check.

Panel A in Table 5 displays the estimation results of the robustness checks described

[15]For example, suppose a homeowner bought a \$500,000 house in 1995. He put in \$50,000 to renovate the house right after he bought the house. Ten years later, house values doubled in his area and the house is worth \$1,100,000. As a result, his capital gains are \$550,000 (=\$1,100,000-\$500,000-\$50,000). But because I don't observe his renovation activities, I predict his house value to be \$1,000,000 and his capital gains \$500,000 in 2005.

above. In column (2), I drop observations if actual sale prices and predicted sale prices differ by 100% or if FY2007 assessment values and predicted values at the end of the sample period differ by 100%. Before this procedure, the correlation between predicted values and actual values was 0.80. It increased to 0.91 after this procedure. Column (2) shows that the estimated marginal effects of TRA97 on home sales are very similar to the main results obtained from the original sample. In column (3), I drop observations when actual sale prices and predicted sale prices differ by 50% or when FY2007 assessment values and predicted values at the end of the sample period differ by 50%. After this procedure, the correlation between predicted values and actual values increased further to 0.95. In this case, the estimated marginal effects of TRA97 change somewhat in magnitude, but they are not statistically different from the main results shown in column (1). Column (4) shows the estimated marginal effects when the cut-off point is changed from $500,000 to $450,000. The results remain largely the same. The results of robustness checks shown in panel A of Table 5 suggest that the measurement error problem associated with taxable capital gains is unlikely to be severe and the main findings shown in the previous section are robust to potential measurement errors.

4.3.2 Alternative Sample and Specification

The results shown in Table 4 may have over-estimated the effect of TRA97 if homeowners in 1996 anticipated the passage of TRA97. If they indeed knew that TRA97 was going to pass, homeowners who intended to sell their homes might have delayed selling their homes until after May 7, 1997 if their capital gains were relatively low. Alternatively, if they had huge capital gains, they might have accelerated selling their homes to take advantage of the roll-over rule. Under these circumstances, the findings presented in Table 4 would be artificial rather than real. To deal with such "anticipation" effects, I drop the 1996 observations and

re-estimate equation (4). The estimation results are shown in column (2) of panel B in Table 5. The estimated coefficient $\hat{\beta}_1$ is still not statistically different from zero. $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ are all positive and statistically significant. $\hat{\beta}_5$ is negative and also statistically significant. The magnitudes of the marginal effects are similar to the main results shown in column (1). Thus, the main findings shown in column (1) are unlikely to be driven by the anticipation effect. In fact, Dai, Maydew, Shackelford and Zhang (2006) suggest that the passage of TRA97 was unexpected.

The dataset analyzed in this paper contains houses that were sold between 1982 and 2006. If houses that were first sold before TRA97 are systematically different from houses that were first sold after TRA97, the results shown in the previous section may be driven by such compositional changes. To deal with this concern, I estimate the model using houses that were sold at least once before TRA97. This procedure makes sure that the pre-TRA97 and post-TRA97 observations in the sample are the same parcels. Column (3) of panel B in Table 5 presents the estimation results. The estimated marginal effects are remarkably similar to the results obtained using the original sample, suggesting that it is unlikely that systematic differences exist between houses whose first observed sales occurred before TRA97 and those whose first observed sale occurred after TRA97.

In estimating equation (4), I control for polynomials of real purchase prices, real current prices, and time since purchase up to the fourth order. To allow for more flexibility in these polynomial controls, I include polynomials up to the sixth order as a robustness check. Column (4) of panel B in Table 5 displays the estimation results of this more flexible specification. The estimated marginal effects of TRA97 are virtually identical to the results obtained when I control for polynomials only to the fourth order, suggesting that fourth order polynomials provide ample flexibility for these control variables to affect home sale rates in the main specification.

4.3.3 Dynamic Effect of TRA97

The effect of TRA97 on home sales during the years immediately following 1997 may be different from the effect of TRA97 many years after the law change. In fact, Biehl and Hoyt (2007) find intriguing evidence of the dissipating effect of TRA97. Comparing mobility of homeowners over and under 55 before and after TRA97, they show that TRA97 reversed the lock-in effect of capital gains taxes for homeowners under age 55. However, they find that such an impact of TRA97 disappeared in a few years after 1997, suggesting that the effect of TRA97 was temporary rather than permanent. To investigate the short-term effect of TRA97, I estimate equation (4) using only data within a narrow window of the law change.

In column (2) of Table 6, I limit the sample to 18 months before and 18 months after TRA97.[16] The estimated marginal effects suggest that home sale rates of parcels with capital gains between \$0 and \$500,000 increased within 18 months of the law change. The magnitudes of these effects are similar to the main findings shown in column (1). Home sale rates of parcels with capital gains above \$500,000 also appear to increase, but the effects are not statistically significant. In column (3), I limit the sample to 3 years before and 3 years after TRA97. For parcels with capital gains between \$0 and \$500,000, the estimated marginal effects of TRA97 are very similar to the results obtained using the 18-month window. However, the estimated effect of TRA97 on parcels with capital gains over \$500,000 becomes essentially zero when using the 3-year window. In summary, it appears that the unlocking-effect of TRA97 on parcels with capital gains between \$0 and \$500,000 manifested shortly after the law change. In contrast, the short-run effect of TRA97 on parcels with capital gains over \$500,000 seems to be insignificant.

[16] Note that the variable $(TCG \leq 0) * TRA97$ is dropped for collinearity reasons in this smaller sample. Therefore, β_1 cannot be estimated in practice.

To further investigate the long-term versus short-term effect of TRA97 on home sales, I estimate the following model where the long-term effect of TRA97 is allowed to be different from the short-term effect of TRA97.

$$\begin{aligned}
\text{Prob}(Sale_{ict}) = \Phi \Bigg(& \alpha_0 + \sum_{k=1}^{23} \alpha_k \cdot \mathbf{1}(TCG_{ict} \in C_k) + \beta_1^{early} \cdot \mathbf{1}(TCG_{ict} \leq 0) \cdot Early_t \\
& + \beta_1^{later} \cdot \mathbf{1}(TCG_{ict} \leq 0) \cdot Later_t + \beta_2^{early} \cdot \mathbf{1}(0 < TCG_{ict} \leq 125K) \cdot Early_t \\
& + \beta_2^{later} \cdot \mathbf{1}(0 < TCG_{ict} \leq 125K) \cdot Later_t + \beta_3^{early} \cdot \mathbf{1}(125K < TCG_{ict} \leq 250K) \cdot Early_t \\
& + \beta_3^{later} \cdot \mathbf{1}(125K < TCG_{ict} \leq 250K) \cdot Later_t + \beta_4^{early} \cdot \mathbf{1}(250K < TCG_{ict} \leq 500K) \cdot Early_t \\
& + \beta_4^{later} \cdot \mathbf{1}(250K < TCG_{ict} \leq 500K) \cdot Later_t + \beta_5^{early} \cdot \mathbf{1}(TCG_{ict} > 500K) \cdot Early_t \\
& + \beta_5^{later} \cdot \mathbf{1}(TCG_{ict} > 500K) \cdot Later_t + \gamma_1 h_{ict} + \gamma_2 log(lotsize)_{ict} + \delta_c + \theta_t \\
& + \sum_{j=1}^{4} \rho_{1j} (RPP_{ict})^j \sum_{j=1}^{4} \rho_{2j} (RCP_{ict})^j + \sum_{j=1}^{4} \rho_{3j} (T_{ict})^j \Bigg)
\end{aligned} \tag{5}$$

where $Early_t$ is an indicator variable that equals one if t falls within an initial period after TRA97, and $Later_t$ is an indicator variable that equals one if t falls out of the initial period after TRA97. If TRA97 unlocked homeowners with capital gains between \$0 and \$500,000 but the effect is transitory, then we would expect $(\beta_i^{early}, i = 2, 3, 4)$ to be positive and $(\beta_i^{later}, i = 2, 3, 4)$ to be zero. Similarly, if TRA97 unintentionally locked in homeowners with capital gains over \$500,000 but the effect is transitory, then we would expect β_5^{early} to be negative and β_5^{later} to be zero.

Column (4) and column (5) of Table 6 display the estimation results of equation (5) where the initial period is defined as 18 months and 3 years after TRA97, respectively.[17] The estimates presented in column (4) and column (5) show two interesting patterns. First, the estimated coefficients $(\hat{\beta}_i^{early}, i = 2, 3, 4)$ are indeed positive and statistically significant, whereas the estimated coefficients $(\hat{\beta}_i^{later}, i = 2, 3, 4)$ are much smaller in magnitude and

[17]Note that the variables $\mathbf{1}(TCG_{ict} \leq 0) \cdot Early_t$ and $\mathbf{1}(TCG_{ict} \leq 0) \cdot Later_t$ are dropped for collinearity reasons. Therefore, β_1^{early} and β_1^{later} cannot be estimated in practice.

statistically indistinguishable from zero. This pattern suggests that TRA97 raised selling probabilities among homeowners with capital gains between $0 and $500,000 who were presumably locked in by capital gains taxes prior to TRA97. Such an unlocking effect of TRA97, however, was achieved within a short initial period after the passage of TRA97. After this initial period, there was no evidence that home sale rates under the new tax regime were significantly higher compared to the pre-TRA97 period. Second, the estimates of short-term effect of TRA97 on houses with capital gains above $500,000, $\hat{\beta}_5^{early}$, are statistically insignificant. In contrast, the estimates of the long-term effect, $\hat{\beta}_5^{later}$, are negative, statistically significant, and large in magnitude compared with $\hat{\beta}_5^{early}$. This pattern suggests that sale rates of houses with massive capital gains responded little to TRA97 right after 1997. In the long-run, however, TRA97 appeared to reduce sale rates of these houses significantly. Such a delayed effect of TRA97 may be because homeowners were not fully informed about the implications of TRA97 immediately after the law change.[18]

In summary, the robustness checks shown in this section lend support to the main findings that TRA97 increased home sale rates of parcels with capital gains between $0 and $500,000 and reduced home sale rates of parcels with capital gains over $500,000. When I allow the effect of TRA97 to vary over time, I find evidence suggesting that the release effect of TRA97 on parcels with relatively low capital gains was completed within a short period of time after the law change, suggesting that this effect was transitory. For parcels with extraordinary capital gains, however, the effect of TRA97 was initially insignificant but became stronger in later years, suggesting that the unintended lock-in effect of TRA97 on houses with capital gains over $500,000 could remain relevant in future years.

[18] Anecdotal evidence suggests that, as of now, many homeowners still do not fully understand the $500,000 capital gains exclusion provision.

4.4 Estimating the Tax Elasticity of Home Sales

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 use the legislative changes in the top capital gains tax rate during the post-TRA97 period to estimate the tax elasticity of home sales.[19]

To estimate $\frac{\partial \text{Prob}(Sales_{ict})}{\partial Tax_{ict}}$, I need to impute Tax_{ict} for every homeowner in the sample. Since I do not have homeowner characteristics and income data to infer actual marginal tax rates, some assumption on homeowners' marginal tax rates is necessary to impute their housing capital gains tax liabilities. As shown in Table 1, homeowners in the 16 cities and towns studied in this paper are mostly high-income individuals, so it is reasonable to assume that they face the top capital gains tax rate. This assumption allows me to calculate the amount of taxes that a homeowner would owe if he were to sell his house within the next six months:

$$Tax_{ict} = Real(\tau_t \cdot \max(0, TCG_{ict} - 500,000)), \quad \begin{cases} \tau_t = 0.20 & \text{for 1998 - 2000} \\ \tau_t = 0.18 & \text{for 2001 - 2002} \\ \tau_t = 0.15 & \text{for 2003 - 2006} \end{cases}$$

where function $Real(\cdot)$ converts nominal dollar amounts into real 2000 dollar amounts. Panel A in Table 7 displays the summary statistics of such imputed Tax_{ict}. Because the first \$500,000 capital gains can be excluded at the time of a sale and because almost 80% of observations during the post-TRA97 period have capital gains below \$500,000, only 20% of observations have non-zero Tax_{ict}. The average capital gains taxes are \$8,205 for the full

[19] Estimating the tax elasticity of home sales for the pre-TRA97 data is very difficult because capital gains tax liabilities depended on age of the seller and value of the replacement home, neither of which is observed in my data.

post-TRA97 sample and $39,936 for non-zero observations.

To estimate the elasticity of home sales with respect to capital gains taxes, I use the Probit model

$$\begin{aligned}\text{Prob}(Sale_{ict}) = \Phi \Bigg(& \alpha_0 + \sum_{k=1}^{23} \alpha_k \cdot \mathbf{1}(TCG_{ict} \in C_k) + \lambda Tax_{ict} + \gamma_1 h_{ict} \\ & + \gamma_2 log(lotsize)_{ict} + \delta_c + \theta_t + \sum_{j=1}^{4} \rho_{1j} (RPP_{ict})^j + \sum_{j=1}^{4} \rho_{2j} (RCP_{ict})^j \\ & + \sum_{j=1}^{4} \rho_{3j} (T_{ict})^j \Bigg) \end{aligned} \tag{6}$$

where the key parameter to be estimated is λ. Panel B in Table 7 presents the results from estimating equation (6). In column (1), I include all observations in the sample and find the estimated coefficient on capital gains taxes, $\hat{\lambda}$, negative and statistically significant. The estimated marginal effect suggests that a $10,000 increase in capital gains tax liabilities reduces semiannual home sale rates by 0.16 percentage points, representing a 7% decline from the average semiannual sale rate of 2.4 percentage points. Because legislative changes in top rate do not generate variations in tax liabilities if capital gains are below the exclusion level of $500,000, it is useful to examine the tax elasticity among houses with strictly positive capital gains taxes. Column (2) reports the estimation result when observations with zero taxes are dropped. The estimated marginal effect becomes larger. The magnitude suggests that a $10,000 increase in tax liabilities reduces semiannual sale rates by 0.25 percentage points, a 13% decrease from the average semiannual sale rate of these houses.

The estimates of the reduced-form tax elasticity of home sale rates allow us to do policy simulations and to infer the impact of hypothetical changes in housing capital gains taxation on home sales. For example, according to the 2004 Survey of Consumer Finances (SCF), 3.26% of homeowners in the U.S. would be subject to capital gains taxes if they

were to sell their homes at the time of survey. Among these homeowners, the median tax liability is approximately $30,000, assuming a 15% capital gains tax rate. The estimate shown in column (2) of Table 7 suggests that if we eliminate capital gains taxes on housing altogether, the semiannual home sale rate would increase by 0.75 percentage points among the 3.26% homeowners who have positive tax liabilities. Another interesting scenario is when the JGTRRA expires in 2011 and the top capital gains tax rate may increase from the current 15% to 20%. Using the 2004 SCF statistics, this change in top tax rate would increase tax liabilities by $10,000 for the median homeowner among the 3.26% homeowners who have capital gains over $500,000. The estimate shown in column (2) of Table 7 suggests that the semiannual home sale rate would decrease by 0.25 percentage points as a result among these homeowners. In summary, even though we need to be cautious when making out-of-sample predictions using the estimates shown this section, it is helpful to have a tightly estimated elasticity of home sales with respect to housing capital gains taxes for many back-of-envelope calculations.

5 Conclusion

TRA97 introduced the largest change in decades to the tax treatment of housing capital gains in the United States. While researchers have started to use it as a policy instrument to identify the lock-in effect of capital gains taxes in housing markets, existing empirical studies have rarely looked beyond the age-55 rule due to data limitations. In this paper, I collect data from various sources, including local assessment records, ZIP-code level housing price indexes, and sales data on single-family houses, to construct a unique panel of houses where housing capital gains can be imputed more accurately than possible in most survey datasets. Instead of relying on the age-55 rule, I identify the tax lock-in effect by exploiting the cross-sectional variation in accumulated capital gains and the exogenous change in exclusion levels

brought forward by TRA97. This paper is the first study, to my knowledge, to examine the effect of TRA97 on houses with capital gains over $500,000. It is also the first to estimate the tax elasticity of home sales using post-TRA97 data.

I find robust 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 semiannual sale rate of these houses increased by 0.33-0.54 percentage points, representing a 13-22 percent increase from the average semiannual sale rate during the pre-TRA97 period. However, TRA97 appeared to have generated an unintended lock-in effect on houses with capital gains above $500,000. The semiannual sale rate of these houses declined by 0.79 percentage points after 1997, equivalent to a 24 percent decrease from the average semiannual sale rate during the pre-TRA97 period. This empirical finding suggests that although TRA97 raised home sale rates by allowing for a large capital gains exclusion, it also reduced home sale rates of houses with massive capital gains through the elimination of the roll-over rule. Overall, homeowners who had accumulated more than $500,000 capital gains became less willing to sell their houses after TRA97. Furthermore, the unlocking effect of TRA97 on houses with relatively low capital gains dissipated shortly after 1997, but the unintended lock-in effect of TRA97 on houses with massive capital gains appears to be long-lasting. I also estimated the tax elasticity of home sales during the post-TRA97 period, using legislative changes in top capital gains tax rates. The estimation results suggest that a $10,000 increase in tax liability reduces semiannual sale rates by 0.16-0.25 percentage points, a 7-13 percent decline from the average level. These estimates are useful for simulations of hypothetical reforms such as eliminating taxes on housing capital gains or increasing the top capital gains tax rate.

This paper brings new evidence to the literature 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 necessarily representative of the U.S. population. For example, they are better educated and tend to have higher incomes and house values. Therefore, they may have responded to TRA97 differently from homeowners in other places and of different characteristics.

The field calls for more research to fully understand the welfare impact of TRA97. First, this paper does not take into account any general equilibrium effect 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 dataset 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 in the pre-TRA97 tax regime distorted homeowners' mobility and housing consumption decisions and to understand the extent to which TRA97 reversed the lock-in effect of housing capital gains.

References

[1] **Bier, Thomas, Ivan Maric, and Winifred Weizer.** 2000. "A Preliminary Assessment of the New Home Seller Capital Gains Law." *Housing Policy Debate*, 11(3): 645-673.

[2] **Biehl, Amelia, and William Hoyt.** 2007. "The Taxpayer Relief Act of 1997 and Homeownership: Is Smaller Now Better?" Paper presented at the annual meeting of the National Tax Association, Columbus, OH.

[3] **Burman, Leonard, Sally Wallace, and David Weiner.** 1996. "How Capital Gains Taxes Distort Homeowners' Decisions?" Proceedings of the Eighty-Ninth Annual Conference on Taxation held under the auspices of the National Tax Association, Boston, MA.

[4] **Cunningham, Christopher, and Gary Engelhardt.** 2008. "Housing Capital-Gains Taxation and Homeowner Mobility: Evidence from the Taxpayer Relief Act of 1997." *Journal of Urban Economics*, 63(3): 803-815.

[5] **Dai, Zhonglan, Edward Maydew, Douglas Shackelford, and Harold Zhang.** 2006. "Capital Gains Taxes and Asset Prices: Capitalization or Lock-In?" NBER Working Paper 12342.

[6] **Fan, Jianqing, and Irene Gijbels.** 1996. *Local Polynomial Modelling and Its Applications.* London: Chapman and Hill.

[7] **Farnham, Martin.** 2006. "Residential Mobility, the Capital Gains Tax, and the 'Taxpayer Relief Act of 1997'." Unpublished.

[8] **Goodman, John, and John Ittner.** "The Accuracy of Home Owners' Estimates of House Value." *Journal of Housing Economics*, 2(4): 339-357.

[9] **Hoyt, William, and Stuart Rosenthal.** 1990. "Capital Gains Taxtation and the Demand for Owner-Occupied Housing." *The Review of Economics and Statistics*, 72(1): 45-54.

[10] **Hoyt, William, and Stuart Rosenthal.** 1992. "Owner-Occupied Housing, Capital Gains, and the Tax Reform Act of 1986." *Journal of Urban Economics*, 32(2): 119-139.

[11] **Kiel, Katherine, and Jeffrey Zabel.** 1999. "The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing Survey." *Real Estate Economics*, 27(2): 263-298.

[12] **King, Mervyn.** 1980. "An Econometric Model of Tenure Choice and Demand for Housing as a Joint Decision." *Journal of Public Economics*, 14(2): 137-160.

[13] **Newman, Sandra, and James Reschevsky.** 1987. "Federal Policy and the Mobility of Older Homeowners." *Journal of Policy Analysis and Management*, 6(3): 402-416.

[14] **Rosen, Harvey.** 1979. "Housing Decisions and the U.S. Income Tax: An Econometric Analysis." *Journal of Public Economics*, 11(1): 1-24.

[15] **Sinai, Todd.** 1998. "Taxation, User Cost, and Household Mobility Decisions." Wharton Real Estate Center Working Paper 303.

Table 1: Characteristics of the 16 MA Cities/Towns

City/Town	Bachelor's	Graduate or Professional	Median HH Income 1999	Ownership Percentage
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 and Graduate or Professional refer to educational attainment for the population 25 years and older in 2000. Ownership Percentage is owner-occupied housing units as a percentage of all occupied housing units in 2000.

Table 2: Sale Prices and Number of Sales by Year

Year	Mean	Median	N
1982	137,084	123,500	1,405
1983	152,902	135,000	2,444
1984	186,019	164,900	2,405
1985	240,716	217,000	2,365
1986	308,295	272,000	2,345
1987	332,980	285,000	3,542
1988	361,837	300,000	3,160
1989	357,911	300,000	2,936
1990	342,556	287,500	2,612
1991	315,569	267,000	3,601
1992	322,201	275,000	3,955
1993	337,797	290,000	3,906
1994	359,474	311,250	3,960
1995	386,454	330,000	3,517
1996	407,836	348,000	3,845
1997	443,288	374,850	4,016
1998	487,614	410,000	4,246
1999	541,459	450,000	4,333
2000	659,927	532,125	3,848
2001	728,247	587,000	3,144
2002	747,042	615,000	3,650
2003	806,318	665,000	3,469
2004	893,483	720,500	3,990
2005	969,502	785,000	3,569
2006	969,513	781,000	2,621
Total			82,884

Note: Sale prices are shown in nominal terms. The number of sales in 2006 is low because I drop Newton sales in 2006. The number of sales between 1982 and 1986 is low because the 1982-1986 sales data may not be exhaustive. See the Data section for details.

Table 3: Summary Statistics of Key Variables

	Pre-TRA97 (N=507,424)		Post-TRA97 (N=654,474)	
	Mean	SD	Mean	SD
Semiannual Sale Dummy	0.024	0.153	0.024	0.153
Taxable Capital Gains (TCG)	49,616	88,931	344,707	278,541
Real Purchase Price	346,159	201,876	366,904	241,624
Real Current Price	435,740	228,898	733,265	423,058
Real Housing Appreciation Rate	0.010	0.046	0.031	0.037
Log(lot size)	9.678	0.968	9.734	0.986
(TCG$\leq$0)	0.313	0.464	0.001	0.036
(0<TCG$\leq$125K)	0.525	0.499	0.192	0.394
(125K<TCG$\leq$250K)	0.125	0.331	0.259	0.438
(250K<TCG$\leq$500K)	0.035	0.184	0.342	0.474
(TCG>500K)	0.002	0.049	0.205	0.404

Note: Taxable Capital Gains (TCG) are measured in nominal terms. Lot size is measured in unit of square footage. Housing appreciation rates refer to semiannual appreciation rates. Real dollars refer to 2000 dollars.

Table 4: Effect of TRA97 on Home Sales - Main Results

	(1)	(2)	(3)
(TCG≤0)*TRA97	1.20	0.53	0.60
	(0.75)	(0.62)	(0.64)
(0<TCG≤125K)*TRA97	0.71***	0.54***	0.54***
	(0.14)	(0.13)	(0.14)
(125K<TCG≤250K)*TRA97	0.62***	0.44**	0.45**
	(0.16)	(0.15)	(0.16)
(250K<TCG≤500K)*TRA97	0.39*	0.29	0.33
	(0.19)	(0.19)	(0.19)
(TCG>500K)*TRA97	-0.83**	-0.82**	-0.79**
	(0.29)	(0.29)	(0.30)
Housing Appreciation Rate		-1.89**	-1.99**
		(0.58)	(0.64)
Log(lot size)		-0.43***	-0.44***
		(0.03)	(0.03)
23 TCG Category Dummies	Y	Y	Y
Real Purchase Price Polynomials	N	Y	Y
Real Current Price Polynomials	N	Y	Y
Time since Purchase Polynomials	N	Y	Y
Year Dummies	Y	Y	N
City Dummies	N	Y	N
City*Year Dummies	N	N	Y
N	1,161,442	1,161,442	1,152,127

Note: All columns are Probit regressions with outcome variable equal to 1 if the house was sold in the next half-year. Marginal effects are expressed in percentage for ease of exposition. Polynomials are controlled to the 4th order. Standard errors shown in parentheses are clustered at house level. ***, **, * denote significance at the 5%, 1%, and 0.1% levels respectively.

Table 5: Effect of TRA97 on Home Sales - Robustness Checks

A. Measurement Errors in TCG

	(1) Original Sample	(2) Difference < 100%	(3) Difference < 50%	(4) Cut-off at 450K
(TCG≤0)*TRA97	0.53	0.39	0.25	0.53
	(0.62)	(0.62)	(0.65)	(0.62)
(0<TCG≤125K)*TRA97	0.54***	0.63***	0.68***	0.54***
	(0.13)	(0.14)	(0.15)	(0.13)
(125K<TCG≤250K)*TRA97	0.44**	0.42**	0.36*	0.44**
	(0.15)	(0.15)	(0.17)	(0.15)
(250K<TCG≤500K)*TRA97	0.29	0.23	0.02	0.33
	(0.19)	(0.19)	(0.20)	(0.19)
(TCG>500K)*TRA97	-0.82**	-0.79*	-1.08**	-0.64*
	(0.29)	(0.33)	(0.35)	(0.27)
N	1,161,442	1,105,591	972,902	1,161,442

B. Alternative Sample and Specification

	(1) Original Sample	(2) Drop 1996	(3) House Sold Before 1997	(4) 6th Order Polynomials
(TCG≤0)*TRA97	0.53	0.52	1.79	0.50
	(0.62)	(0.62)	(1.10)	(0.62)
(0<TCG≤125K)*TRA97	0.54***	0.51***	0.53***	0.52***
	(0.13)	(0.13)	(0.13)	(0.13)
(125K<TCG≤250K)*TRA97	0.44**	0.45**	0.43**	0.44**
	(0.15)	(0.15)	(0.15)	(0.15)
(250K<TCG≤500K)*TRA97	0.29	0.47*	0.28	0.34
	(0.19)	(0.20)	(0.19)	(0.19)
(TCG>500K)*TRA97	-0.82**	-0.76*	-0.77**	-0.81**
	(0.29)	(0.33)	(0.29)	(0.29)
N	1,161,442	1,097,625	1,063,510	1,161,442

Note: All columns are Probit regressions with outcome variable equal to 1 if the house was sold in the next half-year. Marginal effects are expressed in percentage for ease of exposition. Other control variables include housing appreciation rate, log of lot size, 23 TCG category dummies, purchase price polynomials, current price polynomials, time since purchase polynomials, city dummies, and year dummies. Polynomials are controlled to the 4th order unless indicated otherwise. Standard errors shown in parentheses are clustered at house level. ***, **, * denote significance at the 5%, 1%, and 0.1% levels respectively.

Table 6: Effect of TRA97 on Home Sales - Extensions

	(1) Original Sample	(2) 18-Month Window	(3) 3-Year Window	(4) Early = 18 Months	(5) Early = 3 Years
(TCG≤0)*TRA97	0.53	-	-		
	(0.62)	-	-		
(0<TCG≤125K)*TRA97	0.54***	0.66***	0.59***		
	(0.13)	(0.15)	(0.15)		
(125K<TCG≤250K)*TRA97	0.44**	0.58*	0.41*		
	(0.15)	(0.28)	(0.20)		
(250K<TCG≤500K)*TRA97	0.29	0.28	0.39		
	(0.19)	(0.38)	(0.28)		
(TCG>500K)*TRA97	-0.82**	0.50	-0.01		
	(0.29)	(1.02)	(0.61)		
(0<TCG≤125K)*Early				0.52***	0.51***
				(0.14)	(0.14)
(0<TCG≤125K)*Later				0.03	0.11
				(0.52)	(0.53)
(125K<TCG≤250K)*Early				0.37*	0.47**
				(0.18)	(0.16)
(125K<TCG≤250K)*Later				-0.08	-0.11
				(0.50)	(0.50)
(250K<TCG≤500K)*Early				0.66*	0.54*
				(0.28)	(0.22)
(250K<TCG≤500K)*Later				-0.25	-0.34
				(0.50)	(0.48)
(TCG>500K)*Early				0.13	-0.36
				(0.55)	(0.35)
(TCG>500K)*Later				-1.21**	-1.30***
				(0.41)	(0.39)
N	1,161,442	181,704	421,457	1,161,442	1,161,442

Note: All columns are Probit regressions with outcome variable equal to 1 if the house was sold in the next half-year. Marginal effects are expressed in percentage for ease of exposition. Other control variables include housing appreciation rate, log of lot size, 23 TCG category dummies, purchase price polynomials, current price polynomials, time since purchase polynomials, city dummies, and year dummies. Polynomials are controlled to the 4th order. Standard errors shown in parentheses are clustered at house level. ***, **, * denote significance at the 5%, 1%, and 0.1% levels respectively.

Table 7: Estimating the Tax Elasticity of Home Sales using Post-TRA97 Data

A. Summary Statistics

	N	Mean($)	Median($)	SD($)
Taxes (including zeros)	654,474	8,205	0	24,914
Taxes (excluding zeros)	134,463	39,936	26,224	41,881

B. Estimation Results

	(1) All Obs	(2) Drop Zeros
Capital Gains Taxes (in $10,000s)	-0.155*	-0.254**
	(0.073)	(0.091)
N	654,474	134,463

Note: All columns in Panel B are Probit regressions with outcome variable equal to 1 if the house was sold in the next half-year. Marginal effects are expressed in percentage for ease of exposition. Other control variables include housing appreciation rate, log of lot size, 23 TCG category dummies, purchase price polynomials, current price polynomials, time since purchase polynomials, city dummies, and year dummies. Polynomials are controlled to the 4th order. Standard errors shown in parentheses are clustered at house level. ***, **, * denote significance at the 5%, 1%, and 0.1% levels respectively.