TREASURY WORKING PAPER

00/15

A Preliminary Analysis of the Dynamics of Individual Market and Disposable Incomes

Dean Hyslop*
With assistance from Sandra Smith**

ABSTRACT

This paper reports early results from an ongoing analysis of the income dynamics of individuals, using the Inland Revenue Department's income tax database over the four- year period 1994-7.

The first two parts of the paper assess the reliability of the data. The third part reports a preliminary analysis of the dynamic properties of individual incomes. Two types of analysis are used – an analysis of transitions between quintiles of market income over time – and an analysis of the covariance of individuals market and disposable incomes through time.

The author finds that there is a high degree of consistency in the data, although inconsistencies point to the need for care in the handling of outliers. The analysis of dynamics suggests that, for this sample, a large fraction of the observed differences in incomes is transitory. For example, less than 50% of the differences in incomes persist after 3 years. However, if outliers are excluded, the degree of observed persistence rises quite strongly: about two-thirds of market income differences and about 60% of disposable income differences persist after three years in this sample. The latter result is in line with results typically found in US and other overseas analyses using panel data from household surveys. Further analysis, incorporating another year of data and additional modelling techniques, is underway.

*Policy Co-ordination and Development, NZ Treasury
**Forecasting Section, Inland Revenue Department

A Dynamic Analysis of Individual Market and Disposable Incomes

This note summarises a preliminary analysis of the income dynamics of individuals in the IRD income tax database over the four year period 1994-97. The purpose of the note is threefold. First, we provide a description of the IRD database of individual annual income tax returns, and discuss the construction of a panel sample of individuals to analyse the dynamic relationship between individuals' market and disposable incomes. Second, our aim is to provide a brief analysis of the reliability of the available data. We focus on two questions: first, how representative is the data?; and second, we assess the consistency of data from payer and payee sources. Third, we provide a preliminary analysis of the dynamic properties of individual incomes. This analysis focuses, first, on the transitions between quintiles of market income over time; and, second, on the autocovariance structure of individuals' market and disposable incomes. These two types of analyses are complementary in that the transition analysis is robust to outliers, but is insensitive to intra-quintile income dynamics; in contrast, the covariance analysis provides a useful method to characterise the persistence versus transitory nature of income differences, but can be sensitive to outliers.

Data Description

IRD Income Tax Database Description

The IRD database of individual income tax returns provides a stratified random sample of all individual tax returns in any year. The database draws random samples from two strata of tax returns: first, a 2 percent simple random sample of individual-based returns with only PAYE income (ENTYPE='I', and ENCLASS='SW') and, second, a 10 percent simple random sample of other individual-based returns. The former sample includes individuals whose income is subject to tax-withholding (wage and salary earners, beneficiaries, etc.), while the latter sample includes all other individuals with non-PAYE incomes (self-employed, etc.).

These two sampling frames are consistent over time, so that a longitudinal sample can be constructed by combining information pertaining to the same individual in different years. However, as an individual's ENCLASS may change over time, the only consistent method to construct a panel sample is to use the 2 percent 'SW' sample frame, which is also a subset of the 10 percent non-'SW' sample frame. This sample will be referred to as the '2-percent' sample.[1] It is worth emphasising that

---

[1] A second panel sample, referred to as the 'non-SW' sample, is also constructed as follows. First, if there is information on an individual for each year in the panel and their ENCLASS is unchanged throughout the period, they are included in the panel sample. Second, if their ENCLASS changes over time, but there is either IR and/or TDC information for each of the sample years, they are included in the panel sample. The problem with this panel sample is that there will be possibly non-random selection of filers into the sample, resulting in a biased sample with unknown weights. That is, filers may be excluded because their status changed and they are

this panel sample has a different characteristic from most longitudinal surveys. Not only can attrition occur through non-filing over time, but as individuals file their first returns later during the panel period "births" can occur over time, and "attritors" may also return to the sample later in the period as they file returns. Thus, there will be missing data at both ends of the panel and also during the panel period.

For the 1994-97 tax years there are two possible sources of information of individuals' incomes in each year and, for this reason, we focus on this period. The first source of information is an individual's filed tax return (IR3 or IR5). This source provides information of their income from National Superannuation, other combined PAYE earnings (wages and salaries, and welfare benefits), and other income types (interest, dividends, self-employment income, etc.); tax rebates; family support tax credits; and assessed tax. Thus, although the filed returns provide quite a detailed breakdown on income components, they do not enable wage and salary income to be distinguished from (non-Superannuation) welfare benefit income. In addition, low income individuals with only PAYE income are not required to file tax returns, so there is likely to be significant non-random selection of higher income earners in the sample of IR returns.[2] In principle, IR returns provide consistent income information on individuals who either have high wage and salary income, or receive income that is not subject to PAYE withholding tax, such as self-employment income.

The second source of information comes from Tax Deduction Certificates (TDC), which are filed by payers who withhold PAYE tax from individuals' earnings. These payers are primarily employers and the Department of Social Welfare. As each payer is required to file a TDC for each income spell by an individual, individuals can have multiple TDCs filed in a year pertaining to their distinct income spells. From the TDC source it is possible to separately identify wage and salary earnings; National Superannuation; other welfare benefits; earnings related ACC payments; family support paid during the year by NZISS; and tax deductions. In principle, there are three advantages of having the TDC information. First, it enables wage and salary income to be distinguished from welfare benefit income; second, it provides information on the wage and salary, and benefit income of individuals who do not file IR returns; and third, it allows the consistency of the PAYE income reported in IR returns and TDCs to be tested. However, in practice the TDC information tends to be less then ideal, primarily due to the inability to match individual and TDC data.

We use the information from the IR returns and TDCs to construct individual's market and disposable incomes as follows. First, market income is defined as the sum of gross non-benefit, non-superannuation income and non-family support. The simplest way to construct this measure which is consistent with the information we observe is as follows. If we have an IR return but not a TDC, we

---

not picked up in the 'SW' sampling frame, OR because they did not file a return for any reason, although their status was unchanged. For this reason, the 2-percent random sample will be the sample primarily used for the analysis, although the non-SW sample may be used as a secondary sample to supplement the analysis.
[2] The low-income limits were $20,000 in the 1994-96 tax years, and $34,200 in 1997.

assume that the reported PAYE income is entirely earned income (i.e. that the individual received no welfare benefits), and market income is then calculated as the sum of this reported PAYE income and other non-superannuation income.[3] Alternatively, if a TDC but not an IR return is available, we assume that non-PAYE income is zero (in particular, this assumes the individual received no interest or dividend income), and measure market income as the sum of the reported earned PAYE income and earnings-related ACC payments.[4] Finally, if both an IR return and a TDC are available, we measure PAYE earnings as the difference between the maximum of the IR and TDC PAYE income, and the TDC reported welfare benefits, and measure market income as the sum of this PAYE earnings and other non-benefit, non-superannuation income reported in the IR return.

Second, we define disposable income as market income plus benefit and superannuation income and family support,[5] less assessed tax net of rebates and imputation credits from dividends. If an IR return is available (with or without a TDC), we can estimate disposable income directly from the relevant components described above. If only a TDC is available, we measure disposable income as the sum of all the PAYE incomes, plus the family support paid during the year, and minus the tax deductions withheld.

Data Description and Sample Selection

In the first instance, we select individuals from the random 2% sample whose age is known and between 20 and 65 years in 1996. Table 1 provides a comparison of the characteristics of this sample, together with the subsamples of individuals with unknown age; and those aged outside 20-65 years. Those with missing age make up 9.3 percent of the sample, and are more likely to be male, have less available information from either IR returns or TDCs, have lower PAYE earnings than the 20-65 year sample, but have approximately the same taxable income. (The latter suggests individuals in this sample are more likely to be self-employed.) As expected, individuals in the young and old sample have substantially lower PAYE earnings and taxable incomes. Individuals in the sample age 20-65 are on average 38.5 years old, and 51 percent male. Over the four years 1994-97 there is, on average, IR return information for 2.8 years and TDC information for 3 years. The taxable income (from the IR returns) is on average $27,100 (in 1998 dollars); the average PAYE earnings from the IR returns is $21,900, which is slightly greater than the $21,600 in the TDCs. (This discrepancy may perhaps be explained by the inability to identify all TDCs associated with an individual.)

Table 2 presents income and filing characteristics of the 2% random sample aged 20-65 for three subsamples of individuals: those who have positive market and disposable income in each of the

---

[3] In principle, this seems reasonable as any benefit income should be reported via TDCs. Also, although the identity of individuals are unidentifiable on many TDCs, the incidence of missing individual IRD Numbers is substantially lower in the case of DSW filed TDCs (see appendix table A1).
[4] In principle this may seem unreasonable however, in practice, the non-PAYE income for this group is likely to be quite low.
[5] Family assistance in 1997.

four years; other individuals for whom there is IR and/or TDC records in each of the years;[6] and those for whom there is missing IR and TDC records in at least one year.[7] Some 64 percent of males and 54 percent of females have positive incomes in each year. Of these, columns (1) and (4) show there are both IR and TDC records for each year for over half of this group, and another 20-25 percent have either IR records for each year or TDC records for each year. Also, for this sample, the average male market and disposable incomes are $38,000 and $28,900 respectively, while the corresponding female averages are $23,500 and $19,200.

The subsamples in columns (2) and (5) include those individuals with either negative or zero market income in any year of the sample period. On average 52 percent of males and 58 percent of females in these samples have zero market income, while only about 4 percent have negative income. Also, disposable incomes are substantially higher than market incomes in these samples, and are substantially lower than their counterparts in columns (1) and (3). These factors imply this sample consists largely of welfare beneficiaries. Lastly, the samples with missing information over the sample (columns (3) and (6)) are younger, and have lower incomes than the samples of individuals with positive incomes each year. The characteristics of this sample are more similar to those of the sample with zero or negative incomes than the sample with positive incomes, which again is consistent with non-random selection of higher income individuals.

In figures 1A and 1B we present the median market and disposable incomes of males and females over the lifecycle. These medians are based on the income data stacked across the four sample years. For males, figure 1A shows a steady increase in income through the 20s and early 30s, before reaching a peak and then declining steadily from the early 50s through to retirement. Market incomes reach a peak of about $40,000 between age 40 and 45, while disposable income peak earlier at about $25,000. Except during retirement, males are net payers of tax over transfers received. Broadly similar patterns are observed for mean incomes of males; the principal difference is that mean incomes tend to be about $5,000 higher than medians, due to the skewed nature of the income differences.

The lifecycle profiles of female market and disposable incomes are presented in figure 1B. These profiles are considerably different from those of males, reflecting lifecycle fertility patterns during the 20s and 30s. Market incomes achieve a local peak in the mid-20s of about $13,000, before falling to around $8,000 in the early 30s, and then rising to a high of about $18,000 in the mid-40s. In contrast, disposable incomes exhibit a much smoother pattern than market incomes. Disposable incomes rise with market incomes during the teens and early 20s, but fall only slightly during the late 20s and early 30s, and then rise again during the late 30s and early 40s to a peak of about $15,000, before falling during the 50s to retirement. During the peak children bearing and raising ages from the mid 20s to late 30s, the median female is a net recipient of transfer payments, while during the 40s and

---

[6] Note, these individual have either negative or zero income in at least one year.

early 50s she is a net payer of taxes. A similar bimodal lifecycle income pattern is also observed in terms of mean incomes, although the drop in mean income during the late 20s and early 30s is much smaller (about $1,000). However, in contrast to the comparison of median market and disposable incomes, in terms of mean incomes females are net payers of taxes throughout their working lives.

Consistency of IR and TDC PAYE Incomes

Before analysing the dynamic properties of individual's incomes, we first consider the consistency of the income information contained in the IR returns and TDCs, which should shed some light on the accuracy of the market and disposable income measures. Table 3 provides a description of the concordance between the IR and TDC reports of PAYE income for observations with have both types of records. The analysis is presented for the entire two-percent panel sample aged 20-65, and separately for the three subsamples described above. There are a number of interesting points to note. First, the PAYE incomes are exactly equal in 80 percent of matches, and are within 10 log-points (about 10 percent) in a further 8 percent of cases. This suggests that there is a high degree of accuracy (or, at least, consistency) in the available information for the vast majority of individuals' observations. Also, the agreement is greater for the subsample which has positive market income in each of the four years.

Second, the IR reported PAYE income is on average about 8 percent higher than the TDC PAYE income, as might be expected given the problems of unidentified individuals from the TDCs. However, there is also significant under reporting of PAYE income in IR returns relative to TDCs. In fact, of the non-exact matches, TDC income is greater than IR income in slightly more than half of the matches (10.2 percent, versus 10.1 percent lower). A lot of the observed differences can be explained by stratifying the IR and TDC incomes by the number of TDCs filed in a year. For example, when a single TDC is available (in 60 percent of matches), the concordance between IR and TDC income is better. In these cases, the incomes are equal in 89 percent of the matches; also, IR exceeds TDC income in 8.8 percent of matches, and TDC exceeds IR income in only 2.2 percent. This concordance drops off steadily as the number of TDCs increase, while the relative fraction of matches in which IR incomes are lower than TDC income increases. For example, in cases with 5 or more TDCs, IR and TDC income is equal in only 36 percent of matches, IR exceeds TDC income in 20 percent of cases, while TDC exceeds IR income in 44 percent of cases. These facts suggest that individuals are more likely to file correct income information if they have fewer jobs, and are more likely to underreport income if they have more jobs, perhaps because of the "paperwork" effort required to file accurately.

Third, although the agreement in reported income on IR returns and TDCs is generally good, the relatively small number of cases with poorly matched information may generate substantial measurement error "noise" in the data. For example, the standard deviation of the log difference

---

[7] Note, individuals whose sex has been imputed have also been deleted from these samples.

between IR and TDC PAYE income is 0.6, which is quite large. This suggests that this is a potential source of outliers, and that judicious care may be required in the subsequent analysis. However, the method described above for estimating market and disposable incomes should help reduce the effects of any systematic errors in either IR income, as a result of underreporting income from multiple sources, or TDC income, resulting from missing individual identifiers.

Analysis of Income Dynamics

Transition Probabilities of Market Incomes

We next provide a descriptive analysis of the dynamics of individuals' incomes by examining transitions over time between quintiles of market incomes. In order to control for possible life cycle effects on income dynamics, we have computed age-specific quintile rankings of all available market incomes, separately for males and females. The quintile transitions are then based on these rankings.

Table 4A presents the one-year, two-year and three-year quintile-transition probabilities averaged over the available years for males, and table 4B presents the corresponding results for females. The likelihood of remaining in the same quintile from one year to the next ranges from about 50 percent for the second quintile to nearly 80 percent for the fifth quintile. Similarly, the likelihood of being in the same quintile two years later ranges from 38 percent to 68 percent; while three years later, the figures range from 32 percent to 62 percent.[8] The vast majority of transitions are between adjacent quintiles: for example, for the three middle quintiles, the one-year transition probabilities to adjacent quintiles range from 30 – 38 percent; while the two-year and three-year transition probabilities' ranges are 35 – 43 percent and 37 – 45 percent respectively.

The second point of interest in these tables is the likelihood of subsequent non-filing for various quintiles. For males, the one-year probabilities decline monotonically from 6.5 percent of those in the lowest quintile to 1 percent of those in the highest. This again indicates the non-random selection of individual tax filers in the cross-sectional and balanced panel samples. Over time however the probabilities show signs of converging: the three-year probabilities vary from about 6 percent of those in the lowest two quintiles to 2 percent of those in the highest quintile. For females, the patterns of the probabilities are similar to those of males, except that the second-quintile has a noticeably higher dropout rate than the lowest, and the dropout rates flatten more across the distribution over time. Also, quite similar patterns can be observed for transitions from non-filing to filing status.

Covariance Structures of Market and Disposable Incomes

The descriptive analysis discussed above, and presented in tables 2, 4 and appendix table A2, provides strong evidence that the observed cross-sectional samples in the IRD database non-randomly

---

[8] Note, incomes in intervening years may not be in the same quintiles.

selects higher income earners. Furthermore, this non-random selection is exacerbated when we construct a balanced panel sample, consisting of individuals with income in each of the sample years. In this section, we present some descriptions of the autocovariance structure of individuals market and disposable incomes over the sample and consider the effects of the sample selection on the serial correlation structures of individuals' incomes.

Table 5 contains the empirical autocovariances of the logarithms of market and disposable incomes, separately for males and females. This table is based on the balanced sample of individuals who have positive market (and disposable) income in each of the 1994-97 tax years. The covariances are computed after first controlling for age-specific effects, assumed to be constant in each year, and year-specific effects.[9] Panel A of the table presents the estimated autocovariances of male log incomes, and panel B presents the estimates for females. Each panel is organised as follows: the variances of market and disposable incomes for respective years are presented on the leading diagonal, the covariances between market and disposable incomes in different years are presented below the leading diagonal, the corresponding autocorrelations are presented above the leading diagonal, and the estimated standard errors of the variances and covariances are presented in parentheses below the estimates.

Consider first the autocovariances of male log incomes. First, the cross-sectional variance of market incomes is substantially greater than that of disposable income – e.g., the variance of market income lies in the range 0.9 – 1.2, while the variance of log disposable income is between 0.47 and 0.60. This suggests that transfers provide a significant income redistribution across individuals. Second, the autocorrelation structures of market and disposable incomes are quite similar and imply that the differences in incomes is quite transitory in nature. For example, the first-order correlation in either market incomes or disposable incomes is about 0.7, which implies that 70 percent of cross-sectional income differences (as measured by the variance) persist after one year. Furthermore, the third-order correlations are 0.47 for market incomes and 0.48 for disposable incomes, which implies that less than half of observed cross-sectional differences persist for three years.

Third, the cross-correlations between males' market and disposable incomes, presented in the top right-hand block of table 5A, suggest similar serial correlation patterns to the incomes separately. The diagonal elements in this block are the contemporaneous correlations between market and disposable incomes, which are about 0.88. The extent to which these contemporaneous correlations differ from 1 (presumably) reflects the heterogeneity in the population in terms of family composition which affects the eligibility for transfers, and choice decisions. The non-diagonal elements in this block are the non-contemporaneous correlations between market and disposable incomes. For example, the elements in the bottom-left and top-right corners of this block, 0.43 and 0.44, are the

---

[9] That is, a regression which includes a full set age and year dummy variables is estimated for log(market income) and log(disposable income) is estimated separately for males and females. The covariances are then computed using the residuals from these regressions.

correlations between market and disposable incomes three years apart. These reflect both the transitory nature of incomes and also the heterogeneous make-up of the population.

The results for females, presented in table 5B, are broadly similar to those of males. Differences to note are as follows. The variability in incomes (both market and disposable) is greater for females than males; however, the relative persistence in income differences is about the same. The cross-correlations between market and disposable incomes are somewhat lower for females than males, which may reflect greater heterogeneity in the eligibility for (and/or take-up) transfers of females versus males.

We present the corresponding covariance structures of market and disposable incomes in levels in appendix tables A3A and A3B for males and females respectively. The principal differences between these results and the results for log(incomes), is that the serial correlations between income levels are higher than for logs. Also, the contemporaneous correlations between market and disposable incomes are substantially higher in levels – about 0.99 for males, and 0.98 for females. Is it more appropriate to use log income or levels? We favour the use of log income for two reasons. First, the presence of outliers is likely to have a much greater influence on the results for levels than logs. Also, the distribution of incomes in levels tends to be heavily skewed, while log incomes tend to be more symmetrically distributed. Second, much of the focus on income distributions and dynamics pertains to relative rather than absolute differences, and logs provide a better measure of relative differences, while levels measure absolute differences.

In summary, these results suggest that a large fraction of the observed differences in incomes is transitory. In fact, the transitory nature of income differences observed in this sample is significantly greater compared to stylised facts, based on panel survey data for the US – e.g., typically the autocorrelation in individuals (gross) log income is generally greater than 0.5 in panel samples of 6 or more years.

The sample analysed is the balanced sample of individuals with positive incomes in each year. To give a partial evaluation of the effect of this sample selection, we estimated the corresponding covariance structures of incomes using the unbalanced sample of individuals for whom we have <u>any</u> income information over the sample period. The results from this exercise are presented in appendix tables A4A – A4D. The correlation patterns are remarkably similar, although the estimated variances and covariances are larger for this sample, largely due to the presence of more low income earners in the sample. In addition, the analysis was also restricted to samples of prime age workers (aged 25-55); the results from this analysis were essentially the same.

One possible reason for the low persistence in income differences observed here is the presence of outliers in the samples.[10] To explore this possibility, we decided to treat as "outliers" any individual

---

[10] In addition to the fact that the above sample selection only required the presence of positive incomes in each year, an indication that this may be an issue is the higher standard errors associated with the income variances compared to the autocovariances.

who has a market income in any year that is less than 10 percent of their four-year sample average income. This scheme identified 5 percent of males and 7 percent of females. The covariance structures for the reduced samples, which exclude these outliers was then estimated, and is presented in table 6 (panel A for males, B for females). The effect on the results is quite significant, although the fraction of outliers identified by this approach is quite large. First, the estimated variances are on-the-order of one-quarter to one-third lower than those in table 5, and are more precisely estimated. Second, in contrast to the variances, the estimated covariances are relatively unchanged and, consequently, the persistence in income differences as measured by the correlations is substantially higher than previously. For example, about two-thirds market income differences, and about 60 percent of disposable income differences, persist after three years in this sample. This example illustrates both the sensitivity of the results to outliers, and also the tradeoff between using a broader sample and trying to isolate the influence of possible outliers.

Summary and Extensions

The preliminary analysis of the IRD income tax database and the dynamics of individual incomes presented above provides the following broad conclusions. First, as expected, there is significant non-random selection into both the cross-sectional and panel samples of higher income individuals. Second, based on an analysis of the consistency of IR returns and TDC information, there appears to be a high level of accuracy in the reported income, at least that pertaining to PAYE income. Third, the preliminary analyses of income dynamics suggest there is substantial movement within the distribution over time. However, there is some reason to believe that these results may be driven partly by a relatively small fraction of outliers – i.e., individuals who experience extremely large year-to-year changes in income. One interpretation for this is that there is a substantial amount of heterogeneity in the population with respect to income changes over the sample period.

There are several directions to proceed from here. First, the analysis presented above is based on fairly broad population samples. Future analysis perhaps should concentrate on narrower subpopulations – e.g. excluding individuals entering retirement. In addition, a more careful and judicious sample selection to reduce the effect of outliers on the qualitative nature of the results is needed.

Second, a longer panel would be useful in order to see if the rate of decline in the serial correlation in incomes slows down and, assuming it does, to help measure and identify the persistence in income differences over a longer time frame. Ideally, at least another couple of years of data would be helpful for this purpose. However, extending the panel used here will result in greater sample selectivity, partly because of the demands of a longer panel but also because the unavailability of TDC information in earlier years will result in relatively less information being available on low income earners in other years. Third, the covariance structure analysis could be extended to specify and

estimate variance components models of income, in order to help interpret the degree of persistence in income differences and the interaction(s) between market and transfer incomes. This analysis would be the primary beneficiary of a longer panel.

Fourth, there is a need to develop a correction strategy for the sample selection bias induced by the non-random sample selection. This will be particularly important for longer panels, constructed without TDC information. The approach here will follow techniques developed by Heckman and Robb (1985), and Lee (1982, 1983), and used in related contexts by, among others, Ashenfelter and Card (1985) and Hyslop (2000). The main thrust of this approach is to specify a probabilistic sample selection rule,[11] which can then be used to correct for the selection bias.

References

Ashenfelter, Orley and David Card (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs", *Review of Economics and Statistics*, Vol. 67, pp. 648-660.

Heckman, James J., and Richard Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions", in James J. Heckman and Burton Singer (eds), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, Cambridge.

Hyslop, Dean R. (2000), "The Covariance Structure of Intrafamily Earnings, Rising Inequality and Family Labor Supply", Manuscript, Department of Economics, UCLA.

Lee, Lung-Fei (1982), "Some Approaches to the Correction of Selectivity Bias", *Review of Economic Studies*, Vol. 49, pp. 355-372.

Lee, Lung-Fei (1983), "Generalized Econometric Models with Selectivity", *Econometrica*, Vol. 51, pp. 507-512.

---

[11] In this context, the probability of selection will, likely, depend on positively on individuals' income level.

Table 1: IRD Panel Sample Selection Criteria and Characteristics

	Subsample of Two-Percent Client Population		
	Missing Age (1)	Age<20, or Age>65 (2)	Aged 20-65 (3)
Age (in 1996)	.	.	38.47 (.06) [12.27]
Female	0.21 (.01)	0.53 (.004)	0.49 (.002)
Imputed Sex	0.03 (.002)	0.003 (.0004)	0.01 (.0004)
# IR Records (over 4 years)	1.31 (.02) [1.73]	2.06 (.01) [1.66]	2.76 (.01) [1.55]
IR PAYE Earnings[a]	11.81 (.35) [18.55]	8.72 (.08) [8.28]	21.89 (.11) [22.32]
IR Taxable Income[a]	28.19 (.90) [47.60]	12.96 (.20) [21.29]	27.06 (.16) [31.92]
# TDC Records (over 4 years)	1.53 (.02) [1.29]	2.80 (.01) [1.48]	3.01 (.01) [1.42]
TDC PAYE Earnings[a]	10.66 (.24) [17.23]	8.67 (.07) [7.81]	21.55 (.11) [22.24]
Observations	6,356	15,469	46,258

Notes: Estimated standard errors are in parentheses, standard deviations are in square brackets.

(a) Earnings and Incomes are sample-period averages, expressed in 1998 thousands of dollars, adjusted by the CPI.

Table 2: Two-Percent Sample, Aged 20-65

	Males			Females		
	Positive Income Each Year (1)	Non-positive Income Any Year (2)	Missing IR & TDC Any Year (3)	Positive Income Each Year (4)	Non-positive Income Any Year (5)	Missing IR & TDC Any Year (6)
Age (in 1996)	39.10 (.10)	40.61 (.22)	35.31 (.17)	39.13 (.11)	38.92 (.18)	35.70 (.16)
IR & TDC 4 years	0.527	0.176	.	0.510	0.179	.
IR 4 years	0.251	0.183	.	0.235	0.187	.
TDC 4 years	0.207	0.604	.	0.236	0.590	.
IR or TDC 4 years	0.016	0.036	.	0.019	0.044	.
IR or TDC 1-3 years	.	.	0.853	.	.	0.861
No IR or TDC	.	.	0.147	.	.	0.139
Zero Market Income (sample average)	.	0.522	0.141	.	0.575	0.192
Negative Mkt Income (sample average)	.	0.045	0.008	.	0.036	0.010
Market Income[a] (sample average)	38.01 (.35)	6.32 (.21)	15.28 (.39)	23.51 (.17)	3.17 (.08)	9.15 (.20)
Disposable Income[a] (sample average)	28.89 (.25)	9.70 (.15)	12.42 (.27)	19.21 (.12)	9.37 (.08)	8.15 (.21)
log(Mkt Income)[b] (sample average)	10.15 (.01)	8.44 (.03)	8.78 (.03)	9.60 (.01)	7.77 (.03)	8.27 (.03)
log(Disp Income)[b] (sample average)	10.00 (.01)	8.92 (.01)	8.60 (.02)	9.59 (.01)	8.93 (.01)	8.27 (.02)
Observations	14,748	3,820	4,577	12,258	5,361	4,918

Notes: Estimated standard errors are in parentheses. The samples exclude individuals with imputed sex. The samples in columns (1) and (3) consist of individuals who have positive market and disposable income in all 4 years. The samples in columns (2) and (4) are the remainder of the sample described in table 1, column (3), and include those who have negative or missing income in at least one year.

(a) Incomes are expressed in 1998 thousands of dollars, adjusted by the CPI.

(b) log(Income, expressed in 1998 dollars).

Table 3: Comparison of IR and TDC PAYE Incomes

Log Difference[a]	Two-Percent Sample	Subsample		
		Positive Income Each Year	Non-Positive Income	Missing Data
0	0.797	0.806	0.769	0.756
< 0.01	0.025	0.026	0.020	0.025
< 0.05	0.035	0.034	0.038	0.038
< 0.10	0.019	0.018	0.024	0.021
> 0.10	0.124	0.116	0.140	0.160
Mean difference[b]	0.085	0.100	0.006	0.094
Standard deviation[b]	0.599	0.626	0.388	0.648
Fraction of Sample[c] Missing	0.445	0.276	0.568	0.809
Sample Size	185,032	109,868	36,792	38,732

Notes: The fractions are based on all IR and TDC pairs over the sample period 1994-97. The sample sizes refer to the total possible matches – i.e. 4 times the number of individuals in each subsample.

(a) Measures the "relative difference" in the IR and TDC PAYE incomes, equal to the magnitude of log(IR PAYE) – log(TDC PAYE), if both incomes are non-zero.

(b) If IR and TDC PAYE income are both 0, the log difference is set to 0; observations with either IR or TDC PAYE income 0 (but not both) are excluded.

(c) Fraction of the sample size without both IR and TDC information.

Table 4A: Market Income Transition Probabilities, 1994-97 -- Males

First-Year Quintile	End-Year Quintile					
	1	2	3	4	5	Missing Data
One-Year Transitions						
1	**0.663**	0.192	0.051	0.018	0.009	0.067
2	0.163	**0.503**	0.205	0.058	0.028	0.043
3	0.041	0.162	**0.560**	0.180	0.039	0.019
4	0.020	0.054	0.145	**0.612**	0.158	0.012
5	0.015	0.030	0.043	0.133	**0.769**	0.011
Missing Data	0.192	0.114	0.037	0.022	0.018	0.617
Two-Year Transitions						
1	**0.587**	0.214	0.079	0.040	0.022	0.058
2	0.183	**0.408**	0.228	0.088	0.045	0.048
3	0.064	0.180	**0.466**	0.200	0.062	0.028
4	0.035	0.073	0.169	**0.516**	0.184	0.023
5	0.028	0.047	0.063	0.157	**0.683**	0.023
Missing Data	0.245	0.163	0.065	0.046	0.039	0.442
Three-Year Transitions						
1	**0.526**	0.236	0.099	0.052	0.029	0.058
2	0.180	**0.362**	0.229	0.107	0.066	0.056
3	0.073	0.177	**0.412**	0.215	0.086	0.036
4	0.050	0.095	0.177	**0.451**	0.195	0.033
5	0.034	0.062	0.069	0.177	**0.628**	0.031
Missing Data	0.273	0.178	0.087	0.054	0.054	0.355

Notes: Entries are probabilities of being in the "End-year" quintile, conditional on being in the "First-year" quintile – i.e. the numbers in each row sum to 1. Income quintiles are age-specific quintiles, based on all sample incomes.

Sample size = 23,145 individuals. One-year transition probabilities are averaged over the 1994-95, 1995-96 and 1996-97 transitions; two-year probabilities are averaged over the 1994-96 and 1995-97 transitions; and three-year transitions pertain to the 1994-97 transitions. Fractions of missing incomes are as follows. one-year transitions: 11.8% of first-year and 9.9% of second-year incomes; two-year transitions: 13.1% of first-year and 8.9% of third-year incomes; and three-year transitions: 14.3% of first-year and 8.7% of fourth-year incomes.

Table 4B: Market Income Transition Probabilities, 1994-97 -- Females

First-Year Quintile	End-Year Quintile					
	1	2	3	4	5	Missing Data
One-Year Transitions						
1	**0.696**	0.184	0.053	0.019	0.006	0.043
2	0.158	**0.480**	0.225	0.058	0.020	0.060
3	0.055	0.156	**0.506**	0.204	0.047	0.032
4	0.018	0.045	0.155	**0.591**	0.172	0.019
5	0.009	0.020	0.048	0.136	**0.775**	0.012
Missing Data	0.119	0.159	0.067	0.031	0.018	0.607
Two-Year Transitions						
1	**0.582**	0.222	0.090	0.042	0.018	0.046
2	0.185	**0.372**	0.246	0.094	0.043	0.059
3	0.080	0.166	**0.394**	0.237	0.081	0.042
4	0.040	0.067	0.167	**0.481**	0.214	0.032
5	0.022	0.039	0.070	0.158	**0.683**	0.029
Missing Data	0.160	0.205	0.108	0.061	0.042	0.426
Three-Year Transitions						
1	**0.520**	0.232	0.116	0.055	0.029	0.049
2	0.196	**0.315**	0.254	0.118	0.058	0.059
3	0.091	0.167	**0.331**	0.253	0.109	0.049
4	0.057	0.076	0.172	**0.413**	0.241	0.041
5	0.029	0.053	0.090	0.171	**0.618**	0.040
Missing Data	0.183	0.210	0.135	0.079	0.062	0.332

Notes: Entries are probabilities of being in the "End-year" quintile, conditional on being in the "First-year" quintile – i.e. the numbers in each row sum to 1. Income quintiles are age-specific quintiles, based on all sample incomes.

Sample size = 22,837 individuals. One-year transition probabilities are averaged over the 1994-95, 1995-96 and 1996-97 transitions; two-year probabilities are averaged over the 1994-96 and 1995-97 transitions; and three-year transitions pertain to the 1994-97 transitions. Fractions of missing incomes are as follows. one-year transitions: 12.6% of first-year and 10.6% of second-year incomes; two-year transitions: 14.0% of first-year and 9.5% of third-year incomes; and three-year transitions: 15.3% of first-year and 9.1% of fourth-year incomes.

Table 4C: Market Income Transition Probabilities, 1994-97 -- Males

Sample Average Market Income Quintile	Fraction of Time Spent in Annual Income Quintile					Total Years of Data
	1	**2**	**3**	**4**	**5**	
			Males			
1	0.86	0.14	0.004	0.001	0	3.19
2	0.21	0.60	0.16	0.02	0.003	3.58
3	0.04	0.22	0.56	0.15	0.02	3.79
4	0.02	0.05	0.20	0.61	0.13	3.87
5	0.01	0.02	0.03	0.16	0.79	3.89
			Females			
1	0.78	0.22	0.003	0.001	0	3.36
2	0.21	0.58	0.20	0.01	0.001	3.48
3	0.06	0.19	0.56	0.17	0.02	3.69
4	0.02	0.05	0.21	0.59	0.14	3.81
5	0.004	0.01	0.03	0.18	0.78	3.88

Notes: Entries are the fraction of time spent in each annual-income quintile, conditional on being in the Sample average market income quintile. Income quintiles are age-specific quintiles, based on all sample incomes. Sample sizes: 23,145 males; 22,837 females.

Table 5A: Covariance Structure of Logarithms of Market and Disposable Incomes, 1994-97 -- Males

	Log(Market Income) in					log(Disposable Income) in			
	1994	**1995**	**1996**	**1997**		**1994**	**1995**	**1996**	**1997**
log(Market Income) in									
1994	1.171 (.03)	0.687	0.575	0.472		0.867	0.611	0.514	0.426
1995	0.719 (.02)	0.934 (.04)	0.730	0.570		0.607	0.882	0.638	0.504
1996	0.599 (.02)	0.679 (.02)	0.926 (.03)	0.690		0.526	0.666	0.895	0.606
1997	0.532 (.02)	0.573 (.02)	0.690 (.02)	1.080 (.03)		0.444	0.532	0.616	0.890
log(Disposable Income) in									
1994	0.709 (.02)	0.444 (.01)	0.383 (.01)	0.350 (.01)		0.573 (.02)	0.677	0.572	0.478
1995	0.454 (.01)	0.585 (.01)	0.439 (.01)	0.379 (.01)		0.351 (.01)	0.470 (.01)	0.718	0.569
1996	0.396 (.01)	0.439 (.01)	0.613 (.02)	0.456 (.01)		0.308 (.01)	0.351 (.01)	0.507 (.02)	0.663
1997	0.356 (.01)	0.376 (.01)	0.450 (.01)	0.714 (.02)		0.279 (.01)	0.301 (.01)	0.364 (.01)	0.595 (.02)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is a bal anced sample of 14,748 individuals aged 20-65 with positive income in every year.

| | Log(Market Income) in | | | | | log(Disposable Income) in | | | |
	1994	1995	1996	1997		1994	1995	1996	1997
log(Market Income) in									
1994	1.678	0.727	0.592	0.504		0.819	0.588	0.479	0.400
	(.04)								
1995	1.139	1.460	0.739	0.599		0.581	0.819	0.599	0.472
	(.03)	(.04)							
1996	0.887	1.033	1.339	0.730		0.471	0.596	0.824	0.579
	(.03)	(.03)	(.04)						
1997	0.800	0.887	1.037	1.504		0.404	0.483	0.605	0.824
	(.03)	(.03)	(.03)	(.04)					
log(Disposable Income) in									
1994	0.985	0.652	0.506	0.460		0.862	0.687	0.548	0.449
	(.03)	(.02)	(.02)	(.02)		(.03)			
1995	0.641	0.832	0.580	0.498		0.536	0.707	0.703	0.547
	(.02)	(.02)	(.02)	(.01)		(.02)	(.02)		
1996	0.501	0.585	0.770	0.599		0.412	0.478	0.653	0.690
	(.01)	(.02)	(.02)	(.02)		(.01)	(.01)	(.02)	
1997	0.451	0.496	0.582	0.879		0.363	0.400	0.485	0.757
	(.01)	(.01)	(.02)	(.02)		(.01)	(.01)	(.01)	(.02)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is a balanced sample of 12,558 individuals aged 20-65 with positive income in every year.

	Log(Market Income) in					log(Disposable Income) in			
	1994	1995	1996	1997		1994	1995	1996	1997
log(Market Income) in									
1994	0.758	0.794	0.723	0.657		0.910	0.717	0.646	0.585
	(.02)								
1995	0.578	0.699	0.813	0.712		0.713	0.912	0.729	0.634
	(.02)	(.02)							
1996	0.525	0.567	0.694	0.796		0.646	0.740	0.917	0.717
	(.02)	(.02)	(.02)						
1997	0.486	0.505	0.564	0.722		0.594	0.652	0.729	0.923
	(.02)	(.02)	(.02)	(.02)					
log(Disposable Income) in									
1994	0.519	0.391	0.353	0.331		0.430	0.771	0.689	0.623
	(.01)	(.01)	(.01)	(.01)		(.01)			
1995	0.394	0.482	0.389	0.350		0.319	0.399	0.793	0.686
	(.01)	(.01)	(.01)	(.01)		(.01)	(.01)		
1996	0.360	0.390	0.489	0.396		0.289	0.320	0.409	0.773
	(.01)	(.01)	(.01)	(.01)		(.01)	(.01)	(.01)	
1997	0.337	0.350	0.395	0.518		0.270	0.286	0.327	0.436
	(.01)	(.01)	(.01)	(.01)		(.01)	(.01)	(.01)	(.01)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is a bal anced sample of 13,965 individuals aged 20-65 with positive income in every year, and whose annual income is never less than 10 percent of their sample average income.

Table 6B: Covariance Structure of Logarithms of Market and Disposable Incomes, 1994-97 -- Females

	Log(Market Income) in					log(Disposable Income) in			
	1994	1995	1996	1997		1994	1995	1996	1997
log(Market Income) in									
1994	1.141	0.825	0.740	0.681		0.847	0.687	0.614	0.556
	(.03)								
1995	0.906	1.058	0.831	0.740		0.685	0.850	0.698	0.608
	(.03)	(.03)							
1996	0.789	0.854	0.997	0.820		0.598	0.687	0.851	0.673
	(.02)	(.02)	(.03)						
1997	0.738	0.771	0.830	1.028		0.553	0.610	0.692	0.858
	(.02)	(.02)	(.02)	(.03)					
log(Disposable Income) in									
1994	0.706	0.549	0.466	0.437		0.608	0.789	0.683	0.604
	(.02)	(.01)	(.01)	(.01)		(.02)			
1995	0.545	0.650	0.510	0.460		0.457	0.552	0.794	0.674
	(.01)	(.02)	(.01)	(.01)		(.01)	(.01)		
1996	0.469	0.514	0.609	0.502		0.381	0.423	0.513	0.770
	(.01)	(.01)	(.01)	(.01)		(.01)	(.01)	(.01)	
1997	0.440	0.463	0.498	0.645		0.349	0.371	0.409	0.549
	(.01)	(.01)	(.01)	(.01)		(.01)	(.01)	(.01)	(.01)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is a bal anced sample of 11,629 individuals aged 20-65 with positive income in every year, and whose annual income is never less than 10 percent of their sample average income

. Table A1: Distribution of TDCs by Type of PAYE

PAYE Type	Missing IRD Numbers	Valid IRD Numbers	
		Two-percent Sample	Combined Sample
Earnings	0.639	0.699	0.718
DSW Benefits	0.058	0.162	0.120
GRI	0.202	0.120	0.141
ACC Payments	0.101	0.020	0.021
Number of TDCs	748,239	314,482	662,579

Notes: Calculated using all Tax Deduction Certificates (TDC) for the 1994-97 tax years. Estimating the total number of TDCs by rating up the number in the two-percent sample, suggests that IRD numbers are missing on approximately 4.5 percent of TDCs.

Table A2: Market Income Quintiles, 1994-97

| | Two-percent Subsample | | |
Quintile	Positive Income Each Year	Non-positive Income Any Year	Missing Data Any Year
Males			
1	0.049	0.703	0.158
2	0.190	0.176	0.137
3	0.238	0.070	0.073
4	0.259	0.030	0.041
5	0.263	0.020	0.035
Missing Data	.	.	0.556
Sample Fractions	0.637	0.164	0.199
Females			
1	0.008	0.635	0.107
2	0.170	0.213	0.155
3	0.247	0.091	0.089
4	0.280	0.041	0.060
5	0.296	0.019	0.043
Missing Data	.	.	0.546
Sample Fractions	0.550	0.234	0.216

Notes: Sample sizes: Males = 23,367; Females = 22,891. Income Quintiles are determined on an annual basis, and aggregated over the four years 1994-97.

Table A3A: Covariance Structure of Market and Disposable Incomes, 1994-97 -- Males

	Market Income in					Disposable Income in			
	1994	**1995**	**1996**	**1997**		**1994**	**1995**	**1996**	**1997**
Market Income in									
1994	2025 (545)	0.759	0.840	0.726		0.983	0.736	0.846	0.724
1995	1417 (277)	1721 (300)	0.830	0.791		0.685	0.991	0.785	0.774
1996	1767 (410)	1610 (285)	2187 (386)	0.843		0.797	0.806	0.988	0.831
1997	1521 (303)	1528 (278)	1834 (333)	2166 (348)		0.667	0.775	0.814	0.990
Disposable Income in									
1994	1530 (499)	983 (189)	1289 (343)	1074 (220)		1197 (473)	0.668	0.820	0.673
1995	984 (188)	1222 (212)	1120 (194)	1073 (197)		687 (130)	884 (156)	0.770	0.769
1996	1323 (343)	1132 (195)	1606 (300)	1318 (239)		987 (303)	796 (135)	1209 (249)	0.816
1997	1100 (219)	1085 (197)	1313 (236)	1557 (248)		786 (165)	772 (145)	958 (176)	1141 (184)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in thousands of constant (1998) dollars. The sample is a balanced sample of 14,748 individuals aged 20-65 with positive income in every year.

Table A3B: Covariance Structure of Market and Disposable Incomes, 1994-97 -- Females

	Market Income in					Disposable Income in			
	1994	1995	1996	1997		1994	1995	1996	1997
Market Income in									
1994	571	0.694	0.634	0.560		0.972	0.678	0.624	0.532
	(147)								
1995	331	400	0.810	0.688		0.604	0.980	0.789	0.646
	(32)	(31)							
1996	304	324	402	0.751		0.550	0.786	0.982	0.703
	(34)	(27)	(34)						
1997	298	307	336	498		0.484	0.666	0.731	0.980
	(31)	(23)	(23)	(57)					
Disposable Income in									
1994	441	229	209	205		361	0.616	0.561	0.472
	(142)	(25)	(26)	(25)		(138)			
1995	225	272	218	206		162	193	0.792	0.644
	(24)	(22)	(19)	(16)		(20)	(16)		
1996	206	218	272	226		147	152	192	0.707
	(25)	(19)	(23)	(16)		(19)	(13)	(17)	
1997	207	210	229	356		146	145	159	265
	(23)	(16)	(16)	(46)		(18)	(12)	(12)	(38)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in thousands of constant (1998) dollars. The sample is a balanced sample of 12,558 individuals aged 20-65 with positive income in every year.

	Log(Market Income) in					log(Disposable Income) in			
	1994	1995	1996	1997		1994	1995	1996	1997
log(Market Income) in									
1994	1.572 (.04)	0.690	0.563	0.474		0.837	0.570	0.461	0.411
1995	0.894 (.02)	1.545 (.04)	0.712	0.579		0.573	0.848	0.572	0.482
1996	0.731 (.02)	0.909 (.02)	1.655 (.03)	0.686		0.500	0.605	0.877	0.584
1997	0.639 (.02)	0.579 (.02)	0.512 (.02)	0.462 (.04)		0.442	0.505	0.591	0.878
log(Disposable Income) in									
1994	0.873 (.02)	0.579 (.01)	0.512 (.01)	0.462 (.01)		0.830 (.02)	0.673	0.550	0.494
1995	0.577 (.01)	0.885 (.02)	0.617 (.01)	0.529 (.01)		0.508 (.01)	0.904 (.02)	0.678	0.570
1996	0.512 (.01)	0.613 (.01)	1.094 (.02)	0.647 (.01)		0.449 (.01)	0.563 (.01)	1.153 (.03)	0.678
1997	0.476 (.01)	0.539 (.01)	0.646 (.01)	1.139 (.02)		0.414 (.01)	0.486 (.01)	0.617 (.01)	1.185 (.03)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and corresponding correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is an unbalanced sample of individuals aged 20-65 with income in any year; cell sizes vary from 15,560 to 20,697.

Table A4B: Covariance Structure of Logarithms of Market and Disposable Incomes, 1994-97 -- Females

	Log(Market Income) in					log(Disposable Income) in			
	1994	1995	1996	1997		1994	1995	1996	1997
log(Market Income) in									
1994	2.236	0.730	0.598	0.497		0.765	0.529	0.428	0.361
	(.04)								
1995	1.390	2.220	0.718	0.585		0.520	0.760	0.524	0.425
	(.03)	(.04)							
1996	1.110	1.324	2.236	0.727		0.426	0.516	0.775	0.528
	(.03)	(.03)	(.04)						
1997	0.936	1.082	1.357	2.284		0.360	0.426	0.546	0.782
	(.02)	(.03)	(.03)	(.04)					
log(Disposable Income) in									
1994	1.152	0.754	0.614	0.527		1.034	0.657	0.512	0.415
	(.03)	(.02)	(.02)	(.02)		(.02)			
1995	0.731	1.112	0.719	0.605		0.581	1.014	0.658	0.519
	(.02)	(.02)	(.02)	(.01)		(.01)	(.02)		
1996	0.601	0.720	1.210	0.763		0.468	0.592	1.151	0.680
	(.01)	(.02)	(.03)	(.02)		(.01)	(.01)	(.03)	
1997	0.516	0.598	0.740	1.252		0.384	0.467	0.624	1.153
	(.01)	(.01)	(.01)	(.02)		(.01)	(.01)	(.01)	(.02)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and corresponding correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in constant (1998) dollars. The sample is an unbalanced sample of individuals aged 20-65 with income in any year; cell sizes vary from 13,512 to 20,391.

Table A4C: Covariance Structure of Market and Disposable Incomes, 1994-97 -- Males

	Market Income in					Disposable Income in			
	1994	**1995**	**1996**	**1997**		**1994**	**1995**	**1996**	**1997**
Market Income in									
1994	1787	0.763	0.819	0.716		0.981	0.738	0.821	0.697
	(410)								
1995	1278	1538	0.826	0.783		0.684	0.989	0.780	0.750
	(212)	(221)							
1996	1520	1407	1826	0.847		0.771	0.802	0.987	0.819
	(316)	(215)	(274)						
1997	1348	1363	1601	1843		0.653	0.767	0.817	0.972
	(239)	(215)	(249)	(248)					
Disposable Income in									
1994	1329	869	1086	934		1028	0.669	0.793	0.645
	(374)	(145)	(264)	(173)		(355)			
1995	870	1070	962	940		599	762	0.769	0.747
	(144)	(156)	(146)	(152)		(99)	(114)		
1996	1114	971	1316	1128		817	674	974	0.805
	(264)	(147)	(213)	(179)		(233)	(102)	(176)	
1997	957	952	1127	1306		674	667	809	979
	(173)	(152)	(177)	(176)		(130)	(111)	(132)	(134)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and corresponding correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in thousands of constant (1998) dollars. The sample is an unbalanced sample of individuals aged 20-65 with income in any year; cell sizes vary from 18,997 to 21,126.

Table A4D: Covariance Structure of Market and Disposable Incomes, 1994-97 -- Females

	Market Income in					Disposable Income in			
	1994	1995	1996	1997		1994	1995	1996	1997
Market Income in									
1994	484 (97)	0.741	0.669	0.599		0.955	0.701	0.632	0.430
1995	317 (23)	368 (20)	0.823	0.716		0.626	0.957	0.775	0.579
1996	290 (24)	310 (18)	371 (22)	0.792		0.560	0.769	0.958	0.605
1997	283 (22)	295 (16)	325 (16)	438 (36)		0.499	0.665	0.745	0.740
Disposable Income in									
1994	349 (93)	202 (17)	183 (18)	178 (18)		275 (90)	0.647	0.575	0.385
1995	198 (17)	234 (14)	192 (13)	181 (11)		138 (13)	163 (11)	0.787	0.580
1996	181 (17)	192 (13)	236 (15)	202 (11)		124 (13)	129 (9)	164 (11)	0.617
1997	180 (17)	185 (12)	205 (11)	292 (28)		122 (13)	123 (8)	138 (8)	354 (56)

Notes: Sample variances are on the leading diagonal; covariances are below the diagonal; and corresponding correlations are above the diagonal. Estimated standard errors of the variances and covariances are in parentheses below the estimates. Incomes are expressed in thousands of constant (1998) dollars. The sample is an unbalanced sample of individuals aged 20-65 with income in any year; cell sizes vary from 18,419 to 20,757.

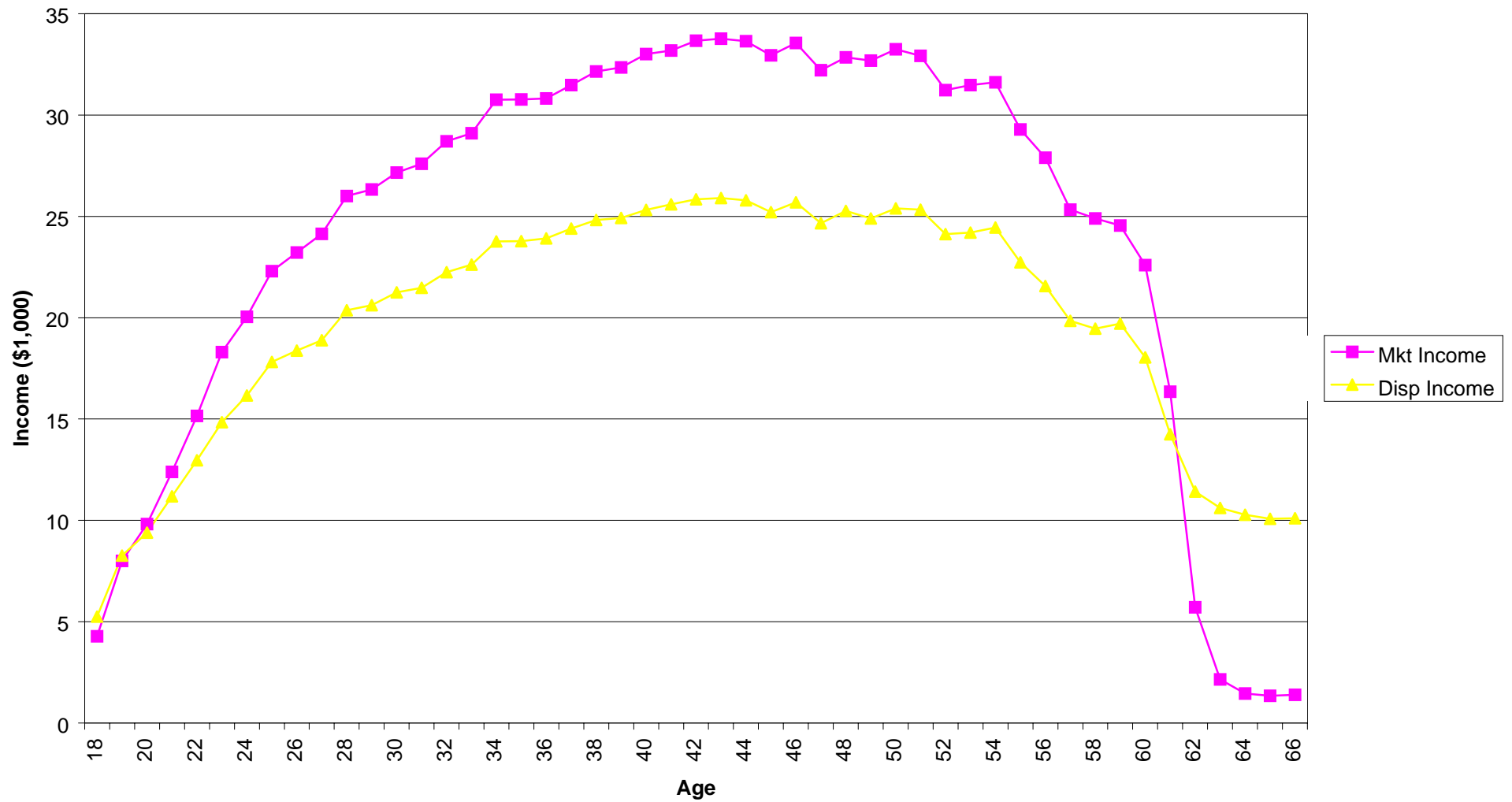**Figure 1A: Median Market and Disposable Incomes, 1994-97 -- Males**

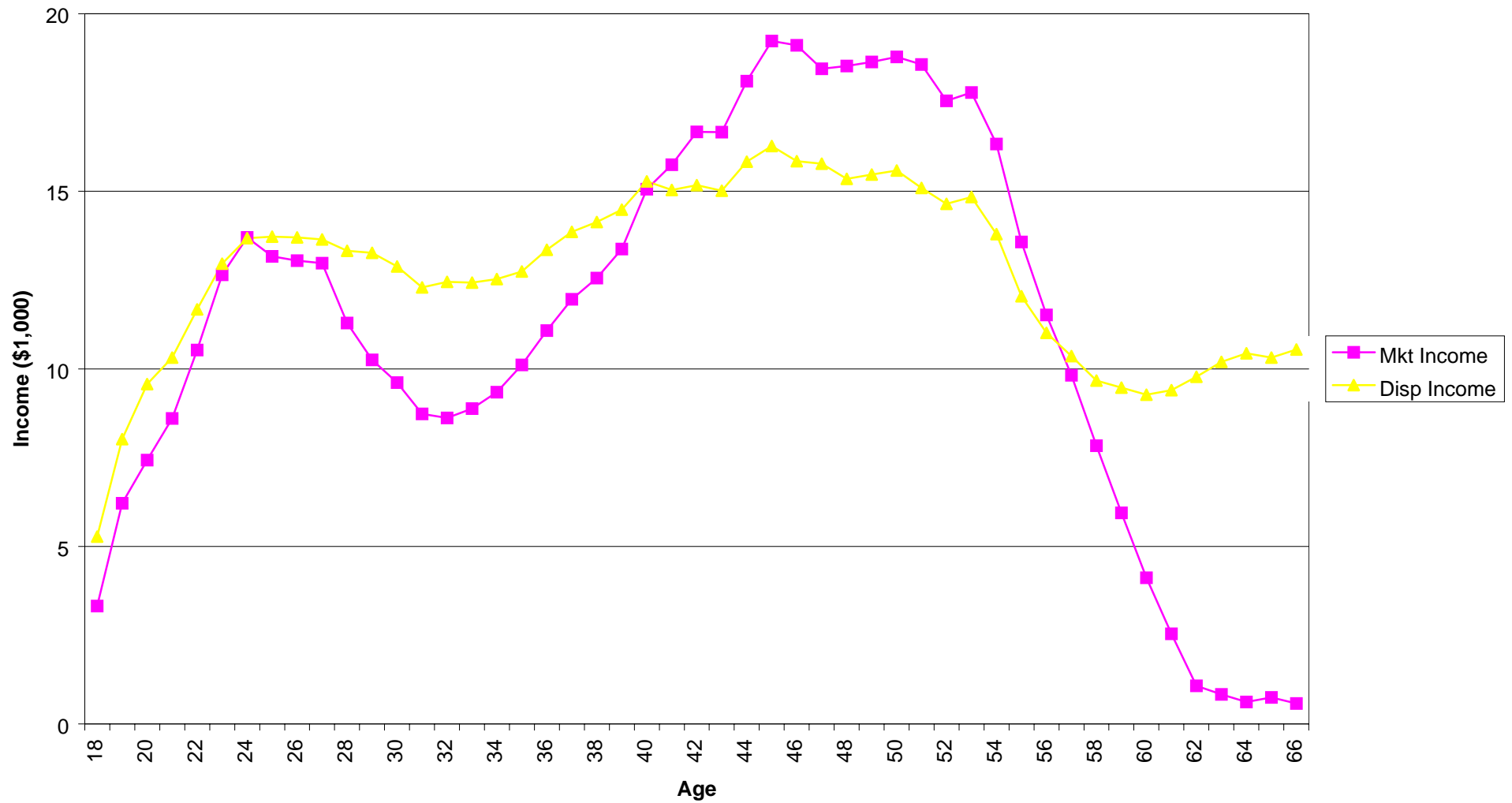**Figure 1B: Median Market and Disposable Incomes, 1994-97 -- Females**

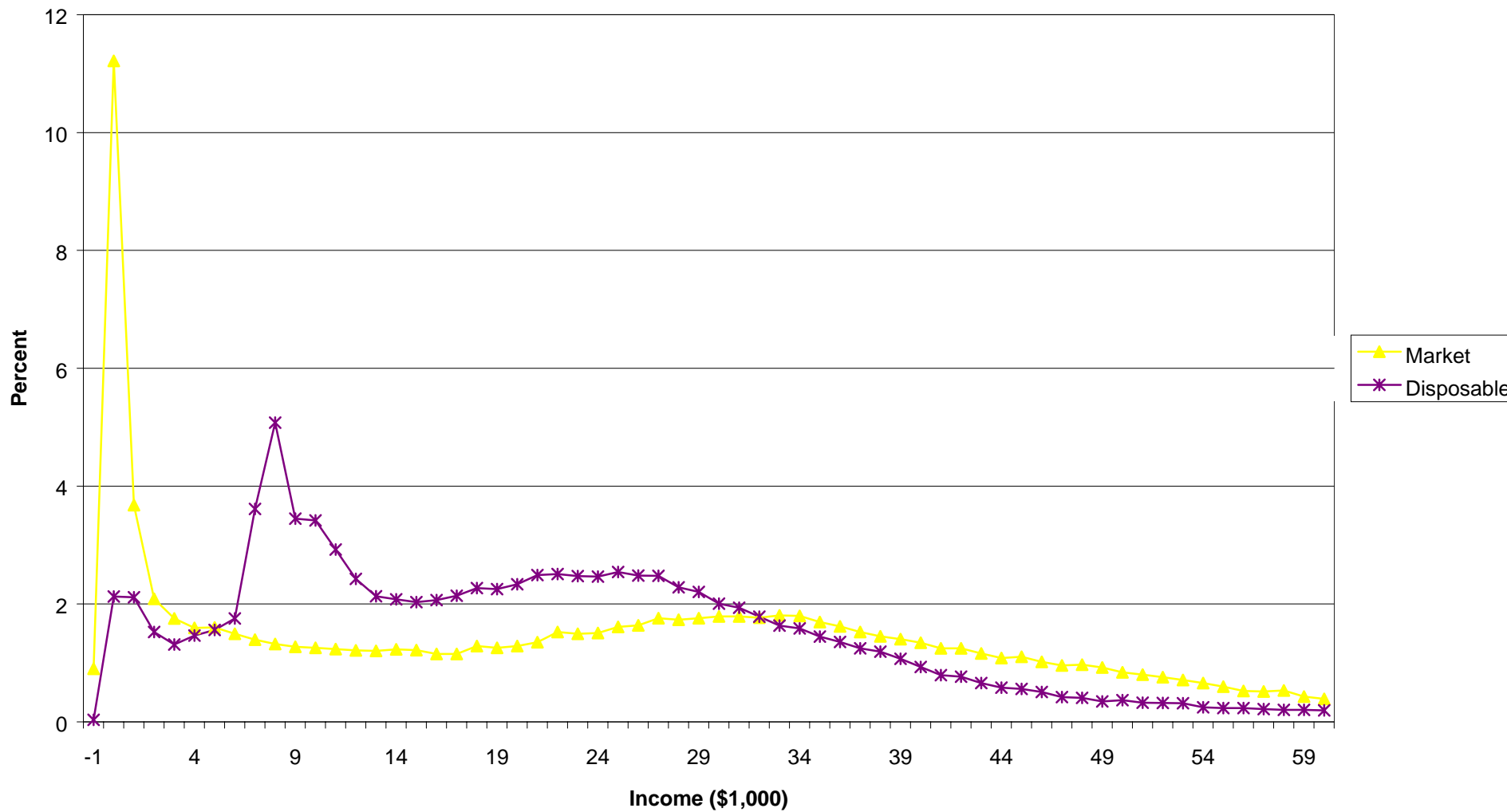**Figure 2A: Market and Disposable Income Distributions, 1994-97 -- Males**

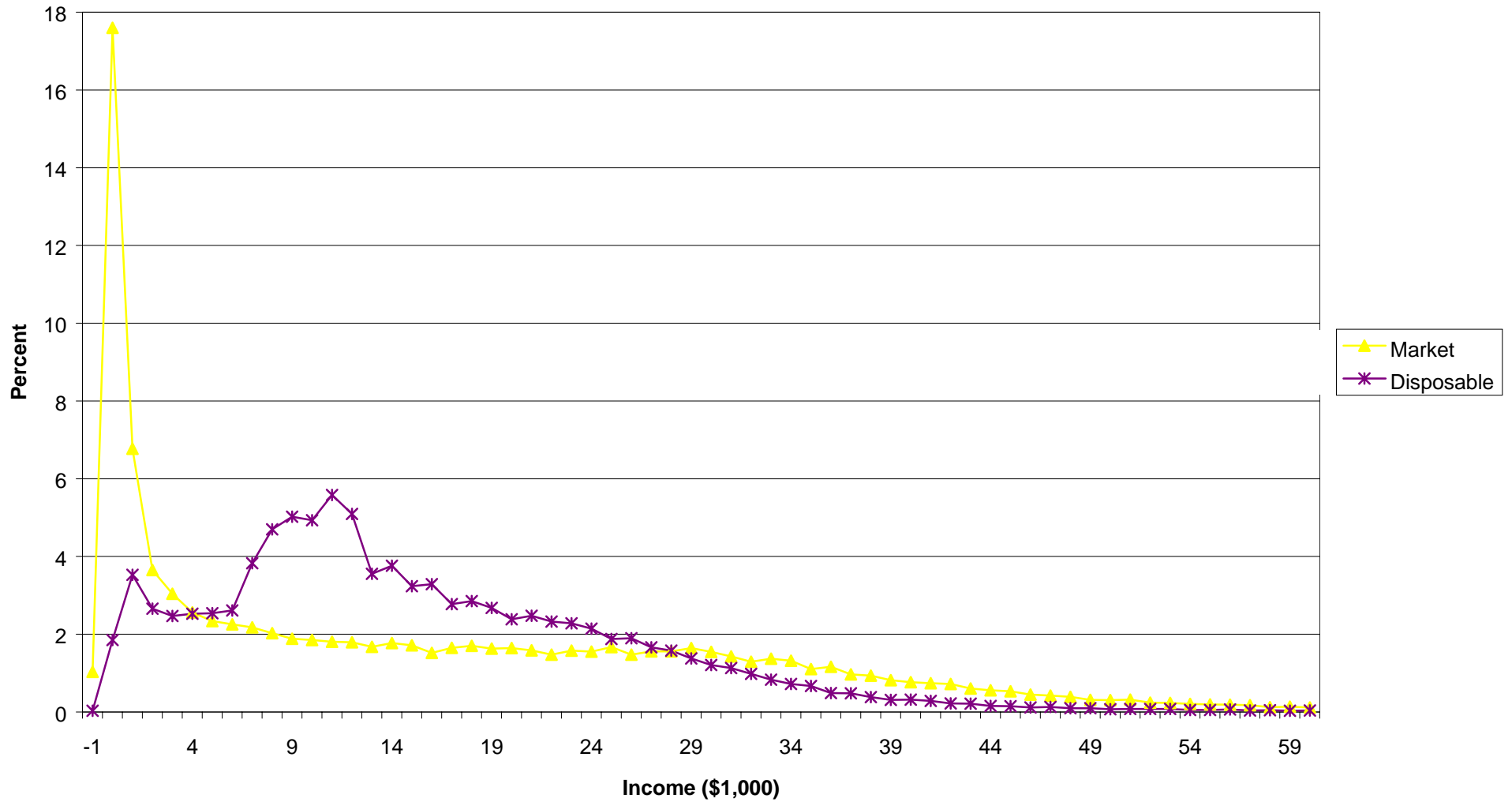**Figure 2B: Market and Disposable Income Distributions, 1994-97 -- Females**

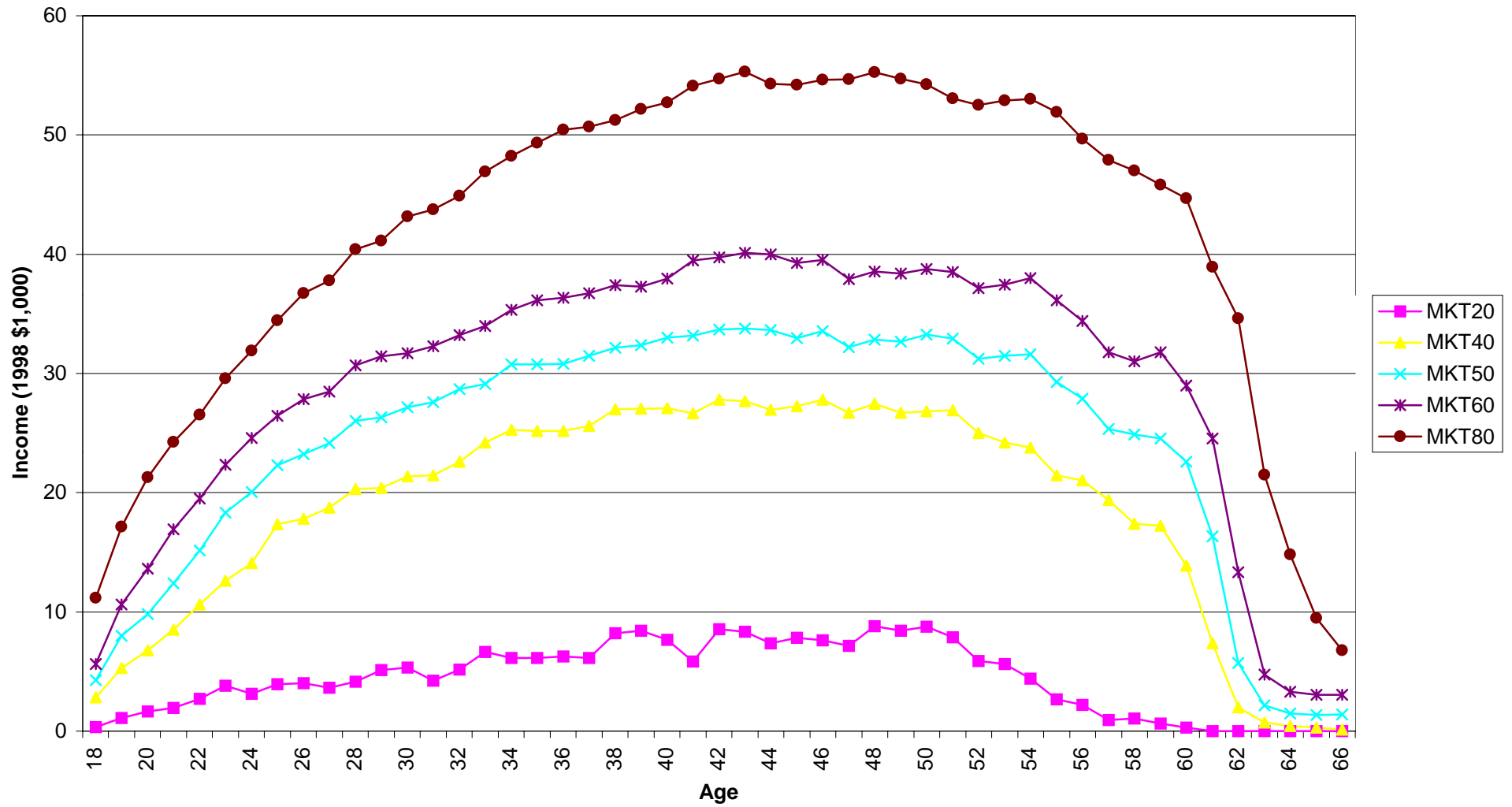# Figure 3A: Market Income Distribution Percentiles, 1994-97 -- Males

Figure 3B: Market Income Distribution Percentiles, 1994-97 -- Females