Male Earnings Volatility in LEHD

before, during, and after the Great Recession

Kevin L. McKinney and John M. Abowd[1]

U.S. Census Bureau

December 31, 2018

Abstract

Using data from the Census Bureau's Longitudinal Employer-Household Dynamics infrastructure files, we study the change in log real labor earnings and measures of its volatility for prime-age men over the period 1996 to 2015. We use a consistently defined population frame to facilitate accurate estimation of temporal changes and comparability to designed longitudinal samples of people. The Great Recession reduced earnings primarily through long spells of non-employment. Prime-age males who did not change employers and worked continuously experienced stable real earnings or growth every year. All other prime-age male workers (about 30% of the eligible population) had a cumulative loss over the same period of -0.288 log points during the Great Recession. Those with stable employment experienced very little change in volatility; whereas overall volatility for prime-age males not stably employed was about 15 times as large as for the stably employed, spiked during the Great Recession, and remained elevated thereafter.

Key words: Earnings loss in recession; earnings variability in recession; active workers; inactive workers; immigrant candidates.

I. Introduction

Using data from the Census Bureau's Longitudinal Employer Household Dynamics (LEHD) program, we estimate earnings volatility trends for prime age males from 1996 to 2015. Unlike the typical longitudinal survey data sources such as the Panel Study of Income Dynamics (PSID), LEHD data contain annual earnings for the virtual universe of private wage and salary

---

[1] The views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau. All tables and figures in this paper have been cleared by the Census Bureau Disclosure Review Board (CBDRB-FY19-118). This version has been cleared for presentation at the 2019 Allied Social Sciences Associations meetings, January 2019. Please contact the authors for updated versions before citing. Contacts: john.maron.abowd@census.gov, kevin.l.mckinney@census.gov.

workers in the United States. The large scale of the data enables a detailed analysis of worker earnings volatility, however, as shown in Abowd et al. (2018), identifier misuse makes it difficult to track some workers over time. To address this issue, we use only earnings associated with "eligible" worker identifiers issued by the Social Security Administration, allowing us to consistently estimate calendar-year time trends.

Unlike many other studies, we estimate earnings volatility both with and without years with zero earnings. This is potentially important, especially post Great Recession, when large numbers of workers were forced to transition to inactive status. One caveat with including years with zero earnings is that unlike survey data, LEHD data contains no affirmative report of zero earnings, the zero is assumed based on the absence of reported earnings. To minimize earnings under-reporting, we impute earnings for workers at firms suspected of under/non-reporting.

Most past studies on earnings inequality use a variation or extension of the estimation framework pioneered in Gottschalk and Moffitt (1994). These studies estimate an error components model with both an annual (biennial for the PSID) permanent and transitory component. Assuming the arguably strong assumptions embedded in the model are correct, estimates of average annual earnings volatility are recovered for both components. However, recent research by Jensen and Shore (2015) and Arellano, Blundell, and Bonhomme (2017) highlights that different data generating processes can create similar changes in average earnings volatility. They find that earnings volatility is not homogenous, the typical worker has relatively low earnings volatility and changes in average volatility are driven by a small subset of workers with repeated large earnings shocks. These results are consistent with our prior research showing that workers above the median earnings rarely have large year-to-year changes.

To investigate these issues, we take a flexible estimation approach. Instead of estimating an error components model or, for example, the computationally intensive estimation strategy proposed by Jensen and Shore, we first estimate a fixed effects model controlling for individual heterogeneity and age effects. The residuals from this regression, represent the difference between the actual and expected earnings change for each worker each year. Using this framework, we decompose the variance of the change in annual earnings into a component representing the composition of the workforce (both the age distribution and the set of eligible or active workers changes each year) and a residual component. Although we make no effort to estimate distinct annual permanent and transitory error components, the structure of the residuals and the autocovariance matrix provide important information about the persistence of earnings shocks. We create autocovariance matrices indexed by calendar time, allowing us to estimate persistence and year specific effects. This approach is computationally tractable, requires few assumptions, and allows us to estimate time trends as well as explore how earnings volatility varies across workers.

II. Data

The empirical work in this paper uses earnings information from the Longitudinal Employer-Household Dynamics (LEHD) infrastructure files, developed and maintained by the U.S. Census Bureau.[2] From this data source, we construct annual person-level earnings files covering the period 1995-2015.

In the LEHD data infrastructure, a "job" is the statutory employment of a worker by a statutory employer as defined by the Unemployment Insurance (UI) system in a given state. Mandated reporting of UI-covered wage and salary payments between one statutory employer and one statutory employee is governed by the state's UI system. Reporting covers private employers and state and local government. There are no self-employment earnings unless the proprietor drew a salary, which is indistinguishable from other employees in this case.

The Office of Personnel Management (OPM) supplied federal jobs data, included from 2000Q1 forward. The OPM data were edited as part of the LEHD infrastructure processing to produce records containing quarterly earnings reports comparable to those reported directly in the UI wage and salary payments. As part of this processing, pseudo-UI account numbers were created using the observed combinations of duty station state and agency/sub-element.[3] The result is a set of state-level employer identifiers conceptually similar to those found on the UI data for private firms.

[2] See Abowd et al. (2009) for a detailed summary of the construction of the LEHD infrastructure.
[3] See https://www.opm.gov/policy-data-oversight/data-analysis-documentation/data-policy-guidance/reporting-guidance/part-a-human-resources.pdf for a list of agency codes.

Due to national security regulations, which suppress certain jobs from the ones released by OPM to the public and other agencies, the coverage of the OPM extract varies by agency. Under-coverage is particularly severe for the Department of Defense (including the Air Force, Army, and Navy), Department of Justice, Department of State, and the Department of Treasury. Although the federal jobs data are typically not included as part of the state-based UI system, in this paper, when we say "UI-covered" employment, we mean "statutory employment" as defined by the UI system or a statutory federal employee.

States and the federal government joined the partnership that supplies input data to the LEHD program at different dates. When a state or the federal government joined, the data custodians were asked to produce historical data for as many quarters in the past, back to 1990Q1, as could be reasonably recovered from their information storage systems. As a result, the date that a data-supplying entity joined the partnership is not the same as the first quarter in which that entity's data appear in the system. The start date for any state or the federal government depends primarily on the amount of historical data the state or federal government could recover at the time it joined. This potential ignorability (in the sense of Rubin 1987 or Imbens and Rubin 2015) of the start date for a segment of the LEHD data is the basis for our methods of constructing nationally representative estimates back to 1995.

Although data are available for some states prior to the start date of our analysis sample, previous research in Abowd et al. 2018 shows that the earnings distribution is not representative of the entire U.S. until 1995. Table 1 shows basic information about the available data for each of the fifty states, plus the District of Columbia, and the federal government (OPM). Almost all of the large states with high earning workers (Illinois, California, Florida, New York, and Texas) are available by 1995Q1 and all states, DC and OPM are available by 2004Q1. One of the key results from Abowd et al. 2018 is that once the large states are available, the annual earnings distribution constructed using only the subset of states available by 1995Q1 is almost identical to the annual earnings distribution from 2004Q1 forward, when all states are available.

By 2004Q1 the LEHD data represent the complete universe of statutory jobs in the U.S.: all fifty states, the District of Columbia, and the federal government are reporting regularly.

Before this date, LEHD data provide a complete frame for the available states, after this date, the LEHD data provide a complete frame for the national population of UI covered jobs, including federal employees. Although the LEHD data provide us with a high-quality jobs frame, individual identifier misuse complicates the time-varying many-to-one assignment of jobs to workers. Therefore, when studying earnings volatility, it is preferable to have a person frame that covers a known population of interest, such as all persons legally eligible to work in the United States. For our analysis, we create a frame of workers using the Census Bureau's edited version of the Social Security Administration's master SSN database (the "Numident"), capturing all reported employment-eligible workers but removing jobs associated with ineligible workers, as we elaborate below. By convention, these data are called the Census Numident to distinguish them from the original SSA version.

LEHD earnings records are reported quarterly by the employing firm. These records contain a nine-digit person identifier, typically assumed to be a Social Security Number. However, at the time the report is received by the state UI office, the nine-digit person identifiers are not verified, resulting in records both with and without a valid SSN. Using the Census Numident, we ascertain if each earnings record is associated with a valid SSN. Records not associated with a valid SSN may have an alternate, valid person identifier, such as an IRS-issued Individual Taxpayer Identification Number (ITIN); nevertheless, we can only distinguish between valid and invalid SSNs. If the SSN is valid, we have access to demographic characteristics, such as sex and birth date, from the Census Numident and other Census sources. We also have an employment history from the UI wage records. If the SSN is not valid, we only have access to the employment history.

Using both the Census Numident and the employment histories from the UI data, we create a "prime-age male eligible-workers" frame, including only jobs and workers that meet the following criteria:

- valid SSN on the Census Numident;
- gender is male;
- individual is between the age of 25 and 59, inclusive;

- the year of the recorded data is greater than or equal to the SSN year of issue and less than or equal to the year of death (if available); and
- has a SSN that was associated with fewer than 12 jobs during the data year.

Every year from 1995 to 2015 in which an individual is between the ages of 25 and 59, an eligible worker is labeled as “active” in the labor market when UI earnings are positive and “inactive” otherwise.

The purpose of the prime-age male eligible-workers frame is twofold. The Census Numident data allow us to consistently identify a set of males legally eligible to work each year, while at the same time implicitly removing earning records from our analysis sample that are not associated with individuals in the covered population. We go a step further. We remove earnings records with valid SSNs where the available data strongly suggest that the SSN is not being used by the person to whom it was issued.[4] These two types of suspect nine-digit person identifiers—invalid SSNs that do not match to the Census Numident and valid SSNs apparently being used by multiple persons and/or for whom the age of the person issued the SSN is inconsistent with labor-market activity—we call “immigrant candidates.”

Table 2 contains counts (rounded to four significant digits) of our analysis sample of prime-age male eligible workers by year, broken down by labor market status—inactive, active, and never worked. The table also contains counts by year for the two largest categories of immigrant candidate records. Figure 1 plots the share of active and inactive workers in our analysis sample by year. The vertical line at 2004 represents the first year when all states, the District of Columbia, and the federal workforce are available. Prior to 2004 there is a large increase in the percent active and a large decrease in the percent inactive, due primarily to state entry. Once the jobs frame is nationally complete in 2004, about 67% of eligible males have positive earnings during the year (active) and about 22% are inactive but observed active sometime during the period 1995-2015 with the remaining 11% of eligible workers never observed with positive earnings. The effect of the Great Recession is clearly seen starting in 2008 and by 2010 the effects are fully realized, with only about 63% of eligible male workers

[4] The use of SSNs not originally issued to the person using the SSN has been documented and studied by Brown et al 2013 and others.

active and roughly 27% of previously active workers inactive. From 2010 forward, the recovery from the Great Recession is long and slow, by 2015 a large percentage of previously active workers are still inactive (26%), although the percent active is approaching the pre-recession peak (66% in 2015 vs 67% in 2007).

Figure 2 shows the immigrant-candidate records as a percentage of the active eligible worker analysis sample records by year. We norm the records using the active analysis sample to highlight the potential impact of including these records in our estimates. The average combined proportion of both UI-only active and Census Numident-active immigrant candidate records is about 9%, however this proportion is not constant over time. The proportion of job-year records where the Census Numident reports the worker as either extremely young (age<16) or extremely old (age>70) is about 7% before the Great Recession, declining to about 5-6% post Great Recession. The UI only active records climb to a peak at the Great Recession and then stabilize at a somewhat lower level post Great Recession. In either case, these records present challenges due to the high likelihood that the SSN either does not consistently identify the same worker or identifies a set of workers. We plan to assess the impact of excluding these records in a later version of this paper, but our previous research in Abowd et al. 2018 shows that the overall earnings distribution is noticeably affected by included these "workers" in our analysis.

In order to study earnings volatility, we are primarily concerned with the change in earnings over time. To facilitate this, we transform the annual dataset (1995-2015) into a year-pair dataset (1996-2015) where each observation contains information from the current and previous years, and in some cases the current year and two years previous years. Table 3 shows the results of transforming the yearly dataset into a year-pair dataset with each observation indexed by the current year.[5] Compared with Table 2, the active category is expanded to include workers with earnings in the first year only, the second year only, or earnings in both

[5] During the period prior to 2004 when states are still entering the LEHD data, we do not include year pair earnings observations with dominant job (the job with the most earnings in a year) earnings until the year pair is fully observable. For example, although data for Ohio becomes available in 2000, data for the 2000 year pair (years 1999 and 2000) is not fully observable. The first fully observable year pair for a worker with dominant job earnings in Ohio begins in 2001 and any earlier year pairs where the dominant job in either year is in the state of Ohio are excluded from the analysis sample.

years while the inactive category and the never worked category are combined, creating the eligible, but not active, category. The year-pair dataset directly shows annual labor force participation dynamics, which are highlighted in Figure 3. We see patterns similar to the analogous categories in Figure 2 for workers active both years and inactive both years in the top panel of the two-year graph. The workers that are only active one of the two years represent workers exiting (active year 1 only) and entering the labor market (active year 2 only). From 2001 to 2010, the number of workers exiting is noticeably greater than the number entering, however, the number of workers inactive both years is relatively stable during this period. This result is probably because a relatively large number of “baby-boomers” turned 60 beginning in 2001 and aged out of our sample. The effect of the Great Recession can clearly be seen in the bottom panel beginning in 2007 and/or 2008 with a large increase in workers exiting active status and a large decrease in workers entering active status. At first there is a relatively strong recovery from the Great Recession (as shown by the large reduction in the gap between the active year 1 only and active year 2 only groups) during 2009 to 2011, but there is never a period from 2011 forward when the proportion of entrants is greater than exiters.

Our primary measure of earnings is based on annual UI job-level earnings reports. We adjust nominal earnings to real earnings using the Consumer Price Index (CPI-U), with 2000 as the base year. Let y_{ijt} be the real earnings for worker i employed at firm j in year t. Person-level annual earnings sum all jobs for each eligible male worker in each year:

$$e_{it} = \sum_j y_{ijt}$$

To examine earnings volatility, we create various measures of the change in annual earnings. Our primary earnings volatility measure is the change in log earnings from year $t-1$ to year t:

$$l_{it} = \ln(e_{it}) - \ln(e_{it-1})$$

The change in log earnings measure, l_{it}, is available from 1996 to 2015 for workers with positive earnings in both years. We also analyze two other earnings measures; the first is the arc percentage change, a_{it}, and the second is the two-year change in log earnings, p_{it}

$$a_{it} = \frac{(e_{it}-e_{it-1})}{(e_{it}+e_{it-1})/2} \text{ and}$$

$$p_{it} = \ln(e_{it}) - \ln(e_{it-2}).$$

The arc percentage change allows us to include workers with earnings in only one year and the two-year change allows us to produce results comparable to the PSID where earnings are only available every other year. The data for a_{it} are available from 1996 to 2015, while the data for p_{it} are available from 2006 to 2015.[6]

III. Results

We begin our discussion by examining the trends in the mean one-year change in earnings shown in both Table 4 and Figure 4. Although the variance of the one-year change in log earnings is our principal measure of earnings variability, the mean change is also important. For example, workers whose earnings are decreasing may be less able to smooth consumption over time when hit with a large negative earnings shock. Several different measures of earnings change are shown on the graph: the first is the difference in log earnings, l_{it}, for workers with earnings in both years; the second is the arc percent change, a_{it}, using workers with earnings in at least one of the two years; the third is the arc percent change, a_{it}, using workers with earnings in both years; the fourth is the difference in log earnings, l_{it},for workers with earnings in both years and for whom the difference in log earnings is not less than the 1st percentile nor greater than the 99th percentile of the overall distribution of the difference in log earnings; and the final measure is identical to the fourth measure, except that the percentile cutoffs are recalculated each year. The series are clustered into two groups, the first cluster includes only the "Arc Pct Change A1+" measure and the second includes all other series. Not surprisingly, trimming the change in earnings reduces the large negative earnings shocks that occurred due to the recessions in 2001-2003 and 2008-2010, but otherwise the series in the second cluster are very similar. The "Arc Pct Change A1+" measure includes workers moving into and out of active status and paints a somewhat different picture of post-2000 earnings growth. Including workers with relatively long spells of inactivity results in a relatively long period between 2001 and 2010 where earnings are either declining or not increasing (except for a small increase in 2006), with a particularly severe reduction in earnings in 2009.

[6] Although the data for p_{it} could be produced starting in 1997, producing statistics is complicated by state entry, due to time constraints this paper only shows results for the complete data period.

Figure 5 is similar to Figure 4, except that in this figure we calculate the mean two-year change in earnings, p_{it}, (both trimmed using the overall P1 and P99 values and not-trimmed) and compare it with the one-year change in earnings, l_{it}. The mean change in earnings for the two-year measures are somewhat larger, as would be expected due to the longer time interval between earnings measures, but overall the trends are similar.

In Table 5 we show the variance of the change in earnings for each of the measures first discussed in Table 4. As in Table 4, the variance series, shown in Figure 6, can be put into two clusters, however the composition of the clusters differ. In the first cluster we have the “Diff Log Earn” and the “Arc Pct Change A1+” series. These series show the largest earnings variability as well as the largest changes in earnings variability over time. Both series follow a similar trend, with earnings variability declining until 2001, increasing somewhat until 2003, declining leading into the Great Recession, substantially increasing during the Great Recession, and then declining consistently from the Great Recession peak in 2010. The relatively large earnings variability found in the first cluster is due to not excluding extremely large changes in earnings (which are squared when calculating the variance). In the “Diff Log Earn” series a worker must be employed both years, however, a worker who is not active for most of the year but starts a new job at the end of the previous year and is then employed the entire current year would have an extremely large change in earnings solely due to the start date of the job. Similarly, a worker that is inactive the previous year, but starts a new job in the current year, will also have a similarly large change in earnings. The “Arc Pct Change A1+” measure captures both of these cases, while the “Diff Log Earn” measure captures only the first example. While both series are a useful point of reference, the “Arc Pct Change A1+” measure treats workers not continuously employed both years consistently.

The second cluster of series (Arc Pct Change A2, Diff Log Earn Trim, and the Diff Log Earn Trim Yr) in Figure 6 are largely interchangeable (except perhaps in 2009). These series follow a similar, but attenuated, trend as the series in the first cluster, with substantially lower overall earnings variability. We included the Arc Pct Change A2 measure to show how similar this measure of variability is to the trimmed Diff Log Earn measure. Except for very large changes in earnings, both measures produce very similar results.

The “Arc Pct Change A1+” and the “Diff Log Earn Trim” are useful companion measures when interpreting changes in the earnings variability over time. For example, prior to 2001 and once again during the Great Recession, worker entry and exit play a relatively large role in earnings variability compared to the period between 2001 and 2008 when the relationship between the two series was relatively stable. Workers who maintain relatively stable employment (at least some earnings in both years) face substantially reduced earnings variability compared with workers who do not.

Figure 7 resembles Figure 6, except that we calculate the variance of the two-year change in earnings, p_{it}, (both trimmed using the overall P1 and P99 values and not-trimmed) and compare it with the one-year change in earnings, l_{it}. The variance of the change in earnings for the two-year measures is somewhat larger, as would be expected due to the longer time interval between earnings measures, but overall the trends are similar.

In Table 6, we show selected percentiles (P5, P10, P25, P50, P75, and P95) of the distribution of the change in log earnings. Focusing first on the middle of the distribution, workers with earnings changes between P25 and P75, the median earnings change is about 2% a year on average with the P25 value typically around -9% and the P75 values around +15%. Real earnings changes around the median are not symmetric, with increases in earnings typically larger than the decreases in earnings, except during the Great Recession years of 2008 and 2009 when the P25 earnings decrease is larger in absolute value than the P75 increase.

As we move outside the central part of the earnings change distribution, the year to year change in earnings is substantially more variable over time. The P5, P10, P90, and P95 series all show more change in earnings variability over time then the central part of the earnings change distribution. During recessions, large increases in earnings decrease, while large decreases in earnings increase, with the increase in the negative earnings shocks outweighing the decrease in the positive earnings shocks resulting in increasing overall earnings variability. We can see this more clearly in Figure 9 by examining the percentile ranges (P75-P25, P90-P10, and P95-P5) for the tails of the earnings change distribution. The trimmed change in log earnings (Var Diff Log Earn Trim) series and the interquartile range (P75-P25) are very similar both in level and in their relative stability over time. As we move outside the central part of the change distribution earnings variability is noticeably less stable over time, with the P90-P5 range substantially more variable than the P75-P25 range.

So far, we have shown earnings variability for prime age males to be generally declining over time, except for periodic increases that occur during Great Recessions. One possible explanation for the change in earnings variability might be a shift in the age distribution of the active male population along with an assumption that different parts of the age distribution have more or less earnings variability (perhaps by assuming, for example, that older workers experience less wage variability than younger workers). To determine if this hypothesis is consistent with the data, we estimate the following regression model:

$$l_{it} = \beta_0 + \beta_1 age_{it} + \beta_2 age_{it}^2 + \epsilon_{it}$$

Another possible explanation is that earnings changes differ systematically across workers. For example, some workers may consistently experience more earnings variability than others and that a change in the composition of these workers is responsible for the change in earnings variability over time. To test this hypothesis, we replace the overall constant term β_0 with a fixed person effect β_i as shown in the equation below.[7]

$$l_{it} = \beta_i + \beta_1 age_{it} + \beta_2 age_{it}^2 + \epsilon_{it}$$

The results of the model estimation are summarized in Table 7 where we compare the residuals from the earnings change regression models with the dependent variable l_{it}. If the models have a sizeable level of explanatory power, the residuals should differ substantially from the change in log earnings. In Figure 10, we plot the series shown in Table 7. The residuals from the regression model using age and age-squared are virtually identical to the "Diff Log Earn Trim" series. This result strongly rejects the hypothesis that changes in the age distribution of the male workforce explains the trend in the change in log earnings.

In contrast to the results of the age model, the results including a fixed person effect does show some potential explanatory power. Certain workers do appear to consistently have more earnings variability than other workers, especially after the Great Recession (although this may simply be a new worker effect that would disappear with a longer time series). Once we

[7] All models are estimated using the trimmed (P1 and P99) version of the one-year difference in log earnings.

double trim, removing the top and bottom 1% of earnings change residuals, the trends are about the same, but the overall level of earnings variability is substantially less.

The distribution of the log earnings change (Figures 8 and 9) and the residuals from the second earnings change model (Figure 10) suggest at least two important sub-populations. The first group, representing at least the middle 50% of the workers every year, experience a relatively consistent moderate amount of earnings variability, with only minor disruptions during recessionary periods. The second group at the bottom and top of the change in earnings distribution, experience a large amount of year-to-year earnings variability, with a declining trend and large shocks during recessions. The results from the earnings change regression models show that many of the large changes in annual earnings are clustered within certain workers, however the model including a fixed person effect explains only about 15% of the overall (across worker and year) earnings change variation. While person heterogeneity is important (especially in the tails), the vast majority of prime age male workers appear to have a similar probability of a substantial yearly change in log earnings.

To better understand what type of workers or events are associated with large changes in annual earnings we use information about the earnings level and work history to create various yearly sub-populations of prime age male workers. Each sub-population represents a fraction of the overall population and has its own mean and variance. The variance decomposition equation below (for the example of two sub-populations), shows how each of three statistics for each sub-population (fraction of the total, mean, and variance) are combined to calculate the total variance.

$$Var(x) = p * var(x_1) + (1 - p) * var(x_2) + p * (1 - p) * (\mu(x_1) - \mu(x_2))^2$$

When the means across the two sub-populations are equal, $\mu(x_1) = \mu(x_2)$, the total variance is the weighted sum of the sub-population variances. For almost all of the sub-populations we analyze, the means are similar and the last term can safely be ignored. As a check we always compare the yearly total variance calculated using the entire sample with our estimate of the total variance calculated using the weighted sum of the sub-population variances. Any noticeable deviation is evidence that the mean differences between sub-populations are an important part of the total variance and will be noted in our discussion of the results.

The first set of sub-populations (see Table 8) we analyze are based on the overall distribution of maximum real earnings observed for each yearly earnings pair in our trimmed (P1 to P99) analysis sample (1996-2015).[8] Every year, we place each annual maximum earnings value greater than $1,774 (P1) and less than $292,200 (P99) in real 2000 dollars into one of three earnings bins: P1 to P25 – earnings less than or equal to $22,600; P25 to P75 – earnings greater than $22,600 and earnings less than or equal to $58,590; and P75 to P99 – earnings greater than $58,590.

Figure 11 shows the proportion of workers each year in each of the three earnings bins. The composition of workers in each of the three bins changes over time, with the fraction of workers in the middle of the distribution shrinking and those at the bottom and the top growing. Most of the growth in the top of the distribution can be seen to have occurred prior to the Great Recession, while most of the growth at the bottom of the distribution occurred after the Great Recession.

Figure 12 shows the mean change in log earnings for the workers in each max earnings bin. The three series generally track together over time, but up until 2010 or 2011 workers at the top of the earnings distribution rarely faced an earnings decrease and their increases are on average almost always larger than the other earnings groups. Workers consistently in the bottom of the earnings distribution fare especially poorly with a long stretch of decreases in real earnings from 2001 to 2010. These results are consistent with the increasing inequality found between 2000 and 2011 in previous research using LEHD data (Abowd et al 2018).

Figures 13 and Figure 14 show the variance of each earnings category every year, in the first figure we report the unweighted variance, while in the second figure the variance share of the total is computed by multiplying the variance of each sub-population by that sub-population's fraction of the total population and then divide the result by the variance of the overall population.[9] The variance is significantly higher for workers in the bottom quartile of

[8] For example, in 2005 a worker in our yearly earnings pair sample may have earnings in either 2004, 2005, or both years (workers with no earnings in both years are excluded). The maximum of the two non-zero earnings values is our computed maximum earnings value for 2005.

[9] The sum of each share of the total variance may not sum to one due to the omitted "difference in means" variance component, however for our results this "residual" is almost always negligible (any significant deviation will be noted in the text).

the earnings distribution, while the variances of the two other categories are clustered near the bottom of the chart. The variance of workers at the top has generally been declining since 2000 with a small blip during the Great Recession, while the variance for workers in the middle and top stayed relatively constant (excluding increases during recessions) until a noticeable decrease began starting in 2010. The share of total variance results look similar for the middle and top earnings categories, however the results for the bottom look noticeably different. For example, although the variance is generally declining for workers in the bottom after 2010, the share of total variance attributable to workers in the bottom earnings quartile is actually increasing due to the increasing share of workers in the bottom quartile during that time period. This example shows how important it is to look at both the share and the variance of each sub-population. Although the earnings change variance of workers in the bottom earnings quartile was declining, the number of workers in the bottom quartile was increasing, reducing the impact on the total variance of the decline in the variance for workers at the bottom of the earnings quartile.

Given that over half of the total earnings variance is explained by workers in the bottom earnings quartile and that a large determinant of earnings for these workers is whether or not they were employed the entire year and/or had a job change it is worthwhile comparing full year and/or same dominant job workers with everyone else. Table 8 shows the number of observations, mean and variance for workers employed for eight consecutive quarters (every quarter of the two years in each earnings pair observation) and everyone else. Table 9 shows the number of observations, mean, and variance for workers with the same dominant job employer in each of the two years in each earnings pair observation and everyone else. Table 10 shows the results for the interaction of the two categories shown separately in Tables 8 and 9; workers employed every quarter with the same dominant job employer in both years and everyone else.

The modal category for prime-age male workers is employed eight consecutive quarters with the same dominant in both years. As shown in Table 11 and Figure 23, most workers are in a stable multi-year employment relationship and the share of the prime age male workforce attributable to this type of worker is increasing over time, from about 65% of male workers in

1996 to almost 73% in 2015. If we examine the workers employed the entire year (Figure 15) and compare them with the dominant job workers (Figure 19), the trends are similar. Both groups show increases over the period, although the increase is largest for the continuous-active compared to the same-dominant-job workers. The percent employed four quarters in both years increases from about 75% in 1996 to almost 82% in 2015, while workers whose dominant job (employer with the highest earnings) is the same in both years increases from about 81% in 1996 to 84% in 2015.

Considering the changes in mean earnings (Figure 16), the picture looks especially bleak for workers not active every quarter during the eight-quarter window covered by each year-pair observation. Similar to workers in the bottom earnings quartile, earnings growth is never positive from 2001 to 2010 for workers not active every quarter, with consistently less (or in two years, 1998, and 2011, the same) earnings growth compared to workers who are active every quarter during the entire sampled period. The composition of workers in each group is changing over time as more workers shift to consistently working at least some part of every quarter, but it is unclear from the aggregate statistics whether these changes are also associated with a change in the structure of wages for workers active only part of the both years compared with those active the entire period or whether this is purely a labor supply effect.

In contrast to the labor force attachment results shown in Figure 16, workers who change dominant employers during the two-year window have higher earnings growth when labor market conditions are good and negative earnings growth during recessionary periods (Figure 20). The earnings growth for dominant-job changers post Great Recession was exceptionally large and noticeably higher than for dominant-job stayers (almost 7% for changers vs. less than 1% for stayers). However, during the Great Recession job changers faced significant negative earnings growth. For example, in 2009 workers with a dominant employer change between 2008 and 2009, faced a mean earnings decline of over 16%, while earnings for workers that did not change dominant employers declined by slightly less than 5%. During the great recession dominant employer change was much less likely to be voluntary and likely associated with a substantial period of no or low employment, exacerbating the earnings decrease.

The earnings variability results (Figures 17 and 21) for workers not active eight consecutive quarters compared to workers with a dominant job change show increases in earnings variability beginning in 2001 and continuing through 2011. After 2011, the increases in earnings variability either slow down (not active eight quarters) or steeply decline (job changers), however in both cases the earnings variability for these workers is substantially higher than for more stable workers in every year of our sample. Putting the two sets of results together, we compare the most stable workers—those employed all eight quarters who do not change dominant job—with all other workers (Figures 23-25). We see a similar pattern. Earnings variability increases from 2000 to 2011 for the least stable workers and declines moderately for these same workers during the recovery from the Great Recession. Earnings variability for the stable workers is consistently only a fraction of the level for the less stable workers (for example, in 2014, the earnings variance for the most stable workers is 0.0460 compared with 0.7906 for less stable workers).

These results clearly show large differences in earnings growth and variability between the stable workers (active eight consecutive quarters with no dominant job change) and everyone else and would be expected (ceteris paribus) to result in an increase in overall earnings variability during the sample period. However, changes in the earnings variability for each group are largely mitigated by an offsetting positive shift in the proportion of workers in a more stable employment arrangement (work eight consecutive quarters and/or dominant job the same), resulting in a relatively stable variance contribution (Figures 18, 22, and 26) from each of the less stable worker sub-populations over almost the entire sample period.

In summary, although overall earnings variability is slowly declining, our results show substantially different trends in earnings variability for various sub-populations. This result points to the importance of sample composition when estimating overall earnings inequality. If the composition of the analysis sample shifts, for example favoring more stable workers at various point in time, then the resulting trends in the overall sample earnings variability could easily shift up or down.

Table 12 shows the complete autocorrelation matrix for the change in log earnings, trimmed at P1 and P99 overall. The top panel displays the autocorrelations by year and the bottom panel by lag length. As Abowd and Card (1989) found, the change in log earnings is well-modeled by a nonstationary second-order moving average. Since there is no classical measurement error in these data, the negative first-order autocorrelation can be interpreted as a property of the transitory earnings process. The second-order autocorrelation is also negative and one-third the magnitude. The third and fourth-order autocorrelations, also negative, are very small in magnitude although statistically different from zero. The second panel shows that the magnitude of the first and second-order autocorrelations are sensitive to the business cycle—increasing during and just after a recession.

IV. Conclusion

The Great Recession reduced earnings primarily through long spells of inactivity (non-employment) even for prime-age males. Indeed, prime-age males who did not change employers and worked continuously experienced earnings growth every year except 2008 and 2009, but even during the Great Recession, their cumulative earnings loss was less than -0.004 log points. By contrast all other prime-age male workers (about 30% of the eligible population) had a cumulative loss over the same period of -0.288 log points. Those with stable employment experienced very little change in volatility as well. Overall volatility for prime-age males not stably employed was about 15 times as large as for the stably employed, spiked during the Great Recession, and remained elevated thereafter.

References

Abowd John M. and David Card. 1989. On the Covariance Structure of Earnings and Hours Changes. *Econometrica* 57(2):411–45.

Abowd, John M., Kevin L. McKinney, and Nellie L. Zhao. 2018. Earnings Inequality and Mobility Trends in The United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data. *Journal of Labor Economics* 36(S1):183-300.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2009. The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators. In *Producer dynamics: New evidence from micro data*, ed. Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 149–230. Chicago: University of Chicago Press. http://www.nber.org/chapters/c0485.

Arellano, Manuel, Richard Blundell and Stéphane Bonhomme. 2017. Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. *Econometrica* 83(3):693-734

Brown, J. David, Julie L. Hotchkiss and Myriam Quispe-Agnoli. 2013. Does employing undocumented workers give firms a competitive advantage? *Journal of Regional Science* 53:158–70.

Gottschalk, Peter and Robert Moffitt. 1994. The Growth of Earnings Instability in the U.S. Labor Market. *Brookings Papers on Economic Activity* (2):217-272.

Imbens, Guido W. and Donald B. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences*. New York: Cambridge University Press.

Jensen, Shane T. and Stephen H. Shore. 2015. Changes in the Distribution of Earnings Volatility. *Journal of Human Resources* 50(3):811-836

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Table 1 - Data Availability by Source (State UI, DC UI, OPM)

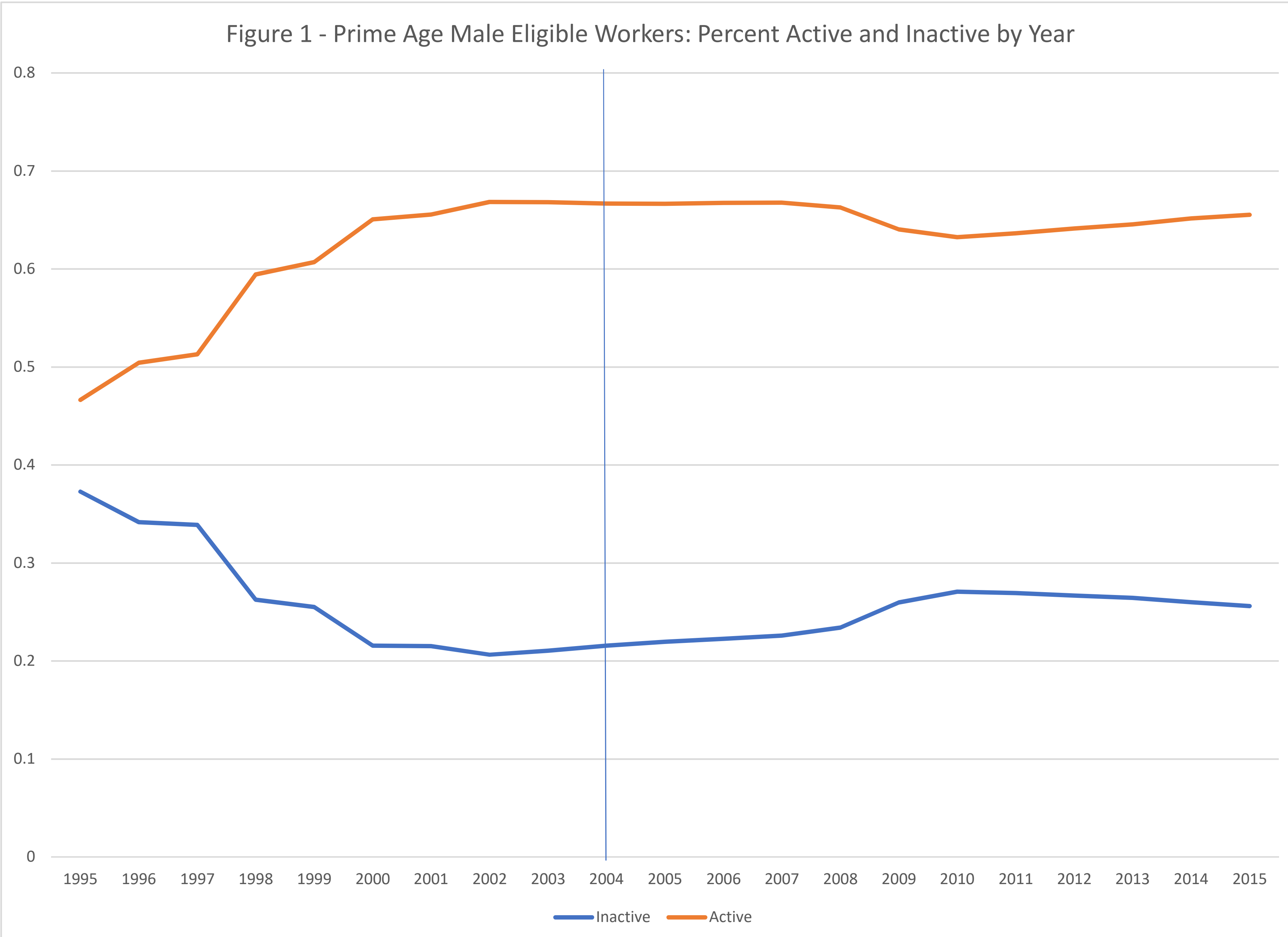
Count	State	First YYYY:Q Available	Last YYYY:Q Available	Pct 2012.1 QCEW Emp
1	Maryland	1985:2	2015:4	1.83%
2	Alaska	1990:1	2015:4	0.22%
3	Colorado	1990:1	2015:4	1.70%
4	Idaho	1990:1	2015:4	0.45%
5	Illinois	1990:1	2015:4	4.38%
6	Indiana	1990:1	2015:4	2.19%
7	Kansas	1990:1	2015:4	0.98%
8	Louisiana	1990:1	2015:4	1.41%
9	Missouri	1990:1	2015:4	1.99%
10	Washington	1990:1	2015:4	2.12%
11	Wisconsin	1990:1	2015:4	2.08%
12	North Carolina	1991:1	2015:4	2.92%
13	Oregon	1991:1	2015:4	1.23%
14	Pennsylvania	1991:1	2015:4	4.44%
15	California	1991:3	2015:4	11.37%
16	Arizona	1992:1	2015:4	1.85%
17	Wyoming	1992:1	2015:4	0.19%
18	Florida	1992:4	2015:4	5.78%
19	Montana	1993:1	2015:4	0.31%
20	Georgia	1994:1	2015:4	2.90%
21	South Dakota	1994:1	2015:4	0.30%
22	Minnesota	1994:3	2015:4	2.05%
23	New York	1995:1	2015:4	6.49%
24	Rhode Island	1995:1	2015:4	0.35%
25	Texas	1995:1	2015:4	8.10%
26	New Mexico	1995:3	2015:4	0.55%
27	Hawaii	1995:4	2015:4	0.44%
28	Connecticut	1996:1	2015:4	1.26%
29	Maine	1996:1	2015:4	0.43%
30	New Jersey	1996:1	2015:4	2.87%
31	Kentucky	1996:4	2015:4	1.32%
32	West Virginia	1997:1	2015:4	0.52%
33	Michigan	1998:1	2015:4	3.04%
34	Nevada	1998:1	2015:4	0.89%
35	North Dakota	1998:1	2015:4	0.31%
36	South Carolina	1998:1	2015:4	1.35%
37	Tennessee	1998:1	2015:4	2.03%
38	Virginia	1998:1	2015:4	2.65%
39	Delaware	1998:3	2015:4	0.31%
40	Iowa	1998:4	2015:4	1.12%
41	Nebraska	1999:1	2015:4	0.69%
42	Utah	1999:1	2015:4	0.91%
43	Ohio	2000:1	2015:4	3.93%
44	Oklahoma	2000:1	2015:4	1.11%
45	Vermont	2000:1	2015:4	0.22%
46	OPM	2000:1	2015:4	4.00%
47	Alabama	2001:1	2015:4	1.34%
48	Massachusetts	2002:1	2015:4	2.55%
49	District of Columbia	2002:2	2015:4	0.43%
50	Arkansas	2002:3	2015:4	0.86%
51	New Hampshire	2003:1	2015:4	0.47%
52	Mississippi	2003:3	2015:4	0.77%

Notes: Each row represents a state, DC, or the federal government (OPM). States are ordered by the quarter their data first became available in the LEHD infrastrucuture files. The last column shows the proportion of each state as a percentage of national 2012 month 1 QCEW employment. States above the horizontal line below row 25 are in the analysis sample the entire period (1995-2015), states below the line enter the sample in the first full year available. The sample is complete in 2004.

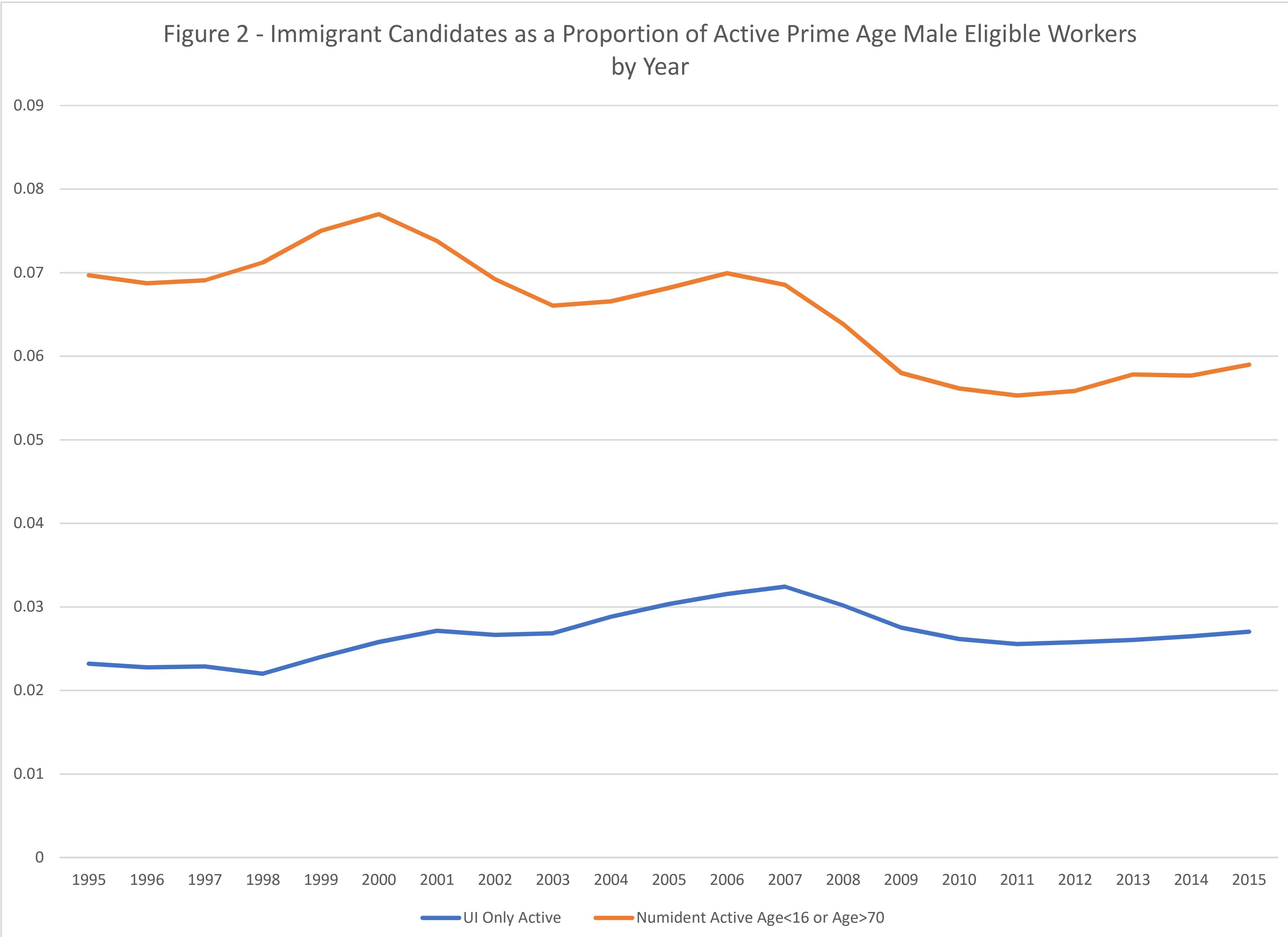
Table 2 - Analysis Sample Composition and (not used) Immigrant Candidates by Year

	Prime Age Male Eligible Workers				Immigrant Candidates	
Year	Inactive	Active	Never Worked	Total	UI Only Active	Numident Active Age<16 or Age>70
1995	26,260,000	32,850,000	11,320,000	70,430,000	762,000	2,289,000
1996	24,440,000	36,080,000	11,020,000	71,540,000	821,000	2,480,000
1997	24,590,000	37,210,000	10,740,000	72,540,000	851,000	2,571,000
1998	19,280,000	43,630,000	10,480,000	73,390,000	960,000	3,107,000
1999	18,920,000	45,010,000	10,230,000	74,160,000	1,080,000	3,376,000
2000	16,160,000	48,730,000	9,999,000	74,889,000	1,257,000	3,752,000
2001	16,260,000	49,540,000	9,762,000	75,562,000	1,345,000	3,656,000
2002	15,700,000	50,830,000	9,517,000	76,047,000	1,354,000	3,518,000
2003	16,140,000	51,210,000	9,289,000	76,639,000	1,374,000	3,382,000
2004	16,660,000	51,490,000	9,065,000	77,215,000	1,483,000	3,427,000
2005	17,050,000	51,720,000	8,818,000	77,588,000	1,569,000	3,526,000
2006	17,320,000	51,940,000	8,550,000	77,810,000	1,638,000	3,632,000
2007	17,660,000	52,160,000	8,305,000	78,125,000	1,691,000	3,575,000
2008	18,370,000	51,980,000	8,076,000	78,426,000	1,569,000	3,320,000
2009	20,430,000	50,370,000	7,845,000	78,645,000	1,387,000	2,921,000
2010	21,340,000	49,840,000	7,635,000	78,815,000	1,303,000	2,798,000
2011	21,270,000	50,260,000	7,436,000	78,966,000	1,285,000	2,779,000
2012	21,110,000	50,730,000	7,273,000	79,113,000	1,308,000	2,833,000
2013	20,960,000	51,160,000	7,127,000	79,247,000	1,333,000	2,957,000
2014	20,650,000	51,740,000	7,022,000	79,412,000	1,371,000	2,984,000
2015	20,410,000	52,240,000	7,059,000	79,709,000	1,412,000	3,082,000

Notes: Counts are rounded to 4 significant digits. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Inactive workers are eligible that year, but have no positive reported earnings. Active workers have positive earnings. Never worked are eligible, but never have positive reported earnings (1995-2015). The immigrant candidate columns show the two largest sources of earnings records excluded from the analysis. All states have entered sample by 2004.



Source: Table 2

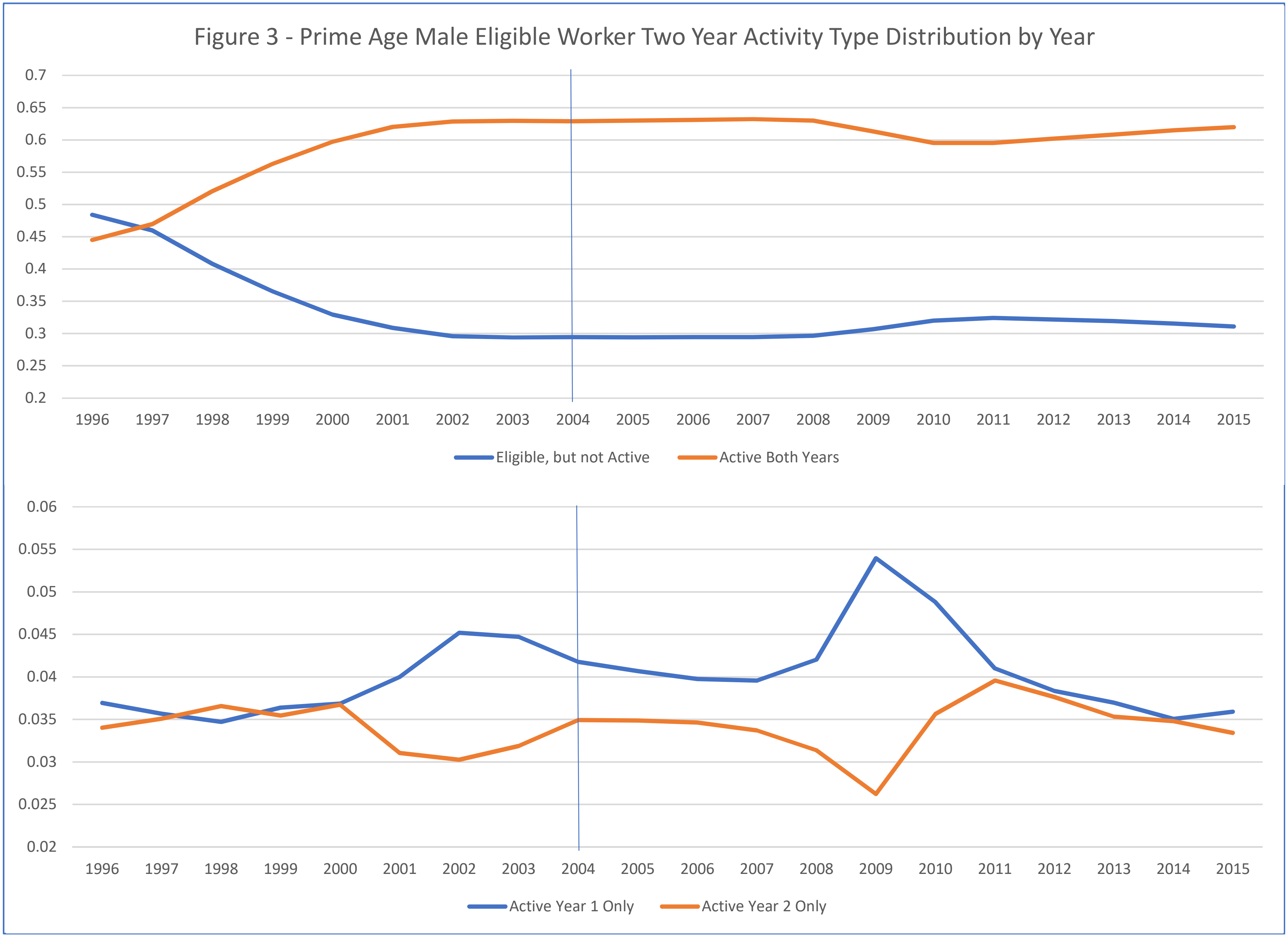


Source: Table 2

Table 3 - Two Year (Current and Previous) Analysis Sample Observations: Prime Age Male Eligible Worker Activity Type by Year

Year	Eligible, but not Active	Active Year 1 Only	Active Year 2 Only	Active Both Years	Total
1996	31,860,000	2,432,000	2,240,000	29,290,000	65,822,000
1997	31,750,000	2,464,000	2,424,000	32,430,000	69,068,000
1998	26,640,000	2,266,000	2,387,000	33,980,000	65,273,000
1999	25,760,000	2,566,000	2,501,000	39,680,000	70,507,000
2000	22,870,000	2,560,000	2,551,000	41,460,000	69,441,000
2001	22,350,000	2,894,000	2,248,000	44,890,000	72,382,000
2002	21,280,000	3,249,000	2,176,000	45,190,000	71,895,000
2003	21,430,000	3,261,000	2,325,000	45,910,000	72,926,000
2004	21,900,000	3,108,000	2,598,000	46,780,000	74,386,000
2005	22,100,000	3,056,000	2,620,000	47,330,000	75,106,000
2006	22,160,000	2,992,000	2,607,000	47,500,000	75,259,000
2007	22,250,000	2,990,000	2,548,000	47,770,000	75,558,000
2008	22,500,000	3,188,000	2,380,000	47,770,000	75,838,000
2009	23,340,000	4,107,000	1,997,000	46,670,000	76,114,000
2010	24,410,000	3,722,000	2,718,000	45,420,000	76,270,000
2011	24,740,000	3,131,000	3,022,000	45,470,000	76,363,000
2012	24,610,000	2,933,000	2,879,000	46,050,000	76,472,000
2013	24,430,000	2,829,000	2,703,000	46,570,000	76,532,000
2014	24,160,000	2,688,000	2,667,000	47,120,000	76,635,000
2015	23,860,000	2,757,000	2,567,000	47,610,000	76,794,000

Notes: Counts are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. For example, 1996 contains information for both the previous year (1995) and the current year (1996). Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead.

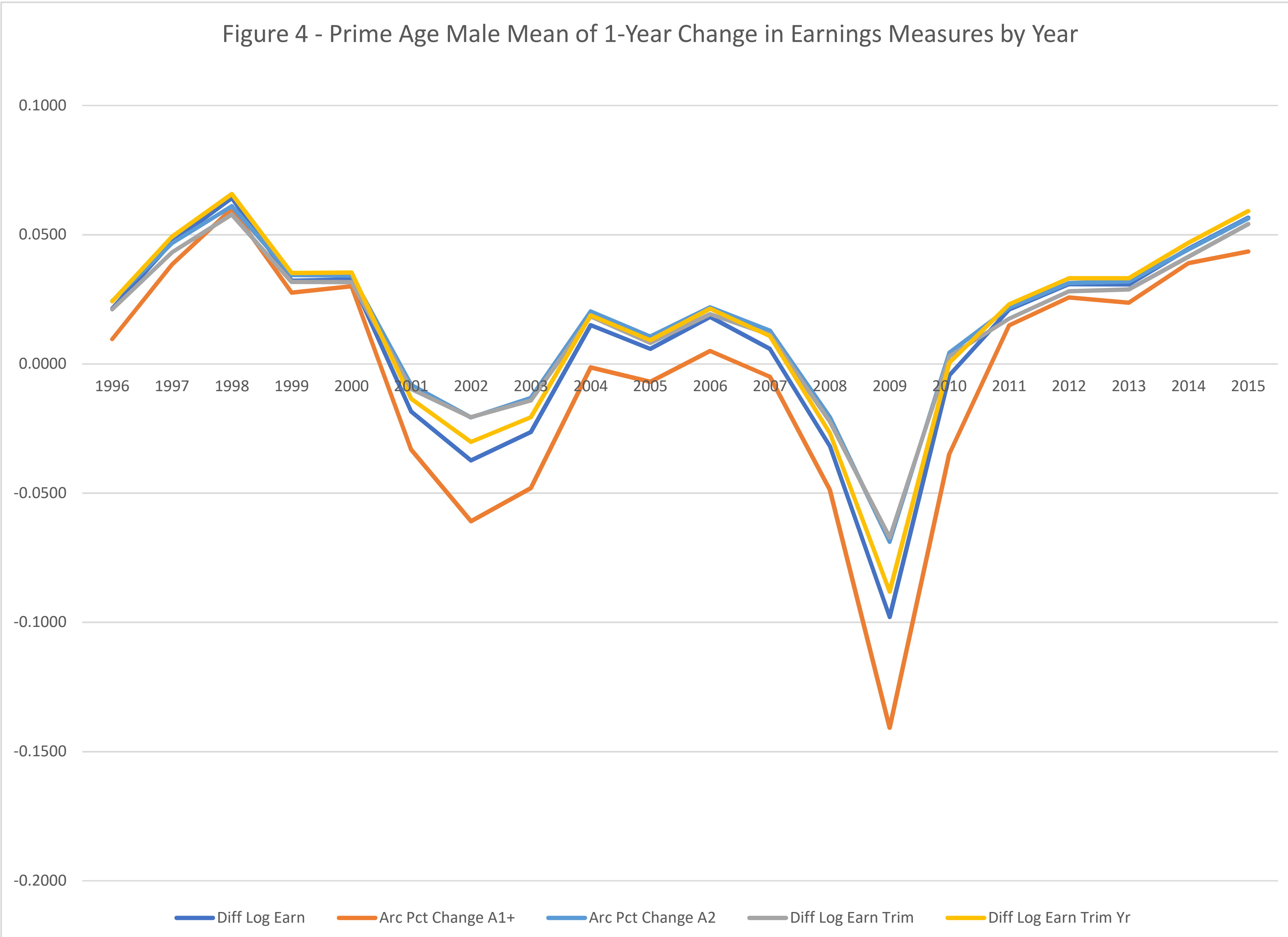


Source: Table 3

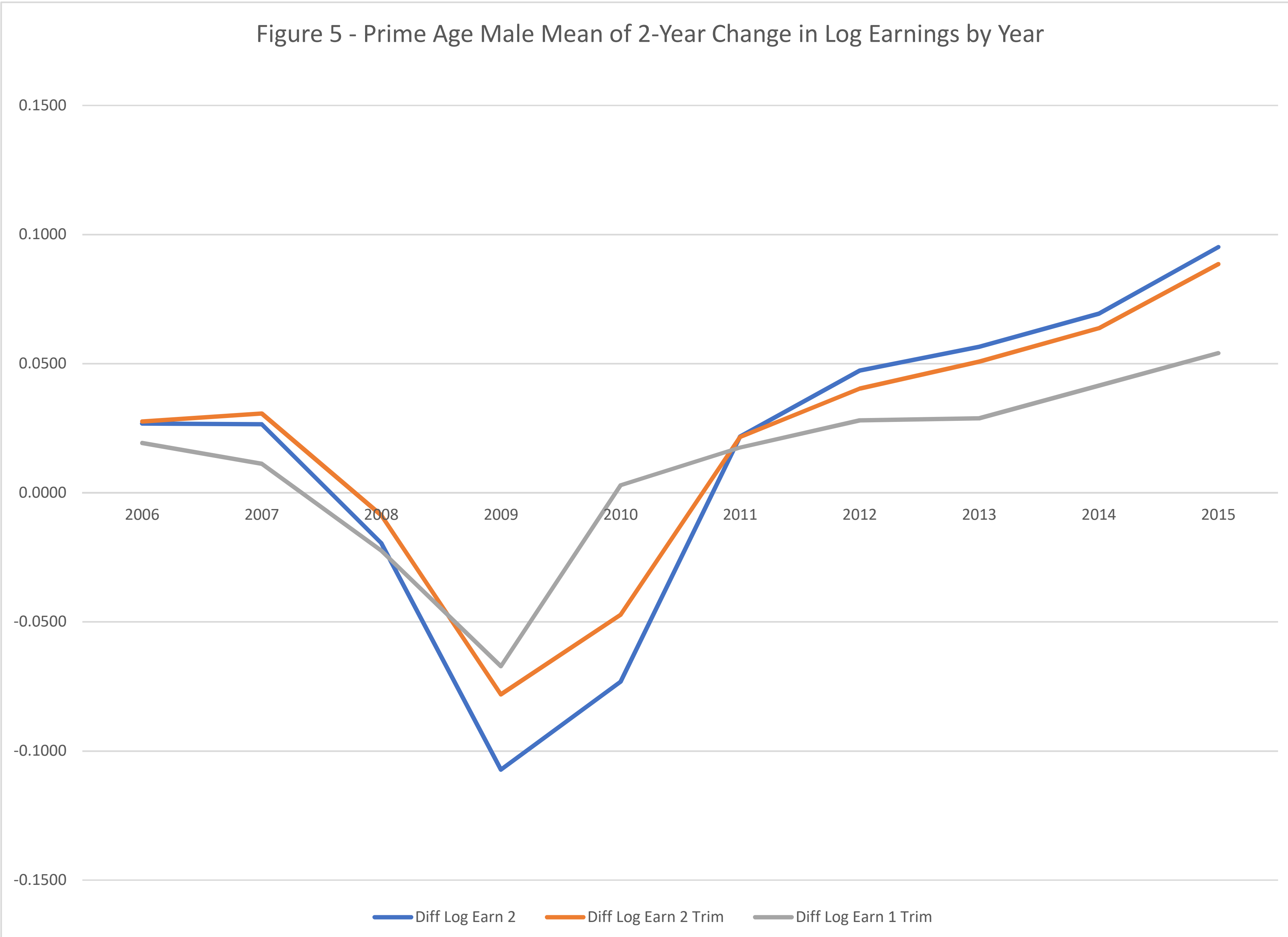
Table 4 - Prime Age Male Number of Observations and Means for Change in Earnings Measures

	Number of Observations							Mean						
Year	Diff Log Earn	Arc Pct Change A1+	Arc Pct Change A2	Diff Log Earn Trim	Diff Log Earn Trim Yr	Diff Log Earn 2	Diff Log Earn 2 Trim	Diff Log Earn	Arc Pct Change A1+	Arc Pct Change A2	Diff Log Earn Trim	Diff Log Earn Trim Yr	Diff Log Earn 2	Diff Log Earn 2 Trim
1996	29,290,000	33,960,000	29,290,000	28,640,000	28,700,000			0.0214	0.0096	0.0243	0.0211	0.0243		
1997	32,430,000	37,320,000	32,430,000	31,730,000	31,780,000			0.0471	0.0385	0.0467	0.0432	0.0493		
1998	33,980,000	38,630,000	33,980,000	33,290,000	33,300,000			0.0642	0.0600	0.0611	0.0578	0.0658		
1999	39,680,000	44,740,000	39,680,000	38,910,000	38,880,000			0.0322	0.0276	0.0344	0.0317	0.0352		
2000	41,460,000	46,570,000	41,460,000	40,650,000	40,630,000			0.0327	0.0301	0.0343	0.0317	0.0354		
2001	44,890,000	50,030,000	44,890,000	44,050,000	43,990,000			-0.0184	-0.0332	-0.0082	-0.0097	-0.0135		
2002	45,190,000	50,620,000	45,190,000	44,290,000	44,290,000			-0.0374	-0.0609	-0.0207	-0.0206	-0.0302		
2003	45,910,000	51,500,000	45,910,000	44,970,000	44,990,000			-0.0263	-0.0480	-0.0131	-0.0142	-0.0206		
2004	46,780,000	52,480,000	46,780,000	45,820,000	45,840,000			0.0150	-0.0013	0.0203	0.0184	0.0189		
2005	47,330,000	53,010,000	47,330,000	46,400,000	46,390,000			0.0058	-0.0070	0.0106	0.0082	0.0092		
2006	47,500,000	53,100,000	47,500,000	46,600,000	46,550,000	45,040,000	44,170,000	0.0182	0.0051	0.0219	0.0193	0.0214	0.0268	0.0276
2007	47,770,000	53,310,000	47,770,000	46,860,000	46,810,000	45,300,000	44,430,000	0.0058	-0.0050	0.0129	0.0113	0.0107	0.0265	0.0307
2008	47,770,000	53,340,000	47,770,000	46,840,000	46,820,000	45,360,000	44,480,000	-0.0317	-0.0487	-0.0205	-0.0223	-0.0267	-0.0195	-0.0087
2009	46,670,000	52,770,000	46,670,000	45,640,000	45,730,000	44,480,000	43,490,000	-0.0979	-0.1408	-0.0688	-0.0672	-0.0881	-0.1072	-0.0781
2010	45,420,000	51,860,000	45,420,000	44,410,000	44,510,000	43,460,000	42,580,000	-0.0045	-0.0349	0.0044	0.0029	0.0004	-0.0732	-0.0472
2011	45,470,000	51,620,000	45,470,000	44,500,000	44,560,000	42,870,000	41,960,000	0.0210	0.0149	0.0217	0.0175	0.0230	0.0218	0.0216
2012	46,050,000	51,860,000	46,050,000	45,120,000	45,130,000	43,130,000	42,220,000	0.0308	0.0257	0.0313	0.0281	0.0332	0.0474	0.0404
2013	46,570,000	52,110,000	46,570,000	45,680,000	45,640,000	43,760,000	42,900,000	0.0308	0.0237	0.0319	0.0288	0.0332	0.0566	0.0508
2014	47,120,000	52,470,000	47,120,000	46,240,000	46,180,000	44,420,000	43,570,000	0.0446	0.0390	0.0443	0.0415	0.0469	0.0693	0.0637
2015	47,610,000	52,930,000	47,610,000	46,760,000	46,660,000	44,910,000	44,070,000	0.0566	0.0435	0.0563	0.0541	0.0591	0.0952	0.0886

Notes: Counts and means are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. For example, 1996 contains information for both the previous year (1995) and the current year (1996). Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. The Diff Log Earn and the Arc Pct Change A2 columns include only workers with positive earnings in both years. The Arc Pct Change A1+ column includes worker active either in year 1 only, year 2 only, or both. Samples with a trim in the name exclude records less than P1 or greater than P99 of either the overall or by year change distribution. Diff Log Earn 2 and Diff Log Earn 2 Trim use two year previous earnings to calculate the earnings change measure.



Source: Table 4

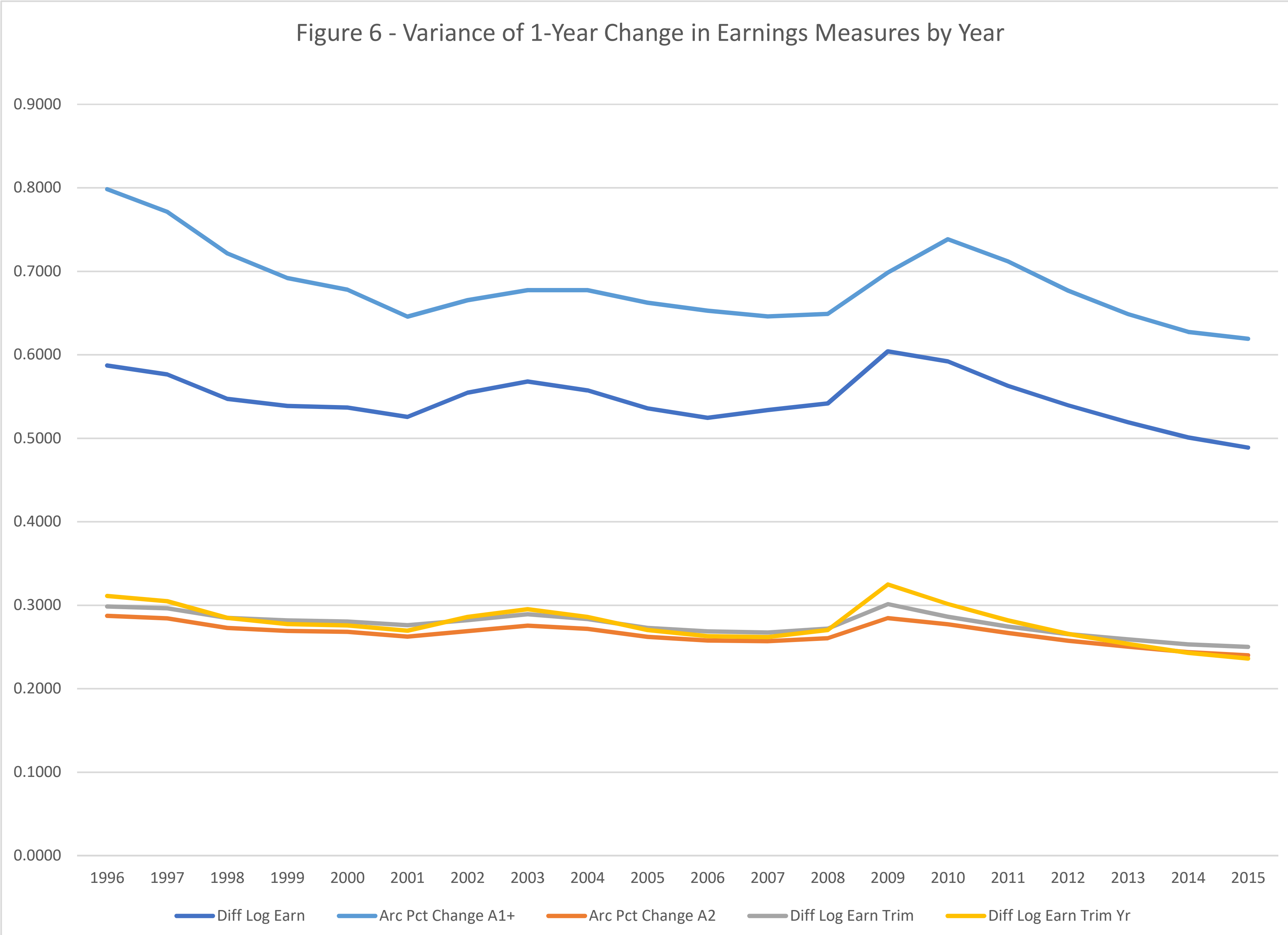


Source: Table 4

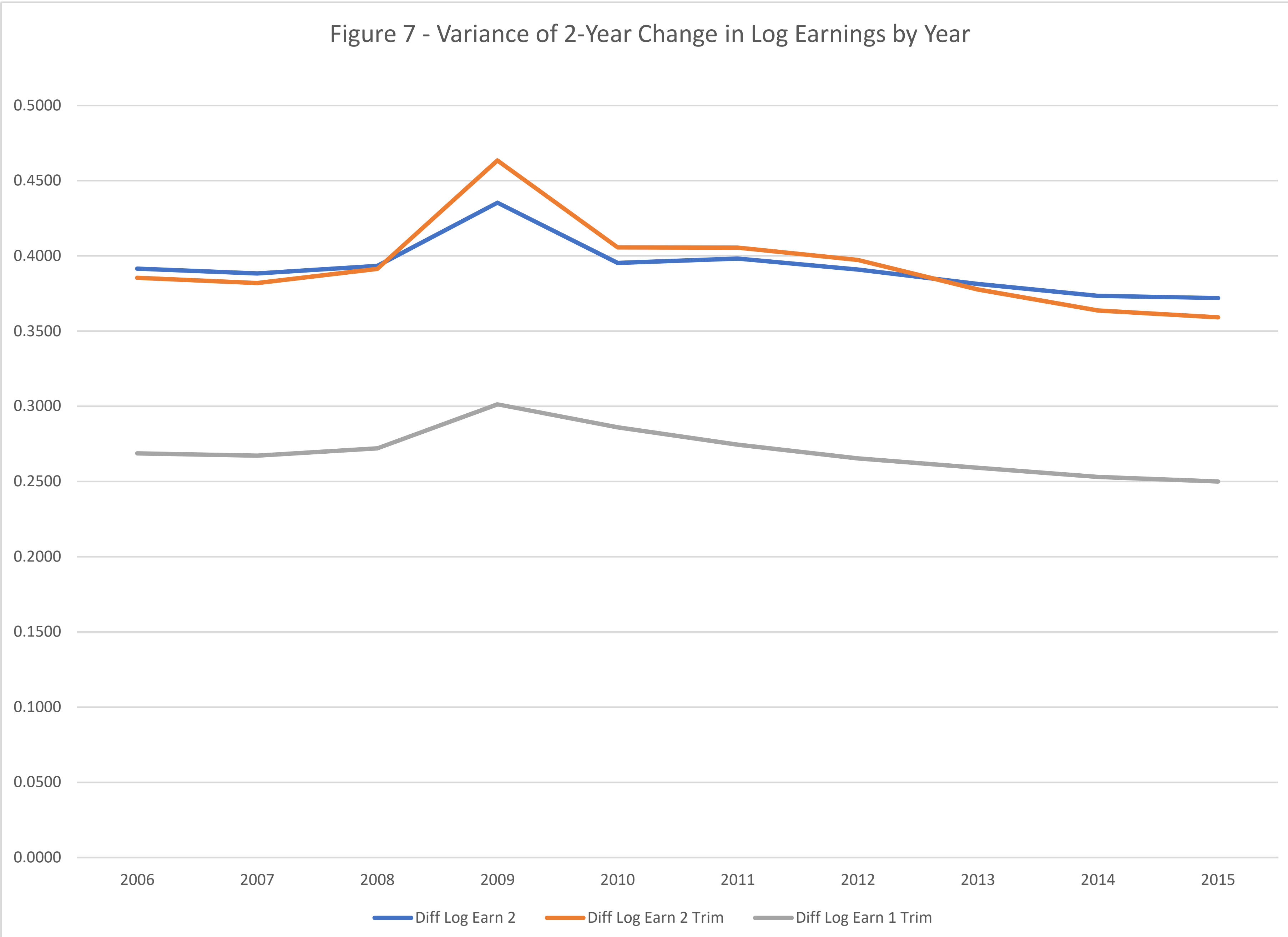
Table 5 - Prime Age Males Variance of Change in Earnings Measures

Year	Diff Log Earn	Arc Pct Change A1+	Arc Pct Change A2	Diff Log Earn Trim	Diff Log Earn Trim Yr	Diff Log Earn 2	Diff Log Earn 2 Trim
1996	0.5872	0.7984	0.2873	0.2986	0.3110		
1997	0.5764	0.7714	0.2844	0.2962	0.3048		
1998	0.5473	0.7215	0.2728	0.2847	0.2848		
1999	0.5387	0.6918	0.2691	0.2817	0.2774		
2000	0.5369	0.6779	0.2682	0.2804	0.2758		
2001	0.5255	0.6456	0.2625	0.2762	0.2695		
2002	0.5547	0.6655	0.2689	0.2820	0.2860		
2003	0.5679	0.6774	0.2755	0.2891	0.2951		
2004	0.5574	0.6775	0.2718	0.2835	0.2860		
2005	0.5357	0.6624	0.2621	0.2727	0.2703		
2006	0.5245	0.6528	0.2578	0.2688	0.2629	0.3915	0.3855
2007	0.5339	0.6459	0.2569	0.2672	0.2619	0.3883	0.3819
2008	0.5417	0.6490	0.2606	0.2721	0.2702	0.3933	0.3914
2009	0.6041	0.6986	0.2845	0.3013	0.3248	0.4354	0.4635
2010	0.5921	0.7384	0.2773	0.2861	0.3015	0.3954	0.4057
2011	0.5627	0.7119	0.2668	0.2745	0.2819	0.3982	0.4055
2012	0.5395	0.6770	0.2574	0.2654	0.2656	0.3910	0.3973
2013	0.5189	0.6488	0.2503	0.2591	0.2537	0.3814	0.3777
2014	0.5010	0.6274	0.2438	0.2531	0.2430	0.3734	0.3637
2015	0.4888	0.6192	0.2400	0.2500	0.2361	0.3720	0.3592

Notes: Variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. For example, 1996 contains information for both the previous year (1995) and the current year (1996). Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. The Diff Log Earn and the Arc Pct Change A2 columns include only workers with positive earnings in both years. The Arc Pct Change A1+ column includes worker active either in year 1 only, year 2 only, or both. Samples with a trim in the name exclude records less than P1 or greater than P99 of either the overall or by year change distribution. Diff Log Earn 2 and Diff Log Earn 2 Trim use two year previous earnings to calculate the earnings change measure. Sample sizes and means by year are shown in Table 4.



Source: Table 5

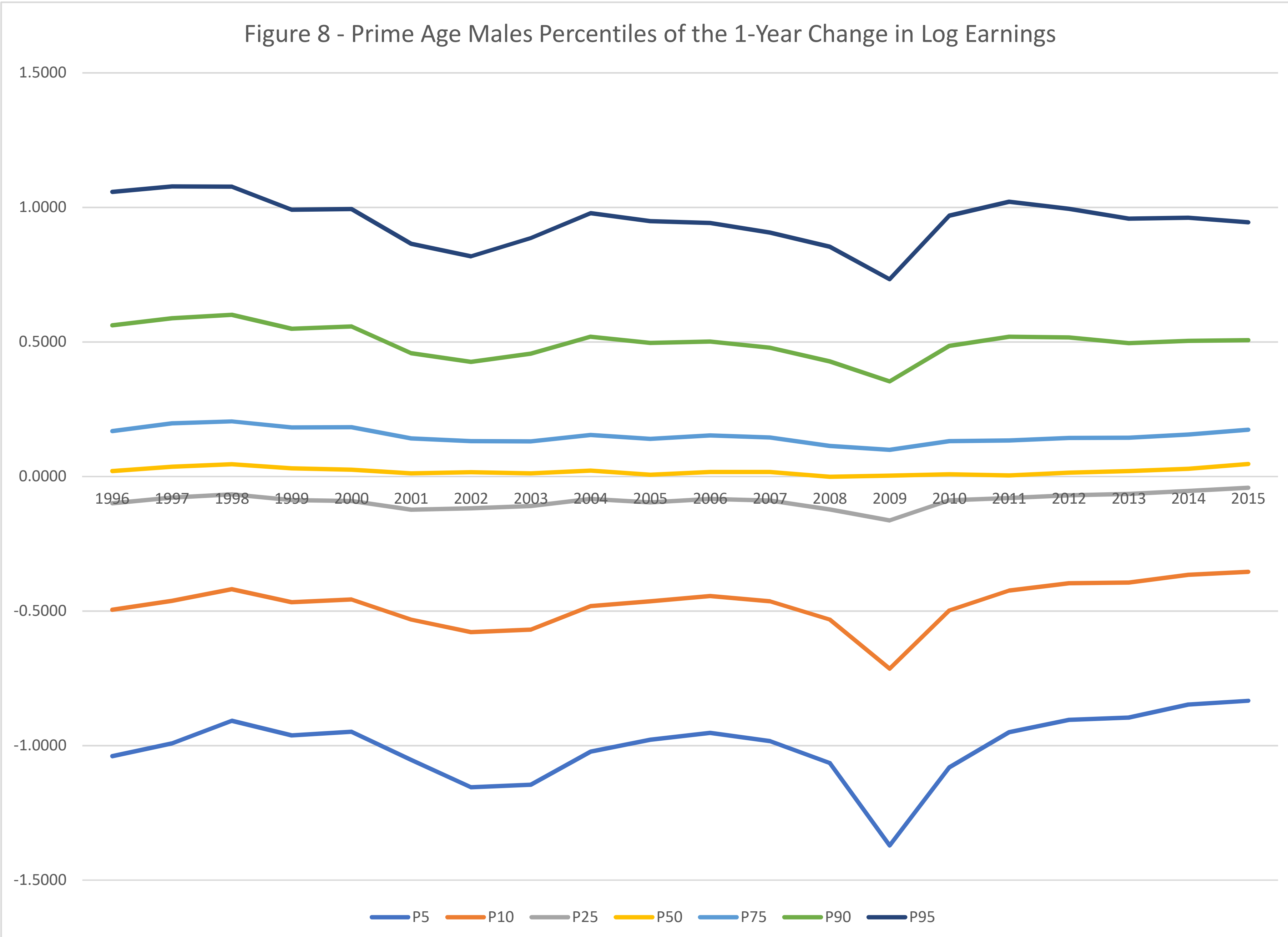


Source: Table 5

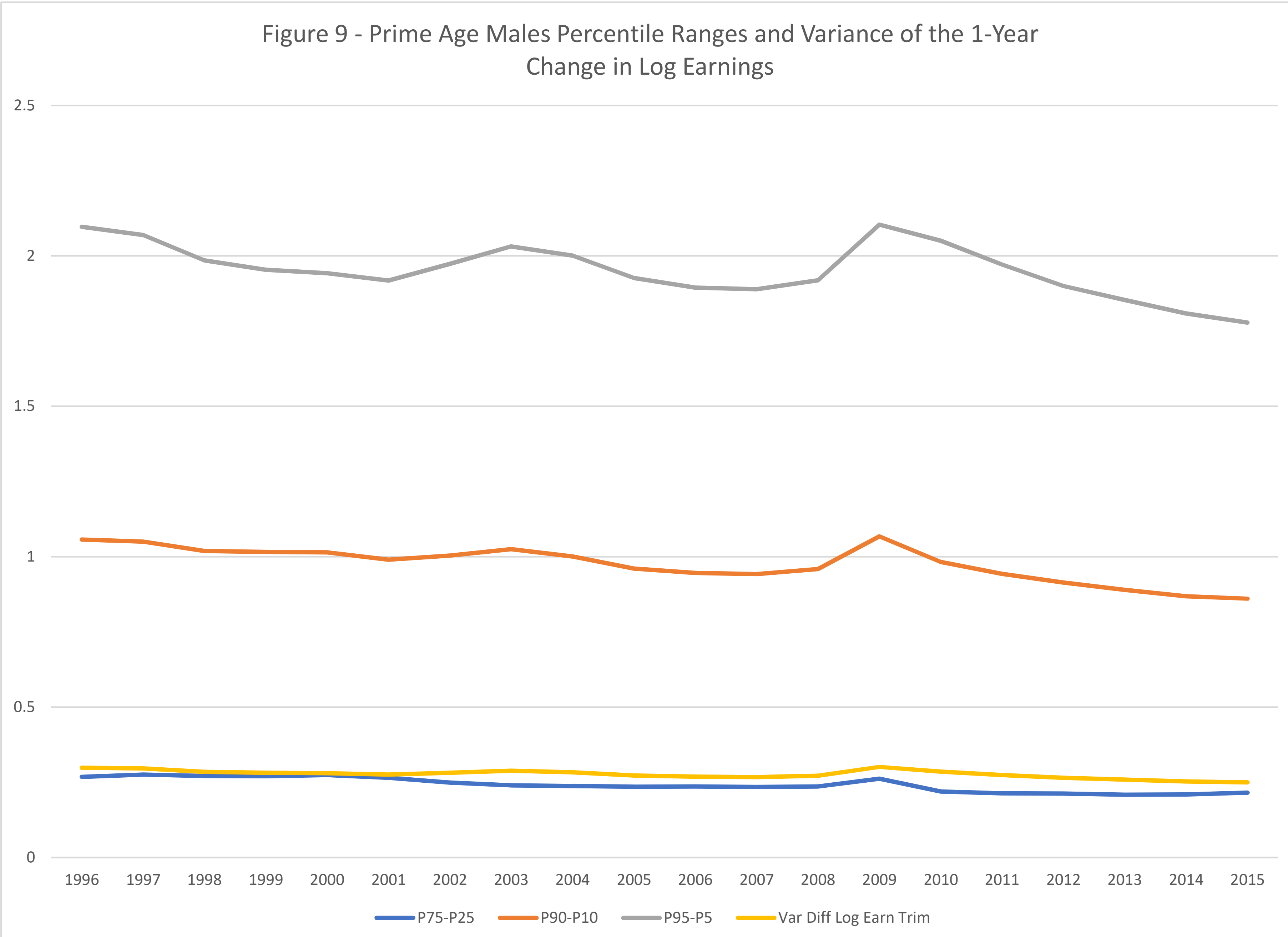
Table 6 - Prime Age Males Percentiles of the 1 Year Change in Log Earnings

Year	Number of Observations	P5	P10	P25	P50	P75	P90	P95
1996	29,290,000	-1.0390	-0.4954	-0.0997	0.0205	0.1687	0.5616	1.0580
1997	32,430,000	-0.9917	-0.4622	-0.0785	0.0363	0.1975	0.5881	1.0780
1998	33,980,000	-0.9079	-0.4183	-0.0666	0.0458	0.2049	0.6010	1.0770
1999	39,680,000	-0.9621	-0.4672	-0.0879	0.0301	0.1827	0.5493	0.9915
2000	41,460,000	-0.9486	-0.4572	-0.0913	0.0255	0.1829	0.5577	0.9939
2001	44,890,000	-1.0530	-0.5318	-0.1236	0.0116	0.1416	0.4584	0.8652
2002	45,190,000	-1.1550	-0.5781	-0.1182	0.0160	0.1312	0.4262	0.8187
2003	45,910,000	-1.1450	-0.5687	-0.1100	0.0120	0.1303	0.4565	0.8864
2004	46,780,000	-1.0220	-0.4814	-0.0837	0.0218	0.1541	0.5196	0.9792
2005	47,330,000	-0.9781	-0.4638	-0.0960	0.0065	0.1398	0.4969	0.9488
2006	47,500,000	-0.9525	-0.4442	-0.0836	0.0171	0.1525	0.5018	0.9419
2007	47,770,000	-0.9828	-0.4637	-0.0896	0.0173	0.1454	0.4785	0.9066
2008	47,770,000	-1.0650	-0.5314	-0.1222	-0.0012	0.1139	0.4278	0.8540
2009	46,670,000	-1.3710	-0.7143	-0.1632	0.0035	0.0994	0.3536	0.7331
2010	45,420,000	-1.0810	-0.4972	-0.0881	0.0080	0.1317	0.4854	0.9697
2011	45,470,000	-0.9504	-0.4235	-0.0802	0.0039	0.1337	0.5198	1.0210
2012	46,050,000	-0.9045	-0.3968	-0.0700	0.0140	0.1430	0.5172	0.9953
2013	46,570,000	-0.8956	-0.3942	-0.0650	0.0205	0.1439	0.4958	0.9583
2014	47,120,000	-0.8477	-0.3649	-0.0535	0.0288	0.1563	0.5039	0.9614
2015	47,610,000	-0.8333	-0.3539	-0.0416	0.0467	0.1740	0.5068	0.9452

Notes: Counts and percentiles rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. For example, 1996 contains information for both the previous year (1995) and the current year (1996). Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years.



Source: Table 6

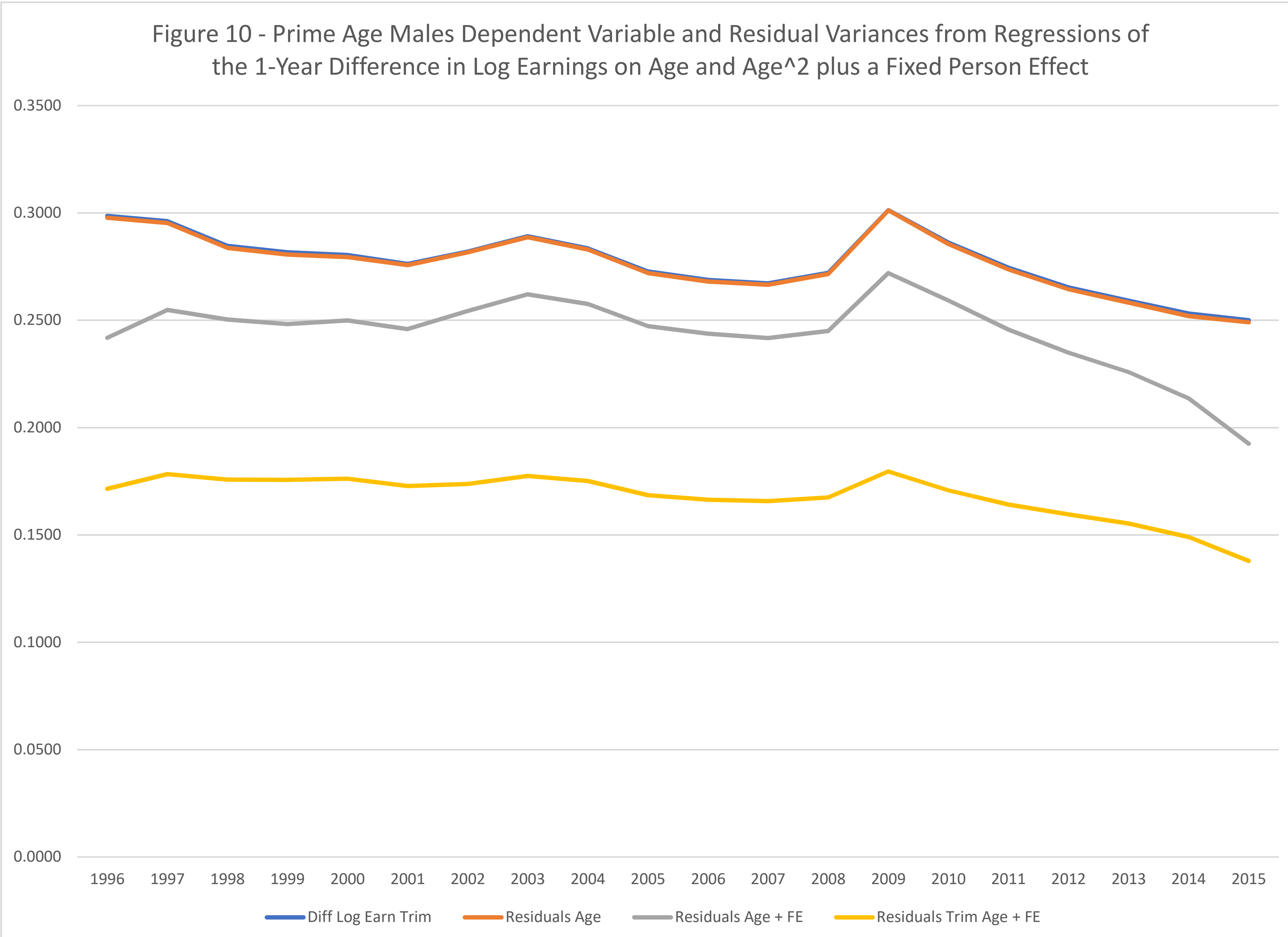


Source: Table 6

Table 7 - Prime Age Males Variance of Regression Residuals by Year

Year	Number of Observations	Diff Log Earn Trim	Residuals Age	Residuals Age + FE	Residuals Trim Age + FE
1996	28,640,000	0.2986	0.2978	0.2418	0.1715
1997	31,730,000	0.2962	0.2953	0.2548	0.1783
1998	33,290,000	0.2847	0.2837	0.2503	0.1758
1999	38,910,000	0.2817	0.2806	0.2482	0.1757
2000	40,650,000	0.2804	0.2794	0.2499	0.1762
2001	44,050,000	0.2762	0.2757	0.2459	0.1728
2002	44,290,000	0.2820	0.2817	0.2543	0.1738
2003	44,970,000	0.2891	0.2887	0.2621	0.1775
2004	45,820,000	0.2835	0.2829	0.2576	0.1751
2005	46,400,000	0.2727	0.2720	0.2473	0.1685
2006	46,600,000	0.2688	0.2680	0.2437	0.1664
2007	46,860,000	0.2672	0.2665	0.2417	0.1658
2008	46,840,000	0.2721	0.2716	0.2450	0.1675
2009	45,640,000	0.3013	0.3012	0.2720	0.1796
2010	44,410,000	0.2861	0.2855	0.2592	0.1708
2011	44,500,000	0.2745	0.2737	0.2457	0.1642
2012	45,120,000	0.2654	0.2644	0.2349	0.1596
2013	45,680,000	0.2591	0.2582	0.2259	0.1553
2014	46,240,000	0.2531	0.2520	0.2136	0.1491
2015	46,760,000	0.2500	0.2490	0.1925	0.1379

Notes: Counts and variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The dependent variable is Diff Log Earn Trim. Observations with Diff Log Earn less than P1 or greater than P99 are excluded from the analysis sample. Residuals Age includes both age and age^2. Residuals Age + FE includes age, age^2, and a fixed person effect. Residuals Trim Age + FE excludes residual values less than P1 and greater than P99.

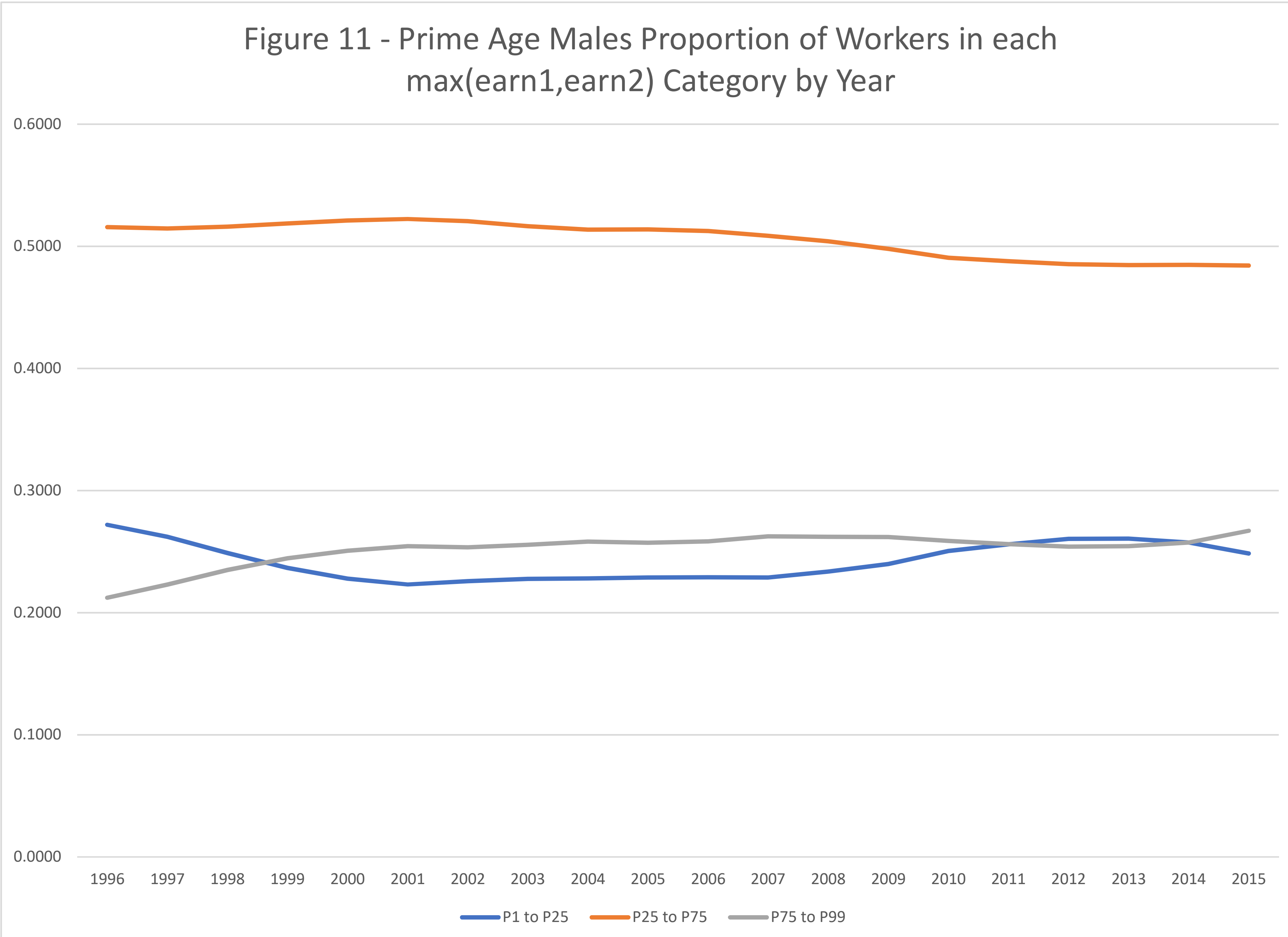


Source: Table 7

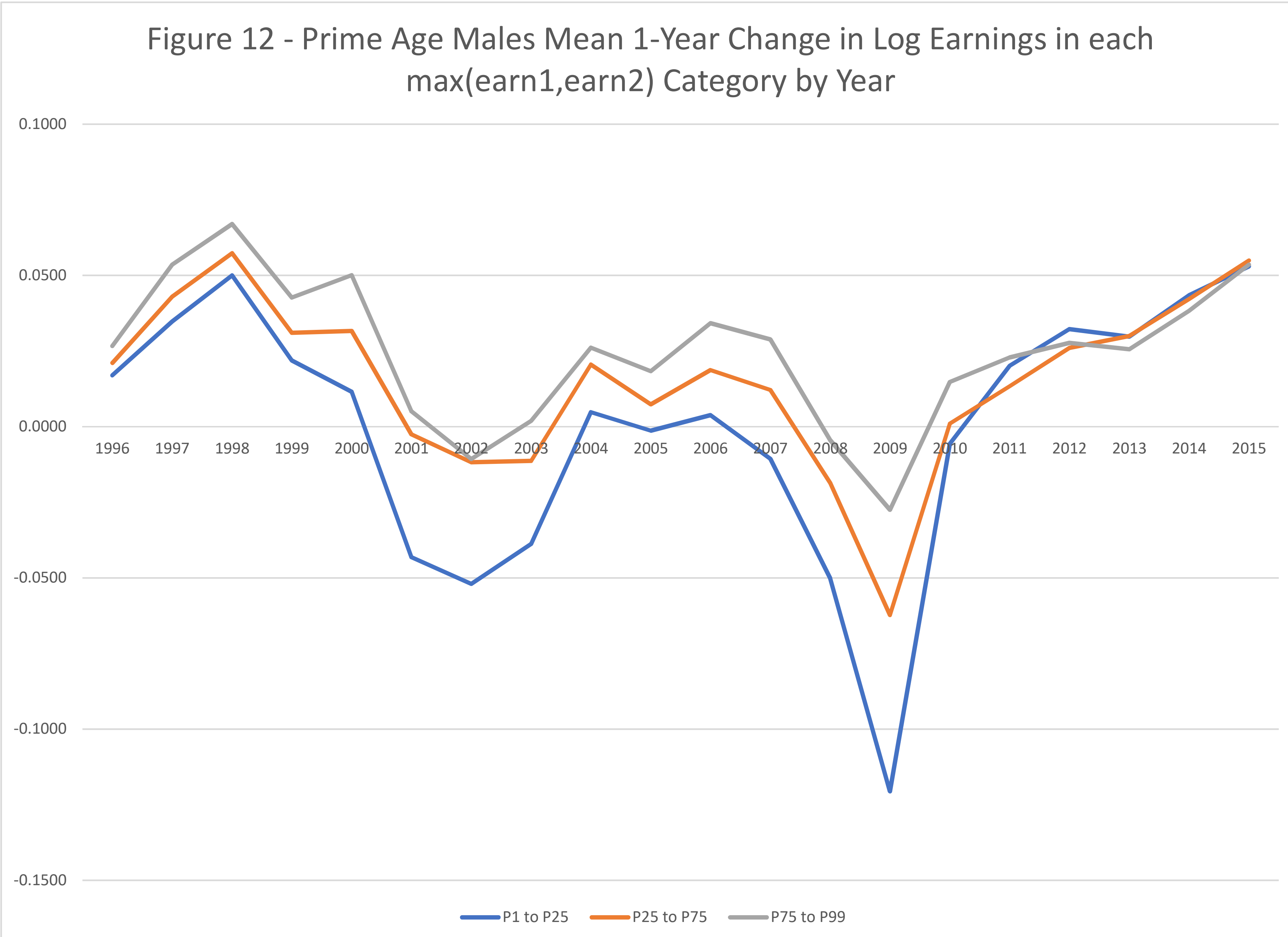
Table 8 - Prime Age Males 1 Year Change in Log Earnings by max(earn1,earn2) Category and Year

	Number of Observations			Mean			Variance		
Year	P1 to P25	P25 to P75	P75 to P99	P1 to P25	P25 to P75	P75 to P99	P1 to P25	P25 to P75	P75 to P99
1996	7,791,000	14,770,000	6,080,000	0.0170	0.0211	0.0266	0.6215	0.1835	0.1646
1997	8,321,000	16,330,000	7,080,000	0.0347	0.0430	0.0536	0.6252	0.1861	0.1632
1998	8,282,000	17,180,000	7,823,000	0.0500	0.0574	0.0670	0.6129	0.1827	0.1612
1999	9,208,000	20,180,000	9,515,000	0.0219	0.0310	0.0427	0.6110	0.1846	0.1686
2000	9,271,000	21,190,000	10,200,000	0.0115	0.0317	0.0501	0.6131	0.1860	0.1736
2001	9,829,000	23,010,000	11,210,000	-0.0432	-0.0025	0.0051	0.6172	0.1851	0.1631
2002	10,000,000	23,060,000	11,230,000	-0.0520	-0.0118	-0.0107	0.6296	0.1913	0.1577
2003	10,240,000	23,230,000	11,500,000	-0.0388	-0.0113	0.0019	0.6456	0.1995	0.1520
2004	10,450,000	23,540,000	11,830,000	0.0048	0.0205	0.0261	0.6395	0.1946	0.1458
2005	10,620,000	23,840,000	11,940,000	-0.0014	0.0073	0.0183	0.6229	0.1845	0.1372
2006	10,670,000	23,880,000	12,040,000	0.0038	0.0187	0.0342	0.6153	0.1812	0.1350
2007	10,720,000	23,830,000	12,300,000	-0.0106	0.0121	0.0288	0.6103	0.1796	0.1370
2008	10,950,000	23,610,000	12,280,000	-0.0502	-0.0187	-0.0045	0.6125	0.1836	0.1377
2009	10,950,000	22,730,000	11,960,000	-0.1206	-0.0623	-0.0275	0.6373	0.2174	0.1486
2010	11,130,000	21,780,000	11,490,000	-0.0056	0.0010	0.0148	0.6272	0.1928	0.1325
2011	11,390,000	21,710,000	11,400,000	0.0201	0.0134	0.0229	0.6030	0.1819	0.1224
2012	11,750,000	21,900,000	11,470,000	0.0323	0.0260	0.0277	0.5811	0.1721	0.1202
2013	11,910,000	22,140,000	11,630,000	0.0297	0.0300	0.0256	0.5683	0.1669	0.1182
2014	11,910,000	22,420,000	11,910,000	0.0434	0.0422	0.0383	0.5614	0.1631	0.1140
2015	11,620,000	22,640,000	12,490,000	0.0530	0.0550	0.0536	0.5603	0.1650	0.1154

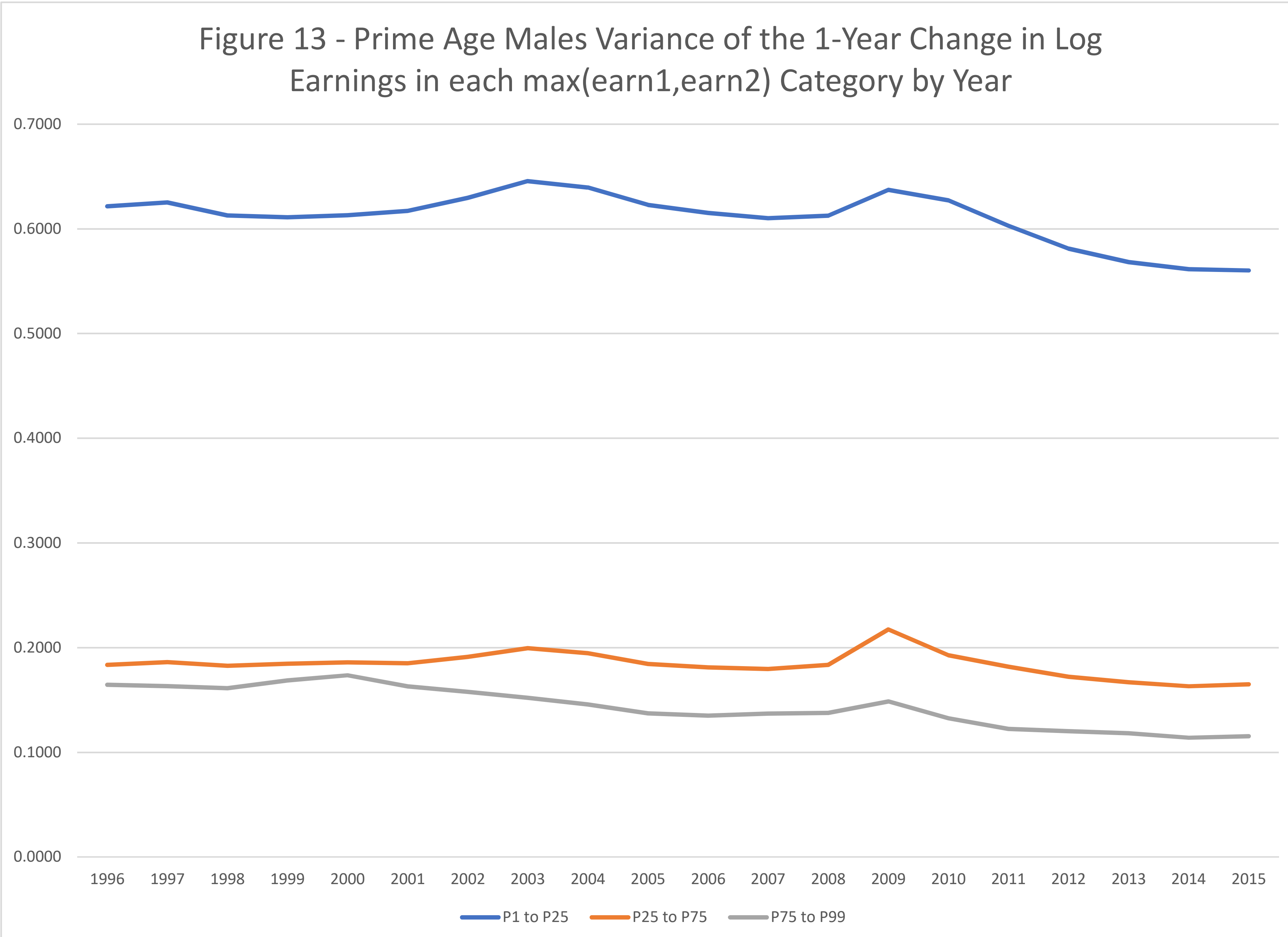
Notes: Counts, means, and variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The analysis variable is the difference between log earnings in the current and the previous year (Diff Log Earn). Observations with Diff Log Earn less than the overall sample P1 ($1,774) or greater than the overall sample P99 ($292,200) are excluded from analysis. Each observation is placed into an earnings bin based on overall sample earnings percentiles: P1 to P25 - max(earn1,earn2)<=$22,600; P25 to P75 - $22,600 < max(earn1,earn2) <=$58,590; P75 to P99 - max(earn1,earn2)>$58,590.



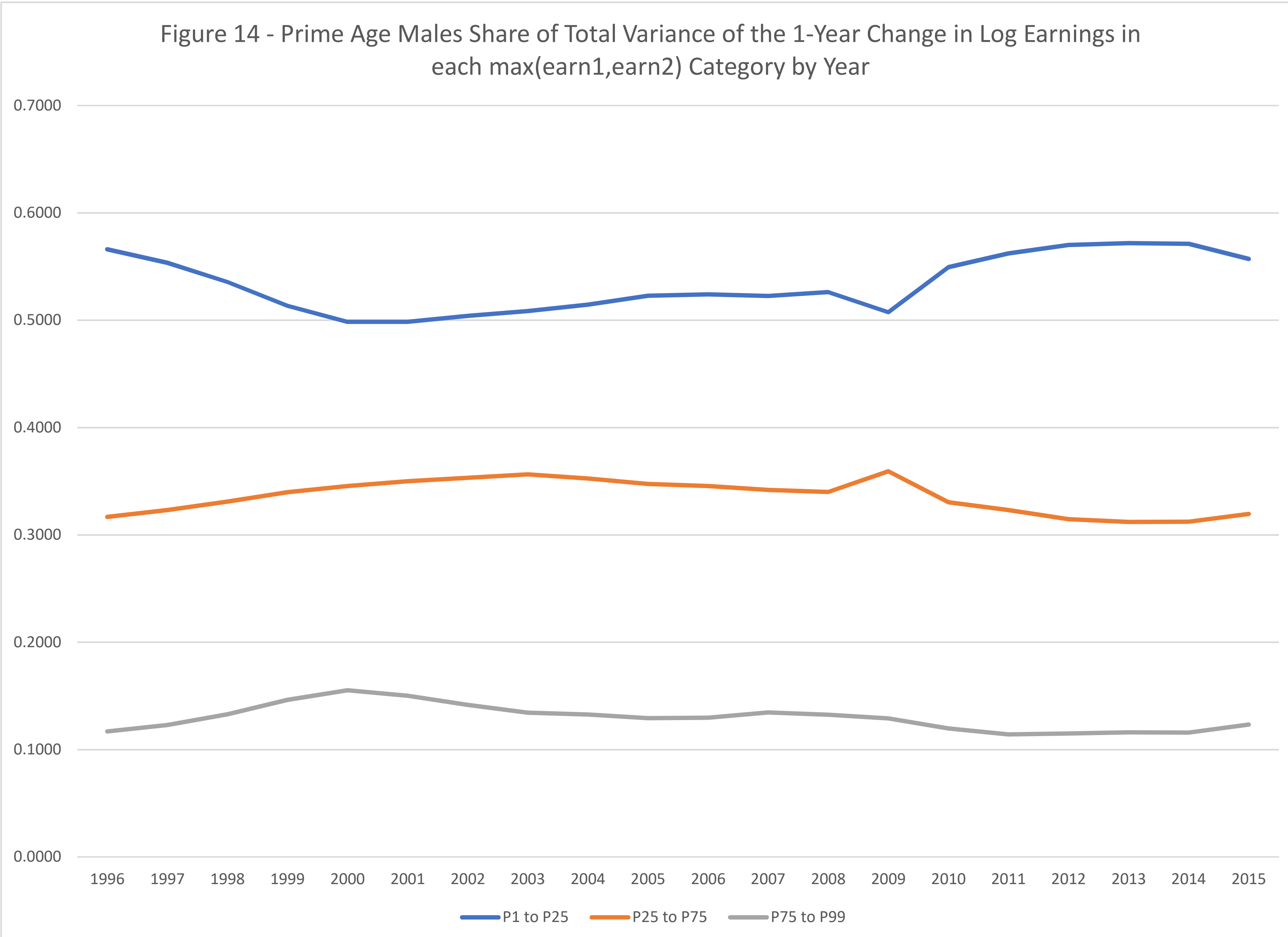
Source: Table 8



Source: Table 8



Source: Table 8

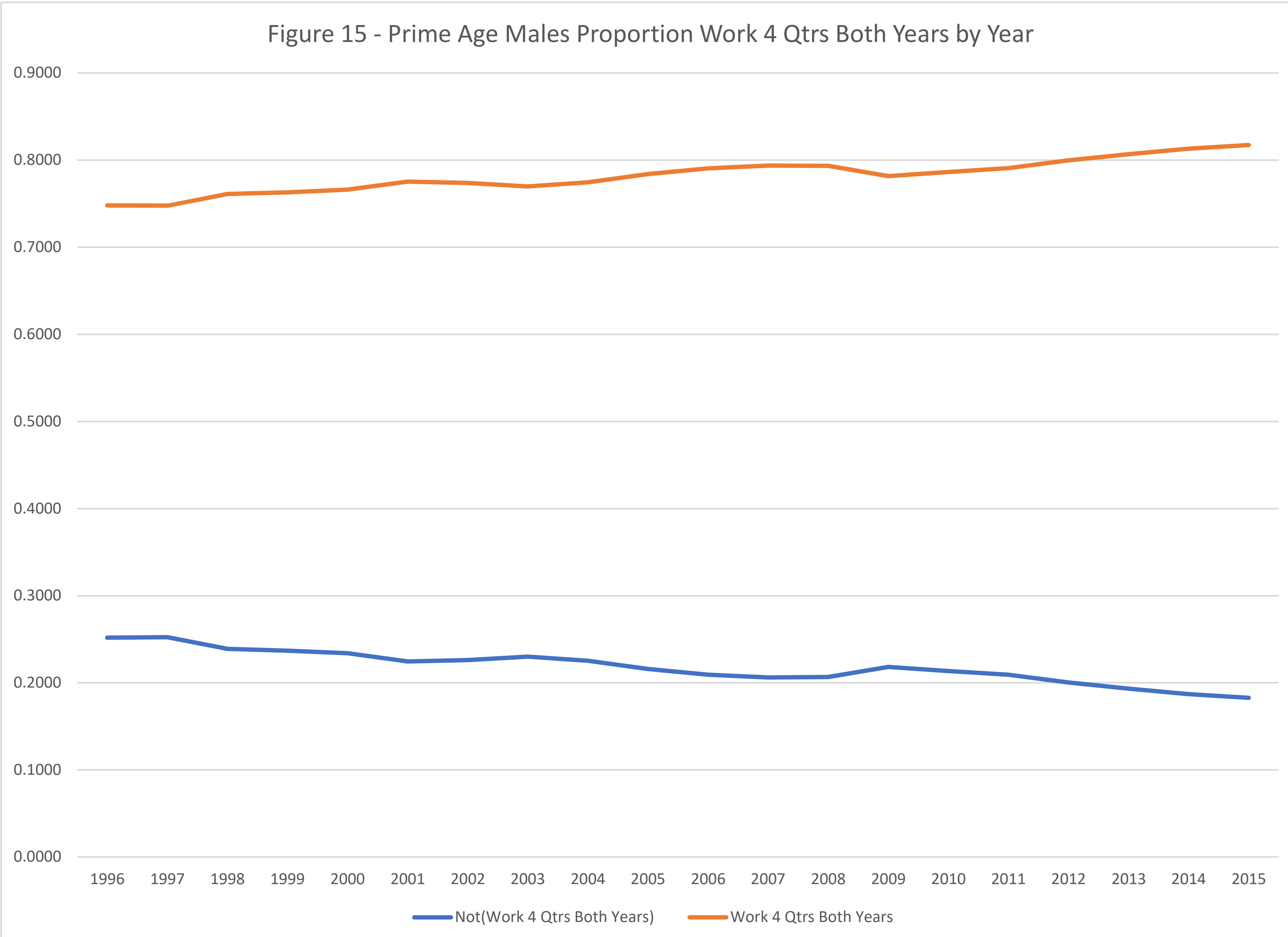


Source: Table 8

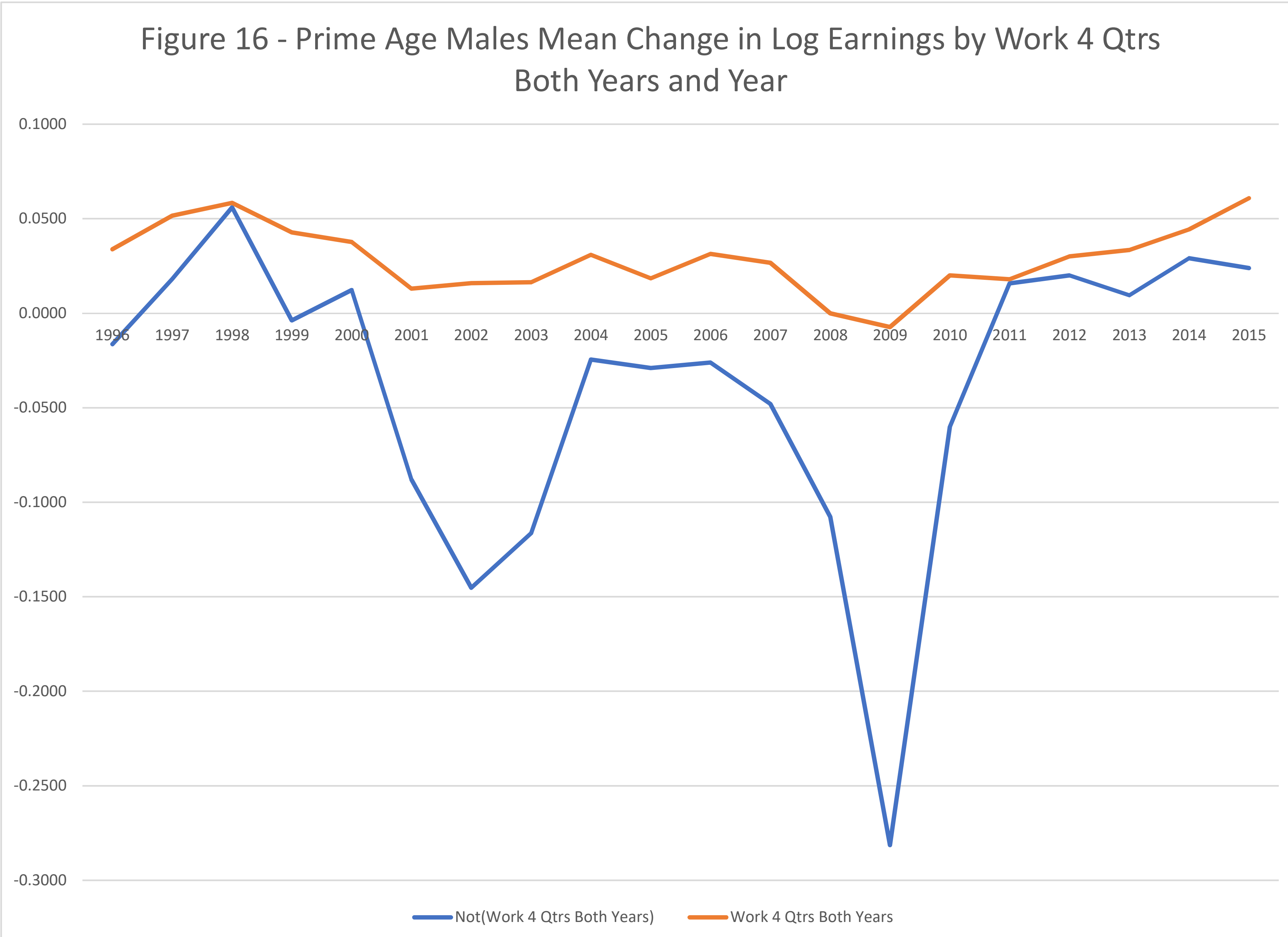
Table 9 - Prime Age Males Number of Observations, Mean, and Variance by Work 4 Quarters Both Years and Year

	Number of Observation		Mean		Variance	
Year	Not(Work 4 Qtrs Both Years)	Work 4 Qtrs Both Years	Not(Work 4 Qtrs Both Years)	Work 4 Qtrs Both Years	Not(Work 4 Qtrs Both Years)	Work 4 Qtrs Both Years
1996	7,215,000	21,430,000	-0.0164	0.0338	0.9812	0.0682
1997	8,005,000	23,720,000	0.0181	0.0516	0.9685	0.0690
1998	7,954,000	25,330,000	0.0560	0.0584	0.9698	0.0696
1999	9,219,000	29,690,000	-0.0038	0.0428	0.9581	0.0711
2000	9,511,000	31,140,000	0.0123	0.0376	0.9578	0.0734
2001	9,892,000	34,160,000	-0.0880	0.0130	0.9754	0.0715
2002	10,020,000	34,270,000	-0.1453	0.0159	0.9978	0.0670
2003	10,350,000	34,620,000	-0.1164	0.0164	1.0240	0.0654
2004	10,330,000	35,490,000	-0.0245	0.0309	1.0300	0.0657
2005	10,020,000	36,380,000	-0.0290	0.0184	1.0240	0.0651
2006	9,757,000	36,840,000	-0.0261	0.0313	1.0330	0.0657
2007	9,665,000	37,190,000	-0.0480	0.0267	1.0370	0.0661
2008	9,677,000	37,160,000	-0.1078	-0.0001	1.0540	0.0661
2009	9,964,000	35,680,000	-0.2814	-0.0073	1.0810	0.0669
2010	9,487,000	34,920,000	-0.0601	0.0200	1.1090	0.0610
2011	9,309,000	35,190,000	0.0158	0.0180	1.0890	0.0588
2012	9,037,000	36,080,000	0.0201	0.0301	1.0870	0.0595
2013	8,825,000	36,850,000	0.0095	0.0334	1.0930	0.0594
2014	8,648,000	37,600,000	0.0291	0.0443	1.0940	0.0595
2015	8,544,000	38,210,000	0.0239	0.0609	1.0940	0.0611

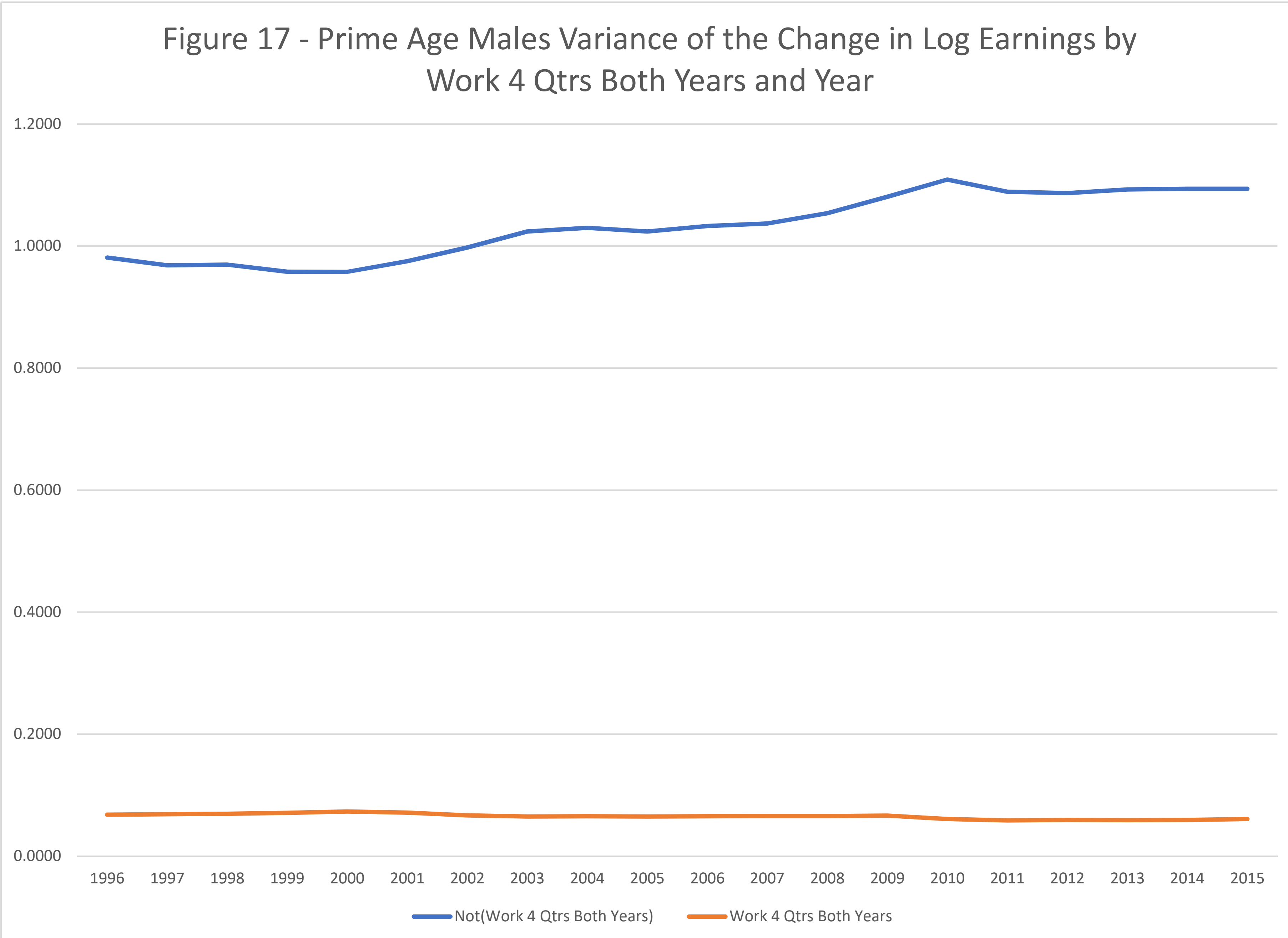
Notes: Counts, means, and variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The analysis variable is the difference between log earnings in the current and the previous year (Diff Log Earn). Observations with Diff Log Earn less than the overall sample P1 ($1,774) or greater than the overall sample P99 ($292,200) are excluded from analysis. Each observation is assigned to one of two categories: positive earnings in each of 8 consecutive quarters starting at the beginning of the previous year and ending in the last quarter of the current year; all other observations.



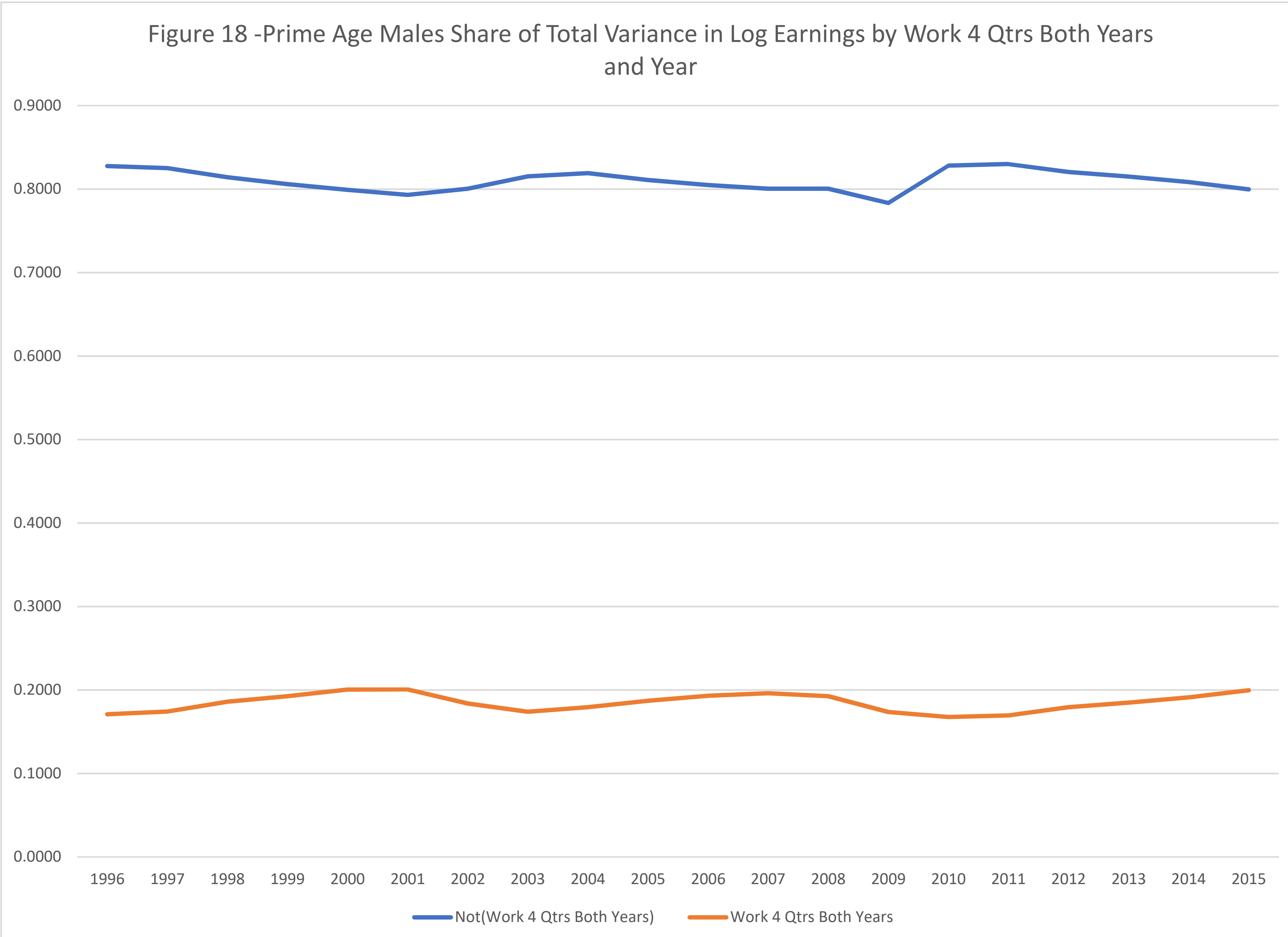
Source: Table 9



Source: Table 9



Source: Table 9

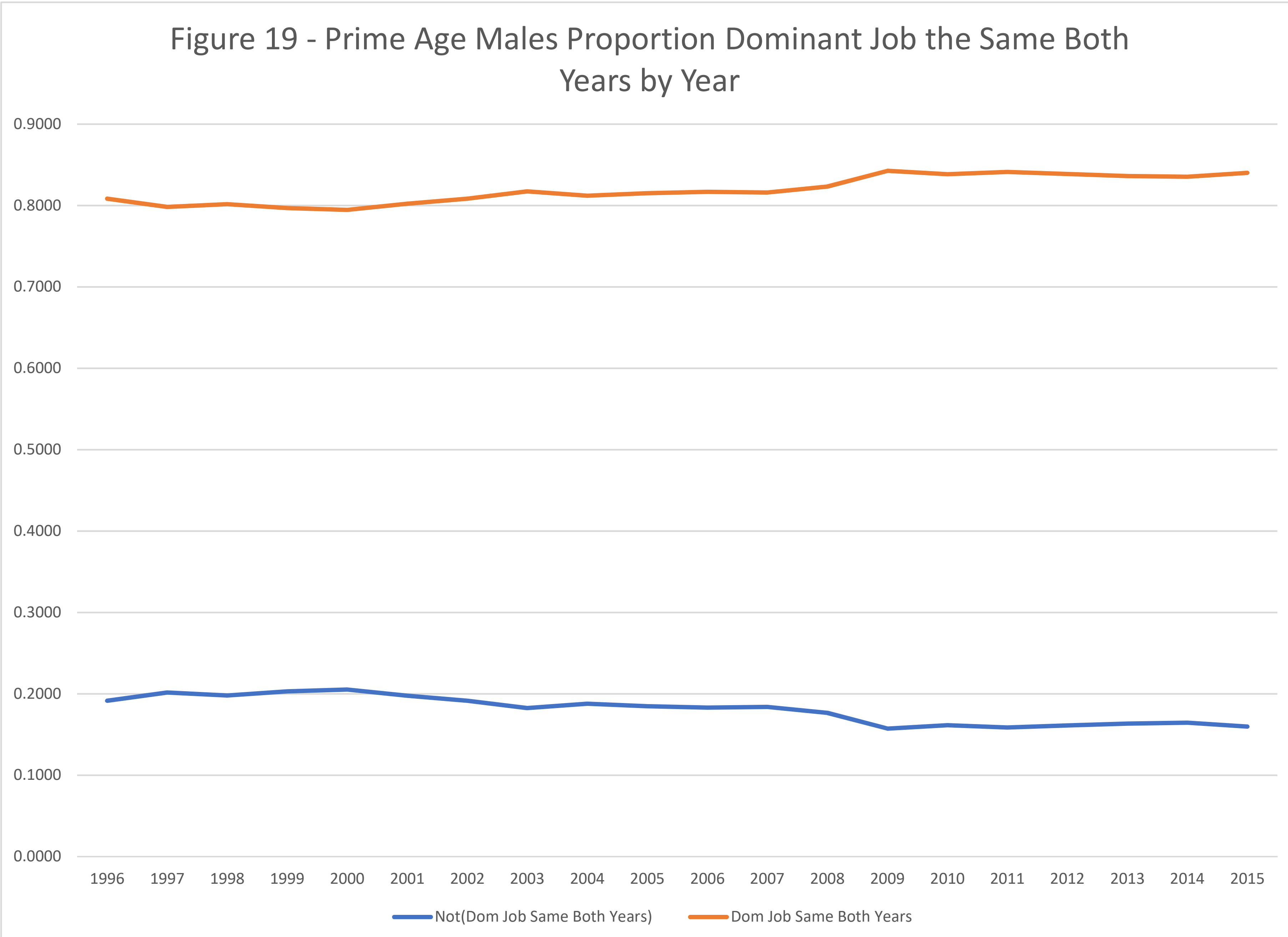


Source: Table 9

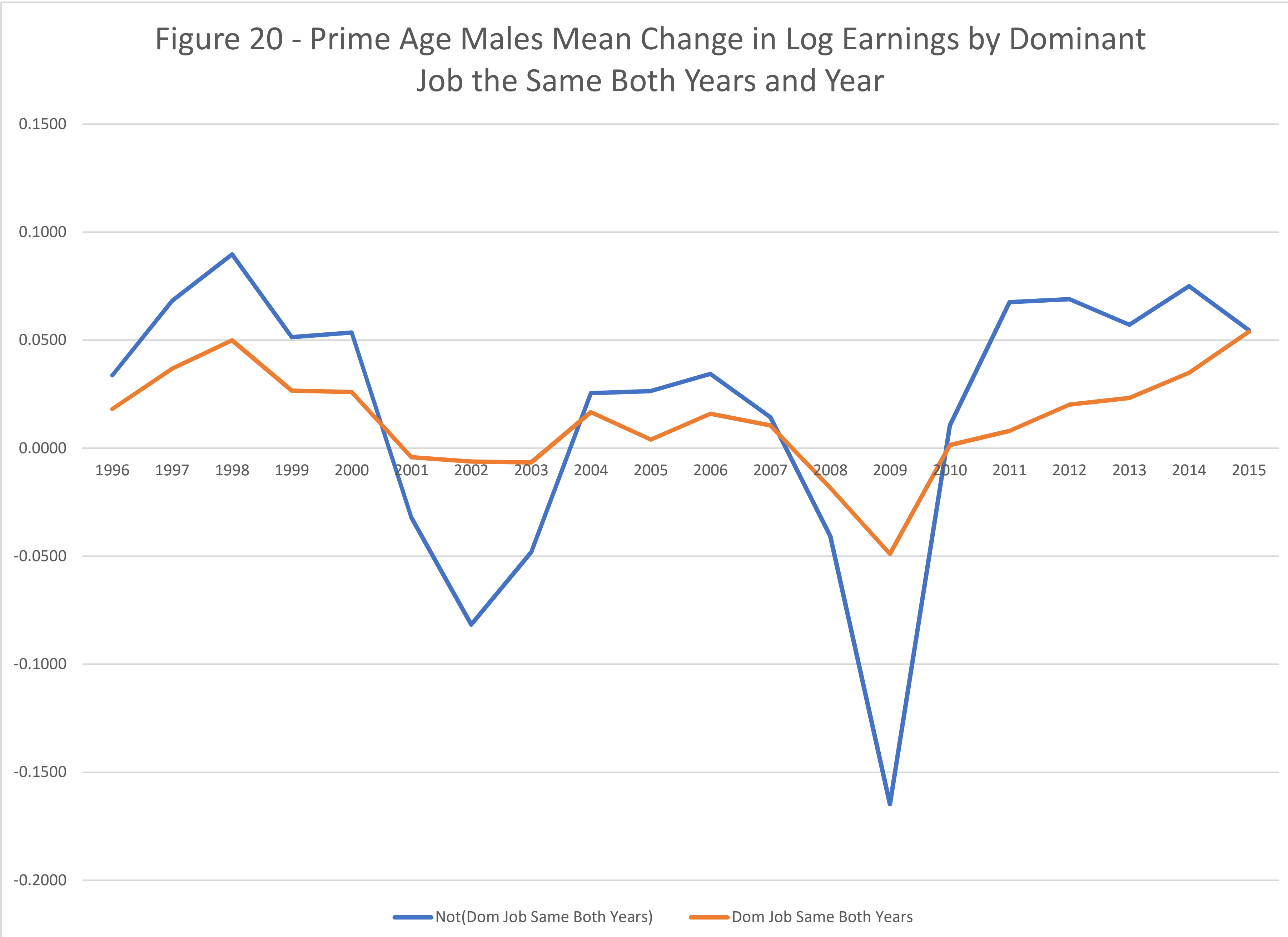
Table 10 - Prime Age Males Number of Observations, Mean, and Variance for the Change in Log Earnings by Dominant Job the Same Both Years and Year

Year	Number of Observation		Mean		Variance	
	Not(Dom Job Same Both Years)	Dom Job Same Both Years	Not(Dom Job Same Both Years)	Dom Job Same Both Years	Not(Dom Job Same Both Years)	Dom Job Same Both Years
1996	5,486,000	23,160,000	0.0337	0.0181	0.6913	0.2056
1997	6,402,000	25,330,000	0.0683	0.0368	0.6577	0.2046
1998	6,598,000	26,690,000	0.0898	0.0499	0.6326	0.1984
1999	7,905,000	31,000,000	0.0514	0.0267	0.6142	0.1968
2000	8,350,000	32,300,000	0.0535	0.0261	0.6027	0.1970
2001	8,716,000	35,340,000	-0.0321	-0.0041	0.6095	0.1939
2002	8,482,000	35,810,000	-0.0817	-0.0061	0.6618	0.1910
2003	8,214,000	36,750,000	-0.0482	-0.0066	0.6967	0.1977
2004	8,615,000	37,210,000	0.0255	0.0167	0.6749	0.1929
2005	8,573,000	37,830,000	0.0264	0.0040	0.6564	0.1857
2006	8,537,000	38,060,000	0.0344	0.0159	0.6461	0.1841
2007	8,626,000	38,230,000	0.0144	0.0106	0.6376	0.1836
2008	8,281,000	38,560,000	-0.0407	-0.0184	0.6547	0.1899
2009	7,183,000	38,460,000	-0.1648	-0.0489	0.7272	0.2196
2010	7,171,000	37,230,000	0.0106	0.0014	0.7540	0.1960
2011	7,061,000	37,440,000	0.0677	0.0081	0.7181	0.1902
2012	7,277,000	37,840,000	0.0690	0.0202	0.6906	0.1832
2013	7,475,000	38,200,000	0.0571	0.0233	0.6744	0.1777
2014	7,612,000	38,630,000	0.0750	0.0349	0.6635	0.1719
2015	7,475,000	39,280,000	0.0546	0.0540	0.6419	0.1754

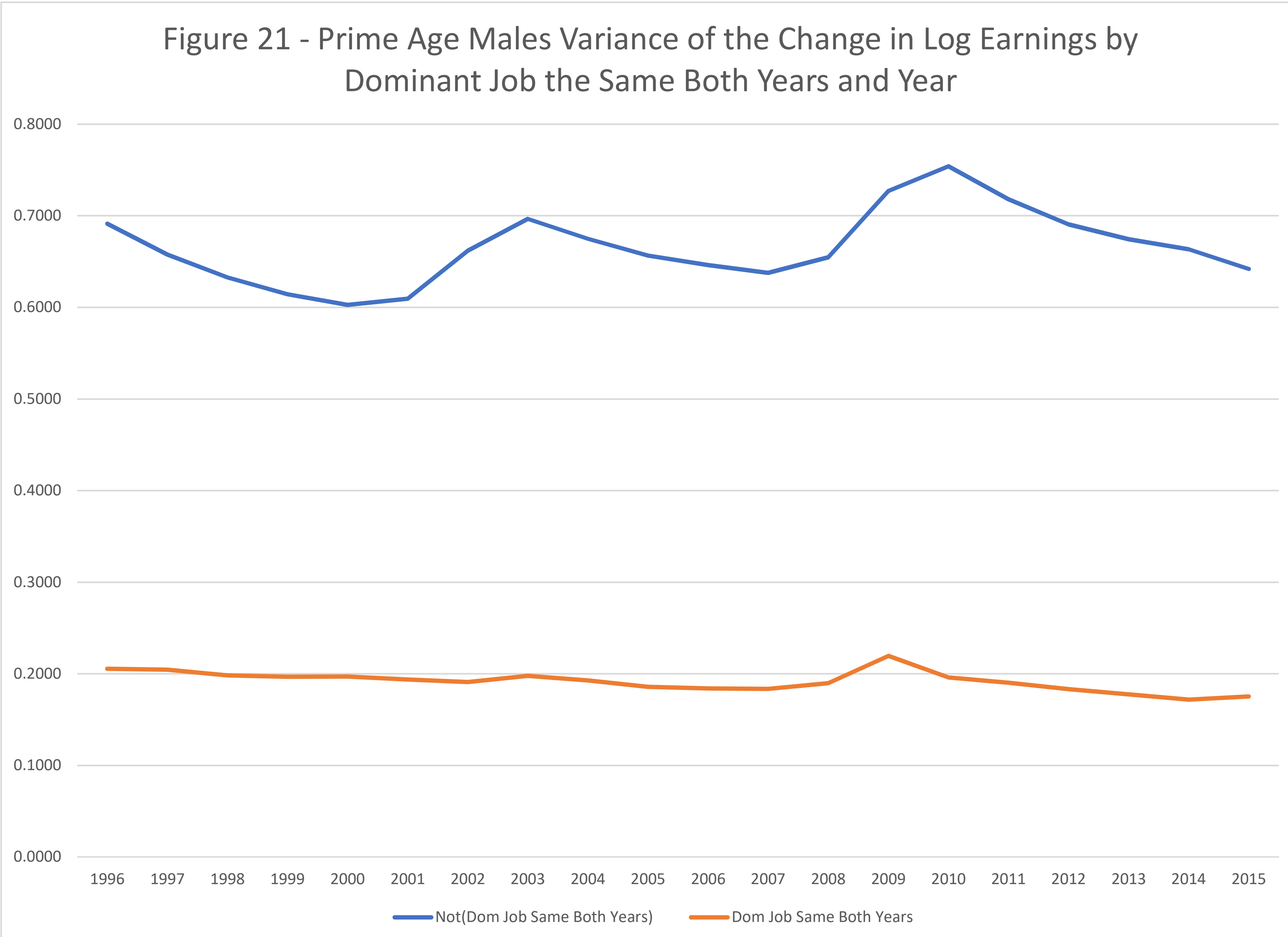
Notes: Counts, means, and variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The analysis variable is the difference between log earnings in the current and the previous year (Diff Log Earn). Observations with Diff Log Earn less than the overall sample P1 ($1,774) or greater than the overall sample P99 ($292,200) are excluded from analysis. Each observation is assigned to one of two categories: a worker's dominant job (employer with the highest earnings during the year) is the same in both years; all other observations.



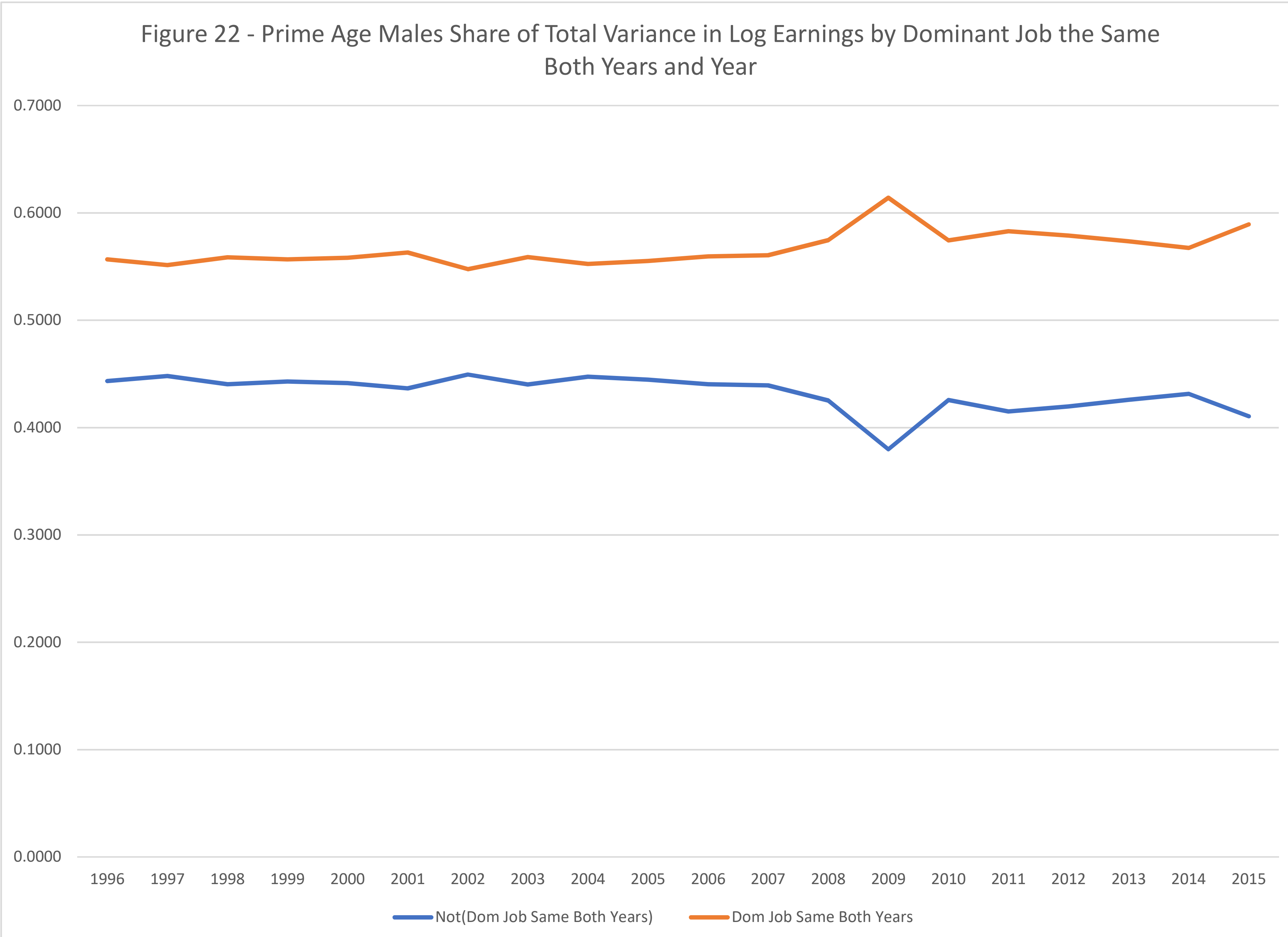
Source: Table 10



Source: Table 10



Source: Table 10

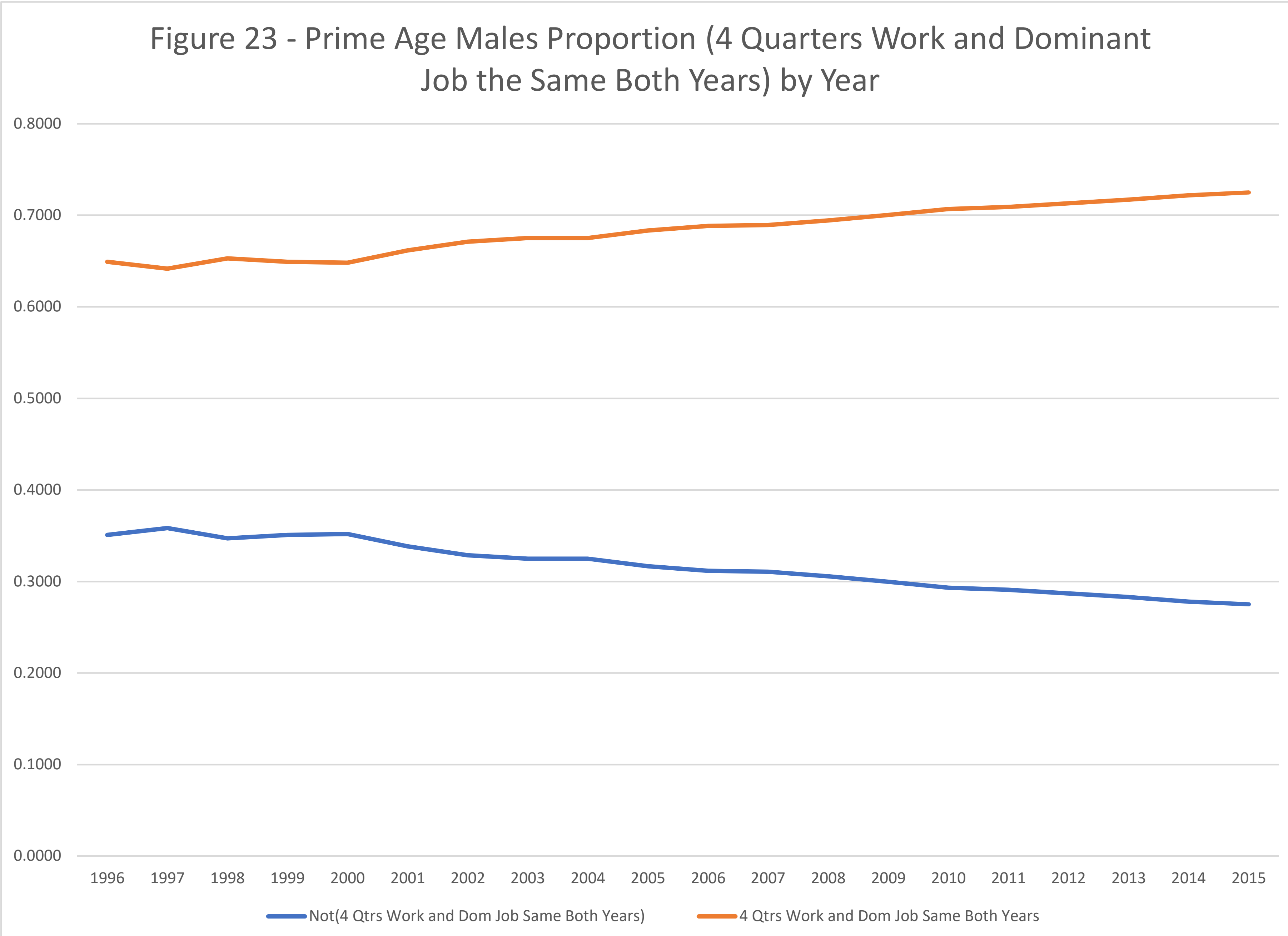


Source: Table 10

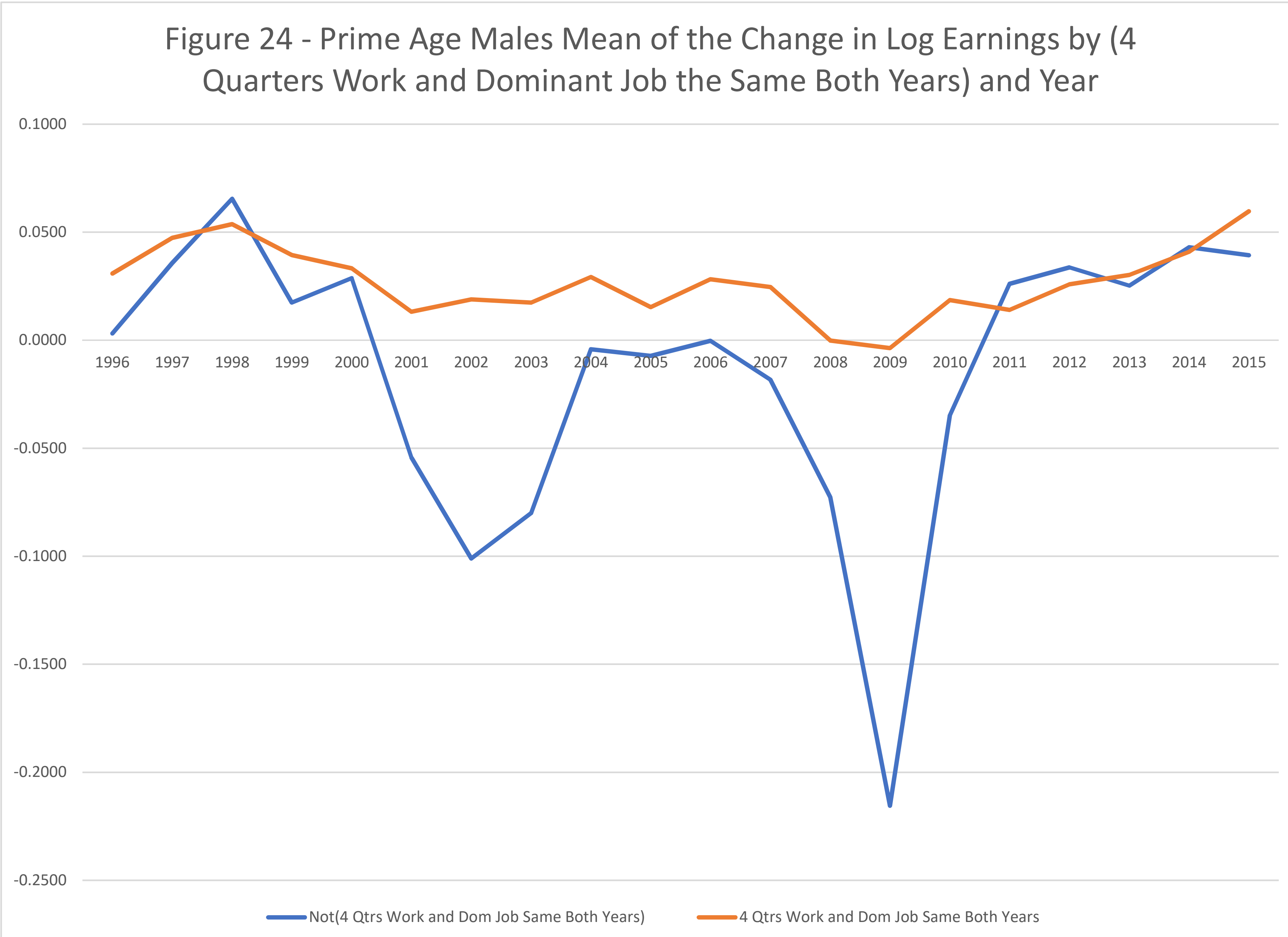
Table 11 - Prime Age Males Number of Observations, Mean, and Variance of the Change in Log Earnings by (4 Quarters Work and Dominant Job the Same Both Years) and Year

	Number of Observation		Mean		Variance	
Year	Not(4 Qtrs Work and Dom Job Same Both Years)	4 Qtrs Work and Dom Job Same Both Years	Not(4 Qtrs Work and Dom Job Same Both Years)	4 Qtrs Work and Dom Job Same Both Years	Not(4 Qtrs Work and Dom Job Same Both Years)	4 Qtrs Work and Dom Job Same Both Years
1996	10,050,000	18,590,000	0.0032	0.0309	0.7516	0.0535
1997	11,370,000	20,360,000	0.0357	0.0473	0.7302	0.0538
1998	11,550,000	21,730,000	0.0655	0.0537	0.7169	0.0549
1999	13,650,000	25,250,000	0.0175	0.0394	0.6983	0.0562
2000	14,300,000	26,350,000	0.0288	0.0333	0.6907	0.0578
2001	14,900,000	29,150,000	-0.0543	0.0132	0.7034	0.0563
2002	14,560,000	29,730,000	-0.1011	0.0189	0.7423	0.0519
2003	14,610,000	30,360,000	-0.0800	0.0175	0.7780	0.0509
2004	14,890,000	30,940,000	-0.0042	0.0292	0.7662	0.0509
2005	14,690,000	31,710,000	-0.0073	0.0153	0.7523	0.0504
2006	14,520,000	32,070,000	-0.0003	0.0282	0.7491	0.0510
2007	14,560,000	32,300,000	-0.0182	0.0246	0.7450	0.0512
2008	14,320,000	32,530,000	-0.0727	-0.0002	0.7697	0.0515
2009	13,680,000	31,960,000	-0.2155	-0.0037	0.8479	0.0537
2010	13,020,000	31,390,000	-0.0348	0.0186	0.8584	0.0479
2011	12,950,000	31,550,000	0.0261	0.0140	0.8316	0.0457
2012	12,950,000	32,170,000	0.0337	0.0258	0.8103	0.0461
2013	12,920,000	32,750,000	0.0253	0.0302	0.7995	0.0460
2014	12,860,000	33,390,000	0.0431	0.0409	0.7906	0.0460
2015	12,860,000	33,890,000	0.0393	0.0597	0.7803	0.0486

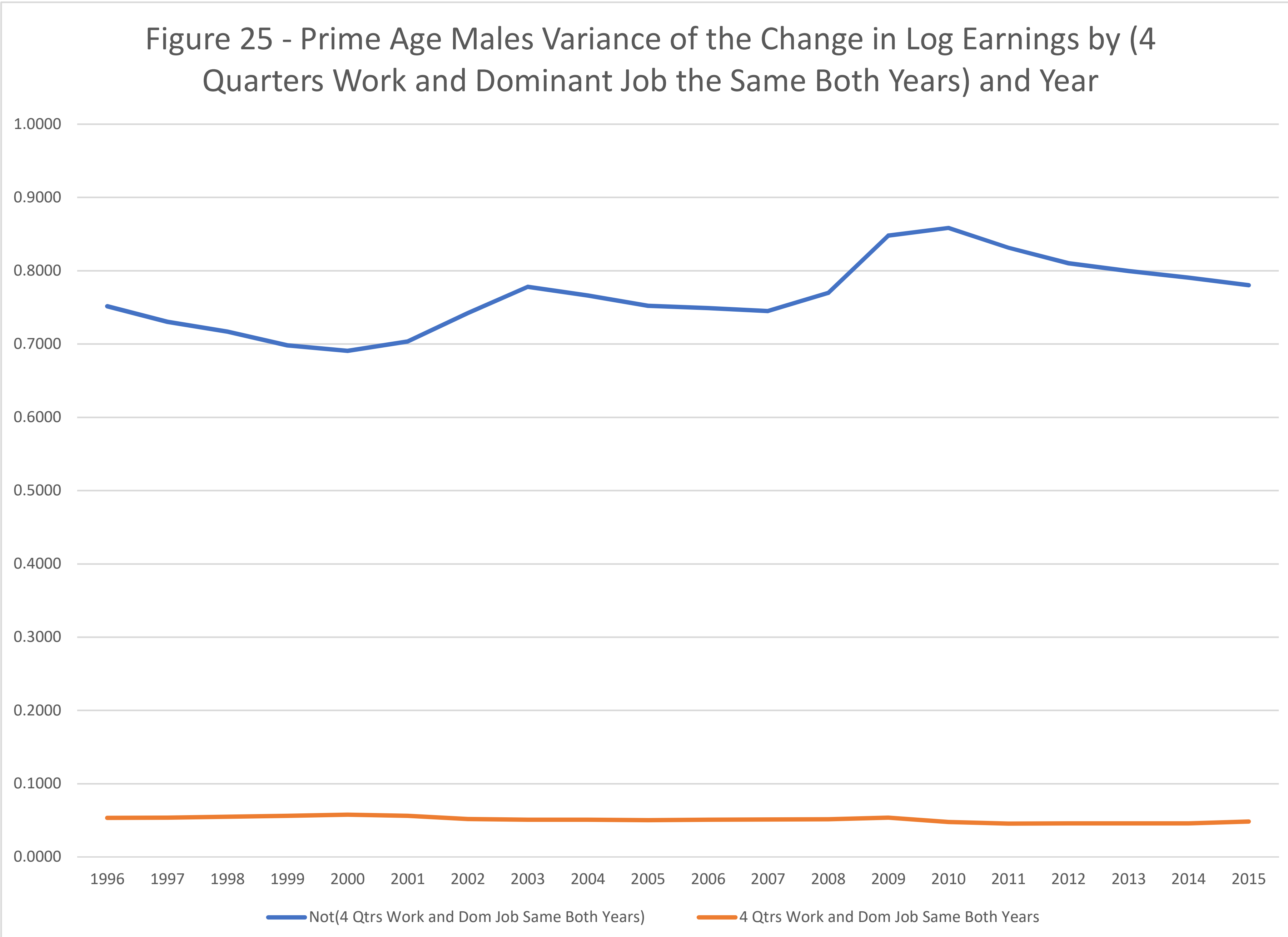
Notes: Counts, means, and variances are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The analysis variable is the difference between log earnings in the current and the previous year (Diff Log Earn). Observations with Diff Log Earn less than the overall sample P1 ($1,774) or greater than the overall sample P99 ($292,200) are excluded from analysis. Each observation is assigned to one of two categories: positive earnings in each of 8 consecutive quarters starting at the beginning of the previous year and ending in the last quarter of the current year and a worker's dominant job (employer with the highest earnings during the year) is the same in both years; all other observations.



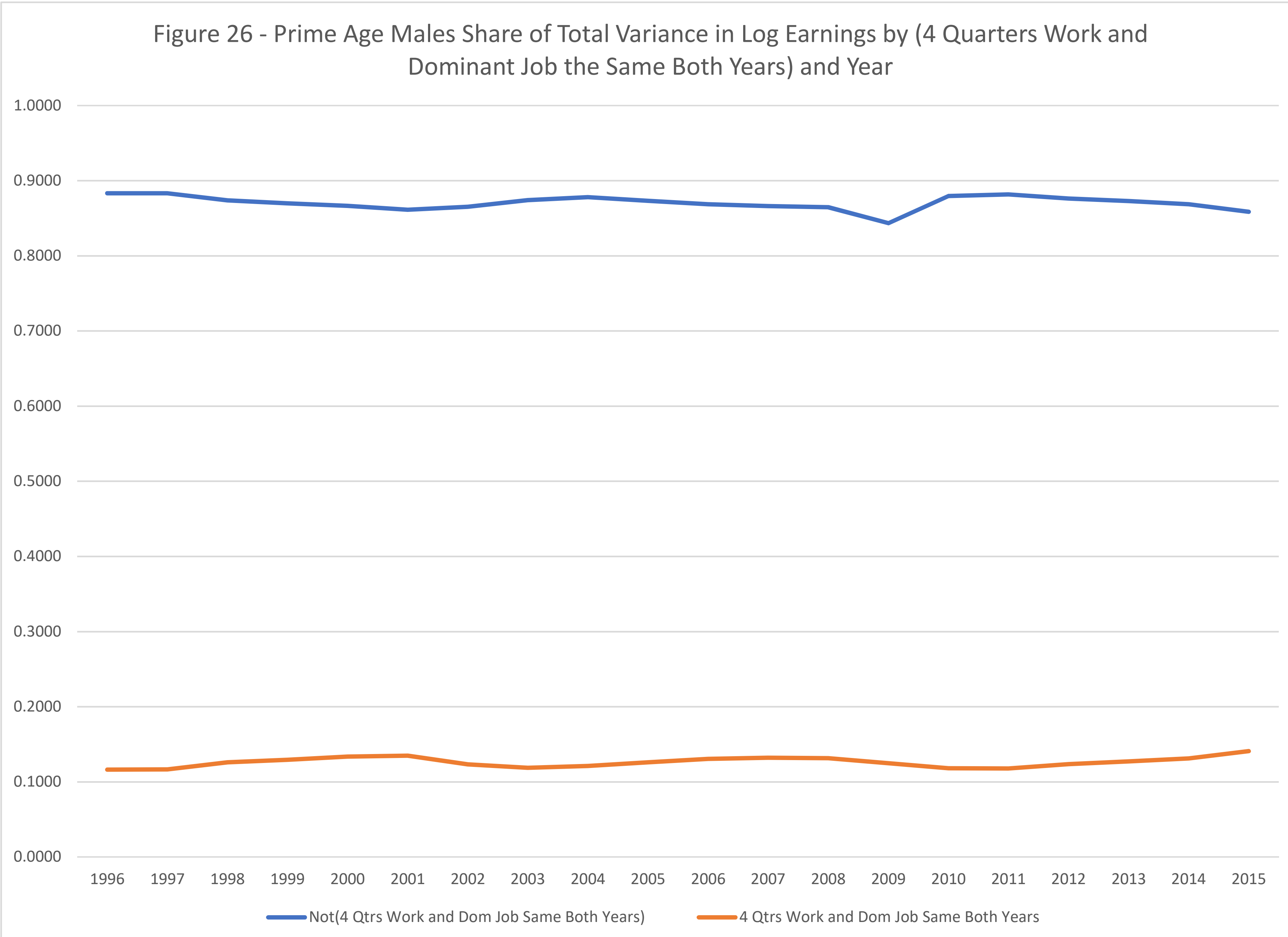
Source: Table 11



Source: Table 11



Source: Table 11



Source: Table 11

Table 12 - Prime Age Males Correlation Matrix of the Change in Log Earnings by Year and Lag Length

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1996	1.00																			
1997	-0.18	1.00																		
1998	-0.06	-0.18	1.00																	
1999	-0.02	-0.06	-0.17	1.00																
2000	-0.01	-0.02	-0.06	-0.18	1.00															
2001	0.00	-0.01	-0.02	-0.06	-0.18	1.00														
2002	-0.01	-0.01	-0.01	-0.02	-0.06	-0.16	1.00													
2003	0.00	0.00	-0.01	-0.01	-0.02	-0.07	-0.18	1.00												
2004	0.00	0.00	0.00	-0.01	-0.01	-0.03	-0.08	-0.20	1.00											
2005	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.03	-0.08	-0.17	1.00										
2006	0.00	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.03	-0.07	-0.16	1.00									
2007	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.02	-0.07	-0.17	1.00								
2008	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.03	-0.07	-0.16	1.00							
2009	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	-0.02	-0.03	-0.06	-0.13	1.00						
2010	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.08	-0.21	1.00					
2011	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01	-0.01	-0.03	-0.12	-0.16	1.00				
2012	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.05	-0.08	-0.15	1.00			
2013	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.03	-0.07	-0.15	1.00		
2014	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.01	-0.02	-0.07	-0.16	1.00	
2015	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.02	-0.02	-0.07	-0.14	1.00
Number of Observations	28,640,000	31,730,000	33,290,000	38,910,000	40,650,000	44,050,000	44,290,000	44,970,000	45,820,000	46,400,000	46,600,000	46,860,000	46,840,000	45,640,000	44,410,000	44,500,000	45,120,000	45,680,000	46,240,000	46,760,000

	Lag Length in Years																			
Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1996	-0.18	-0.06	-0.02	-0.01	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1997	-0.18	-0.06	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1998	-0.17	-0.06	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
1999	-0.18	-0.06	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
2000	-0.18	-0.06	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00					
2001	-0.16	-0.07	-0.03	-0.01	-0.01	0.00	0.00	0.01	-0.01	0.00	0.00	0.00	0.00	0.00						
2002	-0.18	-0.08	-0.03	-0.02	-0.01	0.00	0.01	0.00	-0.01	0.00	0.00	0.00	0.00							
2003	-0.20	-0.08	-0.03	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00								
2004	-0.17	-0.07	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00									
2005	-0.16	-0.07	-0.03	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00										
2006	-0.17	-0.07	-0.03	-0.01	-0.01	0.00	0.00	0.00	0.00											
2007	-0.16	-0.06	-0.02	-0.01	-0.01	-0.01	0.00	0.00												
2008	-0.13	-0.08	-0.03	-0.02	-0.01	-0.01	-0.01													
2009	-0.21	-0.12	-0.05	-0.02	-0.02	-0.01														
2010	-0.16	-0.08	-0.03	-0.01	-0.01															
2011	-0.15	-0.07	-0.02	-0.02																
2012	-0.15	-0.07	-0.02																	
2013	-0.16	-0.07																		
2014	-0.14																			
2015																				
Mean	-0.17	-0.07	-0.03	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Notes: Counts are correlations are rounded to 4 significant digits. The unit of observation is a worker year pair indexed by the current year. Prime age male workers are 25 to 59 years old, have a valid SSN on the Census Numident, the SSN is active, and the person is not reported dead. Sample includes only prime age males with positive earnings in both years. The analysis variable is the difference between log earnings in the current and the previous year (Diff Log Earn). Observations with Diff Log Earn less than the overall sample P1 ($1,774) or greater than the overall sample P99 ($292,200) are excluded from analysis.