Who stays in the District? Who leaves?
Preliminary findings from DC tax filers from 2004

by Yesim Sayin Taylor[1]
Office of Revenue Analysis
Office of the Chief Financial Officer
Government of the District of Columbia
January 28, 2015

Summary of findings

The District of Columbia is a relatively small city, in the middle of a large and growing metropolitan area, with a large transient population including interns, political appointees, and students. A sizable share of the people who have children move to the suburbs, especially after they have their first child, trading one set of benefits (shorter commute, city life, etc.) for another set (such as cheaper housing or childcare). My office recently published a study on whether first-time parents leave the city at rates faster than the rest of D.C. residents, and if this behavior has changed over time. In this short study, we look at the defining characteristics of those who remain in the city.

We tracked the behavior of D.C. residents who filed income taxes for the first time in 2004. We found the following based on their tax filings.

- **The District's population is transient.** Only 23 percent of the 42,257 tax filers who first filed in 2004 remained on the tax rolls in 2012.
- **People tend to stay if there is a change in the family structure.** Singles tend to leave and those who change their filing status, for example, because of a marriage, tend to stay.
- **Family dynamics matter beyond marriage.** We have shown elsewhere that the first child plays an important role in the decision to move out of the city. A second or a third child increases the probability that families will stay.
- **The District attracts high-income residents.** Among those who were in the highest income quintile when they arrived in the city in 2004, 41 percent were still found on the tax rolls. Only a quarter of filers who were in the lowest income quintile, however, were still on the tax rolls in 2012.

Methodology

We use data from the District's tax rolls beginning in 2004—the first year for which we can reliably build a group of "newcomers." We begin with those tax filers who first appeared in the tax rolls in 2004 and

---

track the movement of these filers through 2012—the last year for which income tax data are available.[2] Tax data show that of the approximately 315,000 tax filers in 2004, 56,910 (approximately 17 percent) first appeared in the tax rolls in 2004. We removed from the dataset those who have appeared and disappeared over the years, and only consider continuous filers, leaving 42,257 observations. Table 1 provides information on the composition of the data by year.

Table 1 - Composition of Tax Filers who first appeared in tax rolls in 2004, 2004 through 2012

	Total	65+	Single	Married	Head of Household	Dependent	Changed Status	Added Dependent
2004	42,257	4%	68%	14%	14%	4%	18%	10%
2005	30,563	3%	68%	12%	14%	5%	25%	13%
2006	23,627	4%	68%	12%	15%	5%	28%	16%
2007	19,145	4%	67%	12%	16%	5%	30%	17%
2008	16,169	4%	65%	13%	17%	5%	31%	21%
2009	13,971	4%	65%	13%	17%	5%	32%	22%
2010	12,311	4%	64%	13%	18%	5%	32%	23%
2011	10,848	4%	63%	13%	18%	5%	32%	23%
2012	9,436	4%	64%	14%	17%	5%	32%	24%

Source: DC Individual Income Tax Returns, 2004 through 2012

Only 23 percent of the tax filers who first filed in 2004 remained on the tax rolls in 2012. Those who remained in the same status in subsequent years disappeared from the rolls at a higher rate. By 2012, only 19 percent of such filers were still on the tax rolls.[3] In contrast, 40 percent of those who have changed status across any given two tax years remained by 2012.

Among those filers who added a dependent since they first filed in 2004, the "stay" rate was much higher. The middle right panel of Figure 1 shows the cumulative effect: 55 percent of those who added dependents continued to stay on the tax rolls through 2012. The comparative share for those who did not add dependents (or removed them from their tax filings) is 19 percent.

Income also appears to play a large role in the decision to stay in the city. In the graphs below, we see that among those who were in the highest income quintile when they arrived in the city in 2004, 41 percent were still found on the tax rolls in 2012. Only a quarter of filers who were in the lowest income quintile stayed. However, we see no such differences in departure rates by tax bracket: Those who were in the lowest tax bracket in 2005 (effective tax rates under 3.8 percent, after accounting for various tax credits) exited the rolls at rates similar to those in the middle bracket (between 3.8 percent and 6.1 percent) and the highest tax bracket (above 6.1 percent).

---

[2] This data was collected in 2013 as the final tax payments for tax year 2012 were made in April of 2013.

[3] Not all filers who disappeared from the tax rolls necessarily left the city. Singles who have married would have disappeared from the tax rolls if they became the spouse on the tax form. We discuss this problem more extensively in the appendix.

Figure 1 – Share of population who stayed in the rolls, by various classifications, 2004 through 2012

Source: Individual Income Tax Returns, 2004 through 2012

We use logistic regressions to identify the impact of demographic, income, and tax variables on the decision to leave the city.[4] For example, are singles more sensitive to the changes in the taxes they pay? Would married couples with children be more likely to leave the city if they lose income relative to single residents?

Demographic variables, income change, and poverty

We begin with a model where the likelihood of leaving the tax rolls is determined by demographic characteristics. Included in the analysis are indicators of filing status change (capturing a change in family structure), whether the filer added more dependents (capturing a change in family size), and whether the tax filer is a senior citizen. The analysis also includes two income related classifiers: whether the filer

---

[4] We put our explanation of the model in the Appendix on Logistic Regressions beginning on page 6 since it is a bit technical.

lost income in any of the years in which he filed taxes and whether he was in the lowest quintile for income during his time in the city.

We find that many people move into this city for only a brief period and eventually move out. The estimated odds of leaving the city is 6 to 1 (Figure 2). That is, if we know nothing about a person except that she has been a District of Columbia tax filer since 2004, our best guess is that she has an 86 percent chance of leaving the tax rolls and only a 14 percent chance of staying. The ratio of these two probabilities, 86/14, give us the odds-ratio of 6:1.

Filing a tax form with a status change would cut the odds that the same person would leave by half, to 3.12:1. Of course, the exit rate is still high given the transient nature of this city. 76 percent of filers who have changed their status since they filed their first tax return in 2004 would have moved out by 2012, but this is lower than the overall probability of 86 percent.

Having more dependents cuts the odds-ratios by two-thirds from the baseline of 6:1 to 1.84:1. Among this group, the probability of leaving the tax rolls is only 65 percent. If the person's income was among the bottom two quintiles, now his odds of leaving the tax rolls is much higher than the average at 11:1. 92 percent of first-time filers who reported incomes in the bottom two quintiles have left by 2012.[5]

Figure 2 – Estimated odds-ratios and estimated probabilities conditional on the classification group
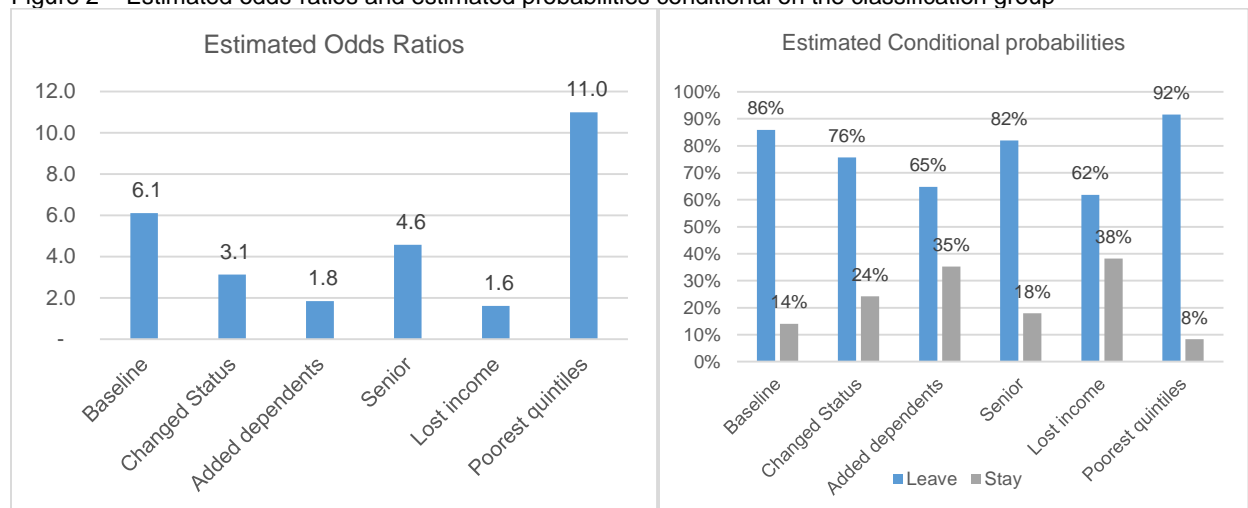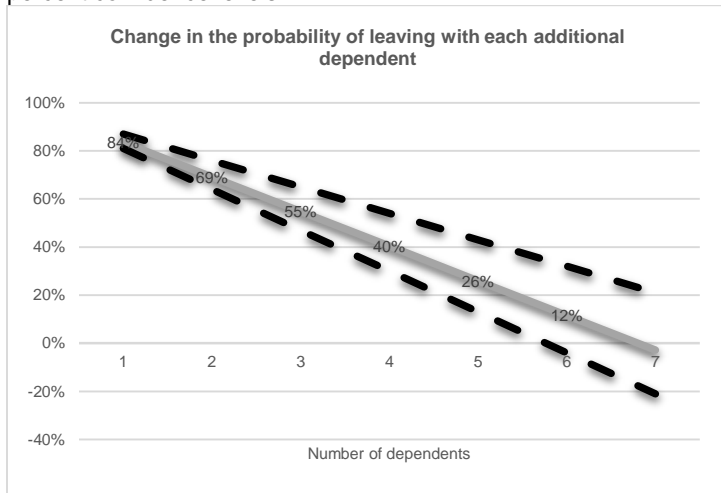

Source: Table 2 in Appendix

Recent work by Moored and Metcalf (2015) show that adding the first dependent is an indicator of leaving the city, but this effect disappears within five years of having the child. The data for the 2004 cohort shows that while the first child might act as a push factor, additional children cement families to the city. Each additional dependent (who are generally children)[6] on the tax filings reduces the probability of leaving the city by 14 percent (Figure 3). Families who do not move after their first child are less likely to

---

[5] This could be the effect of tax filing thresholds, which we plan to address in the next iteration of this paper.

[6] Out of all the dependents claimed in 2012 by people filing federal income taxes in D.C. 85 percent were children at home, 4 percent were parents, and 11 percent were other people claimed as dependents.

move when additional children come along. After all, uprooting and moving families is hard, and it becomes harder as children grow.

Figure 3 – Marginal impact of having an additional child in 2004 on the probability of leaving the city tax rolls, with 95 percent confidence levels

Change in the probability of leaving with each additional dependent

Source: Table 2 in Appendix

Income and tax variables

People from higher income quintiles are less likely to leave the tax rolls relative to those in the poorest quintile, but the effects of income appear to be small if we exclude other demographic variables. The effect is largest for the second-to-top quintile. The odds of leaving for this group is 2.76:1. The effect on the probability of leaving, however, is still muted, going down from 79 percent as measured among the poorest to 73 percent (Figure 4).

While exit rates decline by income quintile, income alone does not explain much. The model fit for income is not very strong, which tells us that we are leaving many important variables out.

Figure 4 – Estimated odds-ratios and observed conditional frequencies for leaving the city tax rolls, by income quintiles
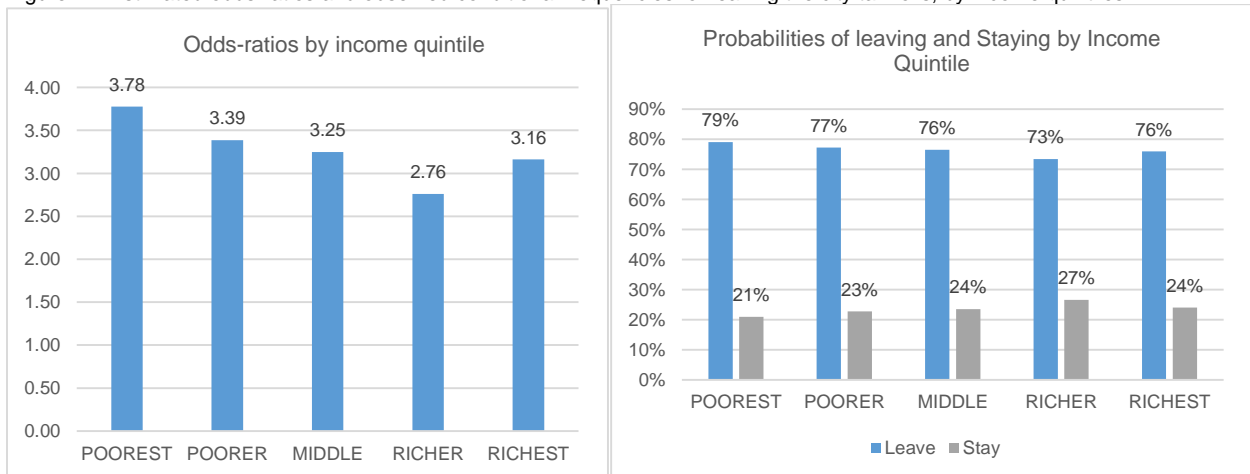
Source: Appendix Table 3

How we group people makes a difference in our ability to explain residents' decision to stay or leave. We considered the question: does it matter if someone was ever among the highest income group in the District or was ever taxed at the highest marginal rate? The analysis shows that having been in the highest income quintile cut the odds of leaving the tax rolls by more than half. Having been in the highest tax bracket, however, reduced the probability of leaving the city from 88 percent to 81 percent.

Figure 5 – Estimated odds-ratios and observed conditional frequencies for leaving the city tax rolls, by income quintiles
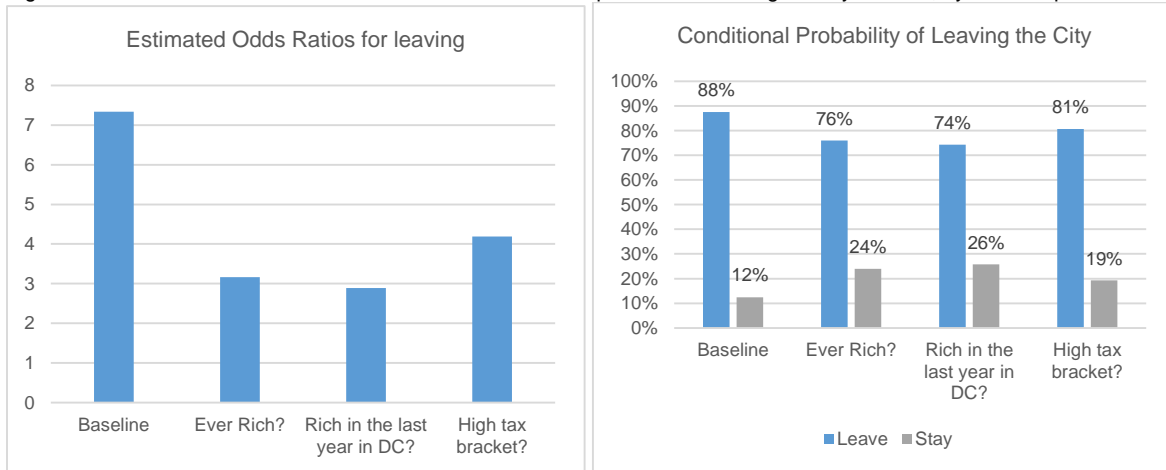
Source: Table 4 in Appendix

Conclusion and next steps

People tend to stay in the District if there is a change in their family structure. Changing filing status – because of marriage for example—and having more children increase the probability that a tax filer will stay, suggesting that family formation matters greatly in anchoring households to neighborhoods. Tax rates at higher income levels do not seem to matter—upper income families, who are in a higher income tax bracket, are less likely to leave the city compared to middle and lower income families, suggesting that they consider things other than tax rates in their decisions to leave or stay.

Our study is limited by the single cohort it investigates—those who first appeared in the tax rolls in 2004. Future investigations will include adding more cohorts including those who have lived in the city for many years.

Data Appendix

Logistic Regressions

The model takes the form:

$$p(X) = \frac{e^{\beta_o + \beta_1 X}}{1 + e^{\beta_o + \beta_1 X}}$$

which, with some arrangement, taxes the form:

$$logit\left(\frac{p(X)}{1-p(X)}\right) = \beta_o + \beta_1 X$$

where $p(X)$ is the conditional probability associated with leaving the tax rolls and β_i is the effect of group i on the logit. It is important to note that the coefficients are logits of odds-ratios and our estimated probabilities can be calculated using the above form. For example, a coefficient of 1 for $\widehat{\beta_1}$ implies a change in the odds ratio of 2.7 (e^1) and if the estimated value of $\widehat{\beta_o}$ is 3, then the probability of observing a taxfiler leaving the tax rolls would be

$$p(X) = \frac{e^{2+1X}}{1+e^{2+1X}} = \begin{cases} 0.95 \; if \; X = 1 \\ \\ 0.88 \; if \; X = 0 \end{cases}$$

It is useful to think of this as a classification exercise: an important feature of the binomial models is that all information concerning the gross effects of, for example, senior status, on staying on tax rolls is contained in the marginal distribution staying on the rolls by age. We can work with data classified by senior status, senior status and filing status, or senior status, filing status and increase in dependent counts. In all cases, the estimated effects, standard errors and the likelihood ratio tests based on differences between deviances would be the same.

The additive model on demographic variables begins with various status variables in filing status, and adds on it two indicators of income—whether the filer was in the lowest income quintile the last year the filer is present in the tax rolls, and whether he has ever lost income (move across quintiles) during his or her presence on the tax rolls. We also run a separate regression using income quintiles only and whether a filer has been in the highest tax bracket. Not included in the paper are a series of regressions on income levels, which proved to have little explanatory power.

Estimated Coefficients for different models

Table 2 – Estimated effects of status change, having more dependents, senior status, lost income and poverty status on the decision to stay or leave the city

Dependent Variable: MOVE_OUT				
Method: ML - Binary Logit (Quadratic hill climbing)				
Date: 11/06/14 Time: 12:13				
Sample (adjusted): 1 42244				
Included observations: 40036 after adjustments				
Convergence achieved after 3 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.809467	0.020585	87.90253	0.0000
CHANGE_STATUS	-0.670240	0.030458	-22.00524	0.0000
MORE_DEPENDENTS	-1.201357	0.036257	-33.13481	0.0000
_65PLUS	-0.290526	0.063927	-4.544687	0.0000

7

LOST_INCOME	-1.328563	0.026763	-49.64115	0.0000
POOR_WHEN_LEFT	0.587153	0.027195	21.59012	0.0000
McFadden R-squared	0.113803	Mean dependent var		0.765211
S.D. dependent var	0.423872	S.E. of regression		0.397019
Akaike info criterion	0.966251	Sum squared resid		6309.696
Schwarz criterion	0.967540	Log likelihood		-19336.42
Hannan-Quinn criter.	0.966659	Deviance		38672.84
Restr. deviance	43639.09	Restr. log likelihood		-21819.55
LR statistic	4966.254	Avg. log likelihood		-0.482976
Prob(LR statistic)	0.000000			
Obs with Dep=0	9400	Total obs		40036
Obs with Dep=1	30636			

Table 3 – Estimated effects of income quintiles on the odds-ratio of moving out

Dependent Variable: MOVE_OUT					
Method: ML - Binary Logit (Quadratic hill climbing)					
Date: 11/06/14 Time: 15:11					
Sample (adjusted): 1 42244					
Included observations: 40031 after adjustments					
Convergence achieved after 4 iterations					
Covariance matrix computed using second derivatives					
Variable	Coefficient	Std. Error	z-Statistic		Prob.
C	1.329295	0.025895	51.33392		0.0000
LAST_QUINTILE_POOR2	-0.109271	0.036847	-2.965565		0.0030
LAST_QUINTILE_MIDDLE	-0.151245	0.036829	-4.106670		0.0000
LAST_QUINTILE_RICH2	-0.313634	0.036511	-8.590032		0.0000
LAST_QUINTILE_RICH	-0.177725	0.038248	-4.646612		0.0000
McFadden R-squared	0.001771	Mean dependent var			0.765307
S.D. dependent var	0.423812	S.E. of regression			0.423423
Akaike info criterion	1.088091	Sum squared resid			7176.126
Schwarz criterion	1.089164	Log likelihood			-21773.68
Hannan-Quinn criter.	1.088430	Deviance			43547.35
Restr. Deviance	43624.60	Restr. log likelihood			-21812.30
LR statistic	77.24408	Avg. log likelihood			-0.543920
Prob(LR statistic)	0.000000				
Obs with Dep=0	9395	Total obs			40031
Obs with Dep=1	30636				

Table 4 – Estimated effects of income and tax rates on the odds of moving out

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.949601	0.023618	82.54600	0.0000
EVER_RICH	-0.797755	0.030900	-25.81745	0.0000
RICH_WHEN_LEFT	-0.889398	0.028388	-31.33021	0.0000
EVER_HIGH_TAX_BRACKET	-0.517532	0.029780	-17.37879	0.0000
McFadden R-squared	0.045127	Mean dependent var		0.765211
S.D. dependent var	0.423872	S.E. of regression		0.413508
Akaike info criterion	1.041008	Sum squared resid		6845.024
Schwarz criterion	1.041867	Log likelihood		-20834.90
Hannan-Quinn criter.	1.041280	Deviance		41669.81

Restr. deviance	43639.09	Restr. log likelihood	-21819.55
LR statistic	1969.285	Avg. log likelihood	-0.520404
Prob(LR statistic)	0.000000		
Obs with Dep=0	9400	Total obs	40036
Obs with Dep=1	30636		

The Singles Problem

It is important to note that not all filers who disappeared from the tax rolls necessarily left the city. Singles who have married would have disappeared from the tax rolls if they became the spouse on the tax form. In fact, of the 28,725 filers who first appeared in 2004 and filed as single, only two continuously remained single year after year. Between 2004 and 2008, the share of singles were relatively stable (Figure 2, right top panel) but between 2008 and 2009, a disproportionately large share of single filers switched their status to married. It is not clear why this is the case, but one plausible explanation is the legislation, which legalized same-sex marriages in the District, which was enacted in 2009. Overall 18,747 continuously single tax filers disappeared from the rolls—again a large share in 2010 and 2011. Another explanation related to the large reduction in filers could be income thresholds for filing tax returns. For some singles, these thresholds could have been binding in the aftermath of the great recession. The continuously married sub-population is much more stable, and the data loss among this group is much smaller (Figure 6, bottom two panels).

Figure 6 – Stability of filing status, singles, and married
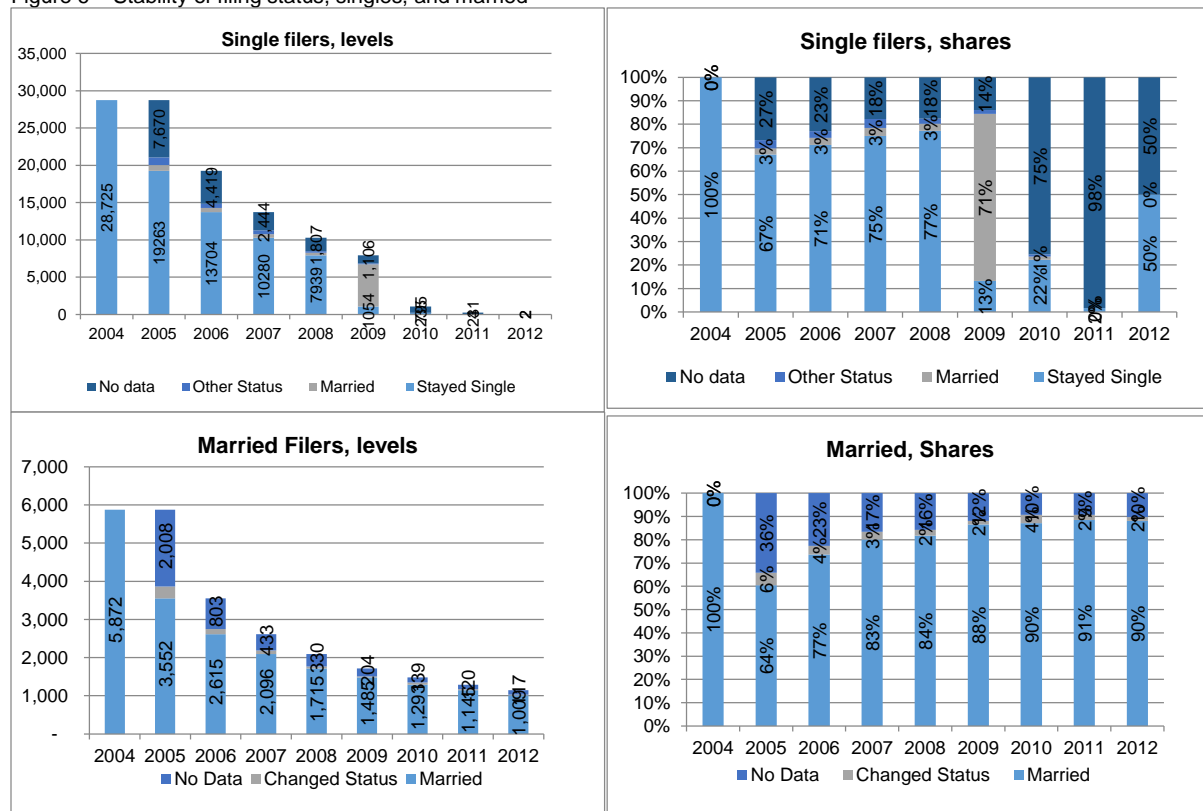
Source: Individual Income Tax Returns, 2004 through 2012 and authors' calculations