Discussion Paper Series

No.237

Volatility in Panel Data of Household Expenditure

Naohito Abe

February 2008

Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences
A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

Volatility in Panel Data of Household Expenditure[1]

Very Preliminary

Naohito Abe[2]

January 22, 2008

Abstract

The variance of growth rates of recall food expenditure is usually greater than the variance of the income growth rate. Many researchers regard the strong volatility as a symptom of measurement error. Comparing two data sets, diary data and recall data, I find that the measurement error cannot account for the observed large variance in the consumption data. Variance decomposition of the consumption and income growth rates reveals that the permanent component of the consumption variance is smaller than that of the income variance, suggesting that the consumption smoothing holds in the long run. Short run fluctuation in consumption, however, is not caused by measurement error. The implication of this finding for the modeling of household behavior is also discussed.

1. Introduction

Consumption smoothing based on the permanent income hypothesis (PIH) has been one of the basic mechanisms in modern macroeconomic theory. According to the hypothesis, fluctuations in consumption are mainly caused by permanent shocks in income, which implies that consumption variance is smaller than that of observed household income. Many empirical studies have confirmed, at least partially, the consumption smoothing hypothesis.[3] For example, household consumption variance in the System of National Accounts is smaller than the

[1] This paper has greatly profited from discussions with Yukio Abe, Noriko Inakaura, and participants in workshops at Keio University. I am grateful to Keio University and the Research Centre for Information and Statistics of Social Science(RSISS) for giving me an opportunity to use their precious survey data. Financial support was provided by the Japan Society for the Promotion of Science (JSPS).

[2] The Institute of Economic Research, Hitotsubashi University. Kunitachi, Tokyo, Japan, 186-8603. E-mail: nabe@ier.hit-u.ac.jp. Phone: +81-425-80-8347, Fax: +81-425-80-8333.

[3] See Attanasio (1999) for a survey on the consumption literature.

variance of income or investment, which is consistent with the hypothesis.[4] Based on many household level data sets, Deaton and Paxson (1994), and Blundell and Preston (1998) found that cross-sectional household consumption variance is also smaller than that of household income.[5]

When we use household-level panel data, however, the situation changes dramatically. The variance of the growth rate of food consumption in the Panel Study of Income Dynamics (PSID) is much greater than the variance of the change in household income. Zeldes (1989) reported that the standard deviation of the growth rate of food consumption in the PSID is 32%, which is extremely large. Strong volatilities in consumption panel data were also reported by Mork and Smith (1989). Using Norwegian household panel data, Mork and Smith found that the variance of consumption change is five times larger than that of the income change. Most previous literature regards the large volatility in household level consumption as a symptom of measurement errors. Hall and Mishkin (1982) wrote:

> "Any study of consumption at the level of individual households needs to include a stochastic element of measurement error and transitory consumption. (Hall and Mishkin (1982), p. 469)"

Zeldes (1989) also wrote:

> "(the standard deviation of the consumption growth) seems implausibly high to be explained by differences in expected growth, expectation errors, or changes in tastes. (Zeldes (1989), p.326)"

Rather than building a behavior model to explain the large consumption variance, most previous studies such as Altonji and Siow (1987) treat the variance as a reflection of measurement error in consumption data and conduct their estimation by imposing certain restrictions on the measurement error.

This paper investigates whether or not the observed large variance in consumption is caused by measurement error. For this purpose, I compare two data sets. One is a household panel survey that contains recall consumption data. The other is a cross-sectional household survey that contains diary records data of household consumption, which is supposed to be very

[4] In the US, the standard deviation of aggregate consumption is about 70% of that of GDP (Kydland and Prescott (1990)). In Japan, the figure is about 80% (Abe (2004)).
[5] See Ohtake and Saito (1998) and Storesletten, Telmar, and Yarron (2004) for recent studies.

accurate. Although the two data sets did not survey the same households, both of them contain a detailed breakdown of consumption such as clothes, medical, transportation, communication, as well as food and income, which gives us an opportunity to consider the measurement issues of not only food consumption, but also other items and income. The age-consumption variance profiles of the two data sets are surprisingly similar to each other in their shape and level. I also found no evidence of a negative correlation between the measurement error and the true consumption or other variables such as household size, which implies that the recall consumption data in the household panel is not seriously contaminated by measurement error. The variances of consumption and income levels are also very similar in the two data sets. The variance of consumption growth, however, is about four times greater than income growth variance, which is inconsistent with consumption smoothing.

Decompositions of the consumption variance into (1) household fixed effects, (2) permanent shocks, (3) transitory shocks, and (4) covariance among fixed effects and other shocks, reveal that half of the variation of consumption is because of household fixed effects, and most of the remaining variation is caused by transitory shocks. Consumption growth exhibits very strong negative autocorrelation, which suggests that after controlling for fixed effects, household consumption is close to a process which is independent and identically distributed (i.i.d.), not a random walk. On the other hand, our data also show that both consumption and income variances in panel data are increasing with age, which is consistent with the life-cycle-permanent income model by Deaton and Paxson (1994). The permanent component of consumption variance is smaller than that of income, which suggests that the permanent income hypothesis is working, at least, in the long run.

The next section of this paper briefly reviews previous studies on measurement issues in recall data. Detailed description of the data is given in Section 3. Comparisons between the two data sets are discussed in Section 4. Section 5 investigates the correlation between the measurement error in food data and household size pointed out by Gibson (2002). Decomposition of the variation in consumption and income into permanent shocks, transitive shocks, and fixed effects through the minimum distant estimator is provided in Section 6. The final section concludes.

2. Measurement Errors in Recall Consumption Data

Many household panel surveys such as the PSID and the British Household Panel Survey (BHPS) contain recall food data, not diary data. Panel data based on diary such as the Consumer Expenditure Survey (CEX) in the United States and the Family Income and Expenditure Survey (FIES) in Japan trace the same household for no longer than one year, which makes it very difficult to identify the effects of seasonal fluctuation. Because the consumption Euler equation has been the basis of empirical studies on household consumption requiring panel data, most research relies on recall food data although they note the large volatility in consumption data. Usually, researchers using household consumption assume that the recall consumption data contain multiplicative measurement error which is not correlated with true consumption, and they log-linearize the Euler equation and put the measurement error into the error term. Recently, such strategies have been criticized. Carroll (2001) and Ludvigson and Paxson (2001) pointed out that by log-linearizing the Euler equation, a significant amount of additional bias is introduced. Recently, an increasing number of consumption researchers avoid using the Euler equation and rely on other aspects such as the consumption-age profile (Gourinchas and Parker (2002)), or asset-age profile (Cagetti (2003)) for their structural estimation so that they do not have to use recall consumption panel data.

There have been several attempts to estimate the degree of measurement error and their correlation with economic variables by directly comparing diary data with recall data. Gibson (2002) used a survey from Papua New Guinea in which a half of the families was asked to recall food consumption and the other half was asked to keep a diary. Gibson found that the recall food data had a larger negative correlation with the family size than the diary data probably because of the negative correlation between the measurement error and family size. In a recent paper, Ahmed, Brzozowski, and Crossley (2006) estimated the degree of the measurement error in recall data by using a unique data set that collects recall and diary food data from the same households in Canada. They found that the measurement error in recall food data is negatively correlated with the diary food data.[6]

3. Data

This paper uses two household level data sets. The first data come from the Keio Household Panel Survey (KHPS). The second data set is the National Survey of Family Income and Expenditure (NSFIE).

3.1. KHPS

The KHPS started collecting detailed information in 2004 on a sample of 4005 Japanese households. Since then, the survey has been conducted annually each January by Keio University. In this paper, I use four waves between 2004 and 2007. The KHPS contains a number of detailed questions about education, job history of wife and husband, as well as family variables such as consumption, real estates, financial assets, etc. Although the KHPS is an

[6] Rodgers, Brown, and Duncan (1993) also found a negative correlation between the true earnings and the earnings reported in the PSID.

annual panel survey, the KHPS contains much more detailed questions on expenditure than other panel studies such as the PSID. The KHPS asks the amount of monthly expenditure in the previous December on (1) food, (2) dining out, (3) housing including rents & repairs, (4) fuel, lights & water, (5) furniture and household utensils, (6) clothes & footwear, (7) transportation, (8) communication, (9) education, (10) reading & recreation, (11) pocket money & social expenses, (12) transfers, and (13) other expenses. Surveyed households are selected by a two-stage sampling method and asked to fill out the questionnaires within one week and return them to enumerators.

Compared with other panel surveys such as the PSID, the KHPS contains far more number of questions. The questionnaire used in 2004 had 67 pages, which is twice as large as the PSID at the beginning of the year. Unfortunately, the costs of asking so many questions appeared in the response rate and the attrition rate. The initial cooperation rate was about 30% in 2004, that is, to obtain 4000 samples, more than 13,000 households are asked to cooperate. The sample size decreased to 3314 in 2005, 2887 in 2006, and 2643 in 2007. The annual attrition rate was about 13%, which is significantly larger than that of the PSID, which was 2-3%.

3.2. NSFIE

The NSFIE has been conducted by the Statistical Bureau in Japan every five years since 1959. In order to compare the KHPS, I use the survey conducted in 2004 for this project. The survey asks various questions such as family composition, annual income in the previous calendar year, and the three-month average expenditure between September and November. Survey households are selected by a two-stage sampling method. About 8000 enumerators visit survey households, give diary books to record all their daily expenditure and revenue for three months, provide instruction on how to keep the diary, and collect them every month. The total sample size of the NSFIE in 2004 is about 54,000 for two or more person households, making it one of the largest household surveys in the world with detailed information on income and consumption. Because the NSFIE is conducted by the Japanese government under the Statistical Law, which makes it obligatory for Japanese citizen to cooperate, the response rate is expected to be high.[7] Because of the human resource costs involved in the survey, the NSFIE is expected to be quite accurate.

The details regarding the survey, such as the sampling procedures and definitions of terms, are described in volumes published for each survey by the Statistics Bureau, Ministry of Public Management, Home Affairs, Posts and Telecommunications. The information is available in both Japanese and English.[8] In this project, 80% of the NSFIE is used.[9]

---

[7] Unfortunately, the response rate of the NSFIE is not available.

[8] Hayashi, Ando, and Ferris (1988) described in detail the features of the NSFIE that make the

4. Comparisons

We select households with male heads, aged 24 to 65, married, and working as employees. Because we compare the variances and standard deviations as well as the means of the two data sets, we exclude households whose consumption or income is below the lowest 1% or greater than the highest 1% in the NSFIE.[10] We also exclude workers in the agricultural sector, self-employed, and unemployed. The selection process produces samples of 1392 families in the KHPS, and 26,749 households in the NSFIE.[11]

4.1. General Comparisons

Table 1 shows the means and the standard deviations of several key variables in the two data sets. As the earnings variable is the earnings of the male head. All figures are in natural logarithms of the expense or income in 1000 yen. By comparing columns (1) to (3), and columns (2) to (4), we can observe that the two data sets are surprisingly similar to each other despite the differences in their methods to record the expenditure. If the KHPS is contaminated by classical measurement error, which is orthogonal to everything in the data set, and if the diary data in the NSFIE are correct, the standard deviations in the KHPS survey must be larger than those in the NSFIE. Table 1 suggests that the measurement error contained in the KHPS is either minor or not the classical one.

The largest differences in the standard deviations and the means in columns (2) and (4) are for Rent & Repair. The average of Rent & Repair in the KHPS is much larger than that of the NSFIE. The difference in the ratio of house ownership between the two data sets might be the reason for the difference. About 77% of the households in the NSFIE own their house, while only 68% of the KHPS own their house. We can also observe a larger standard deviation in the anural earnings in the KHPS than in the NSFIE. The inaccurate earnings records in the KHPS are probably because of the timing of the survey. The KHPS asks for the annual income in the previous calendar year every January, well before the deadline for submitting the tax form to the National Tax Agency, while the NSFIE asks the same question in September, after the deadline.

---

survey suitable for analysis of consumption and saving behaviors.

9 The 80% sample of the NSFIE is provided by the Research Centre for Information and Statistics of Social Science (RCISSS), Institute of Economic Research, Hitotsubashi University. The RCISSS provides academic researchers with many governmental micro data by the cooperation from the Statistical Bureau in Japan.

10 In the NSFIE, the lowest 1% of incomes are below 960, 000 yen, while the highest 1% are above 13,440,000 yen. I apply the same lower and upper bands for the KHPS, so that the minimum and maximum in both samples are identical. The selection of the upper and lower bands is not critical in this study. Applying other criteria does not significantly change the results.

11 In the Appendix, I also show a comparison with the monthly household survey.

Table 1 also reports the means and the standard deviations of the first differenced variables and their autocorrelations. The standard deviations of the first differenced expenditures reported in column (6) are very similar to the standard deviation of the expenditure level reported in column (4). On the other hand, the standard deviation of the earnings growth is much lower than that of the earnings level. As reported in column (7), consumption growth exhibits stronger autocorrelation than earnings growth. These characteristics are difficult to reconcile with the consumption smoothing and random walk hypotheses.

Following many previous studies such as Hall and Mishkin (1982), suppose that household expenditure can be described by the following process:

$$E_{it} = E^*_{it} + \omega_{it}\,, \tag{1}$$

$$E^*_{it} = E^*_{it-1} + \varepsilon_{it}\,, \tag{2}$$

where E_{it} is observed expenditure by household i at time t, E^*_{it} is the true expenditure, ω_{it} is an i.i.d. measurement error, and ε_{it} : an i.i.d. permanent shock in expenditure.

If the diary data are accurate, we can regard the expenditure data in the NSFIE as E^*_{it}. The fact that the variances of E^*_{it} and E_{it} are almost identical suggests that the variance of ω_{it} is very small, or ω_{it} has a strong negative correlation with E^*_{it} so that Cov(E^*_{it}, ω_{it}) +Var(ω_{it}) = 0.

Let us suppose that the measurement error is correlated with the current permanent shock to the true expenditure, but not with the past values, that is, let us assume Cov($\varepsilon_{it}, \omega_{it}$)$\neq 0$, Cov($\varepsilon_{it}, \omega_{it-1}$) = 0, and Cov($E^*_{it-1}, \omega_{it}$) = 0.

Then, simple calculation leads us to:

Var(E_{it}) = Var(E^*_{it-1}) + Cov($\varepsilon_{it}, \omega_{it}$) + Var($\omega_{it}$) + Var($\varepsilon_{it}$),

= Var(E^*_{it}) + Cov($\varepsilon_{it}, \omega_{it}$) + Var($\omega_{it}$) . (3)

The fact that Var(E_{it}) $\approx$ Var(E^*_{it}) suggests that:

$$Cov(\varepsilon_{it}, \omega_{it}) + Var(\omega_{it}) \approx 0. \tag{4}$$

By taking the first difference of (1) and using (2), we can obtain:

$$\Delta E_{it} = \varepsilon_{it} + \omega_{it} - \omega_{it-1}. \tag{5}$$

Therefore,

$$\begin{aligned} Cov(\Delta E_{it}, \Delta E_{it-1}) &= -Var(\omega_{it-1}) - Cov(\varepsilon_{it-1}, \omega_{it-1}) \\ &\approx 0. \end{aligned} \tag{6}$$

Equation (6) is inconsistent with the large autocorrelation in the expenditure growth reported in column (7) in Table 1.[12] That is, both hypotheses: (A): small Var(ω_{it}) and Cov($\varepsilon_{it}, \omega_{it}$) = 0, and (B): large Var(ω_{it}) and Cov($\varepsilon_{it}, \omega_{it}$) $\neq 0$, are inconsistent with the observed large autocorrelation in Table 1. Therefore, we cannot adopt simple explanations such as small measurement error or negative correlation between the measurement error and the true consumption to reconcile between our data sets and consumption smoothing.

4.2. Age Profiles

In this section, I compare the two data sets in detail. Figures 1, 2, and 3 plot the age profiles of average income, total expenditure, and food expenditure of the NSFIE and the KHPS, respectively. As is often pointed out, the average income-age profile is hump-shaped. The profile of the KHPS is very close to that of the NSFIE. The age-expenditure profiles are also hump-shaped as is reported in many previous studies including Attanasio, et al. (1999) and Gourinchas and Parker (2002). Although the age and food expenditure profiles in the KHPS are generally smaller than the profiles in the NSFIE, both profiles in the KHPS and the NSFIE are increasing with age. Even though the average profiles in the two data sets are not identical, we can say that the average profiles or the two data sets share the general characteristics.

Figures 4 and 5 show the age profiles of income and expenditure variances, respectively. Since the seminal works by Deaton and Paxson (1994) and Blundell and Preston (1998), the age-variance profiles of income and consumption have been used intensively to test the life cycle

[12] Using longer household panel data on Japanese households, Abe and Inakura (2007a) found that the correlation between the expenditure growth and its lag disappears quickly as time passes. That is, Corr($\Delta E_{it}, \Delta E_{it-i}$) for i > 1 is almost zero, implying that strong autocorrelation in the measurement error are unlikely.

model of consumption. In Figures 4 and 5, the lines marked with triangles plot the age-variance profiles obtained from the KHPS in year 2004, while the lines marked with quadrangles show the variance profiles obtained from the KHPS between 2004 and 2007.[13] As we can guess from columns (2) and (3) in Table 1, the age-income variance profile in the KHPS is greater than that in the NSFIE although both profiles share the same shape. The three age-consumption variance profiles drawn in Figure 5 are very similar to each other. Until the mid 40s, the consumption variance is constant. In the mid 40s, the variance begins to increase. This age-consumption variance pattern in Japan is also pointed out by Ohtake and Saito (1998) and Abe and Yamada (2006).

In Figure 6, I draw several different age-variance profiles based on the KHPS data simultaneously. The broken lines plot the variance profiles of income growth (marked by quadrangles) and consumption growth (marked by triangles), while the normal lines plot the variance of the levels. The consumption variance is smaller than the income variance except for very young households, which is consistent with consumption smoothing. An increasing trend in the age-income variance suggests that the income process contains significant amount of the permanent shocks, which is also consistent with many previous studies such as Blundell and Preston (1998). The age-variance profile of income growth is much smaller than the age-variance profile of the income level. We can also observe an increase in the income growth variance in the late 40's, which suggests that the permanent shocks to income might depend on age.[14] The age-consumption GROWTH variance profile is very close to the age-consumption LEVEL variance profile, while the income growth variance is much smaller than the income level variance.

The age profiles of the covariance and the correlation coefficient between income and consumption are reported in Figures 7 and 8, respectively. The broken line plots the age profile based on the KHPS, while the normal line plots the profile based on the NSFIE. The age-covariance profiles in Figure 7 based on the two surveys are quite similar to each other. Both are increasing with age, which suggests that the income process contains a permanent component and that consumption moves with the shocks. The profile based on the KHPS exhibits larger fluctuations probably because the sample size of the KHPS is much smaller than the NSFIE. The age-correlation coefficient profile based on the KHPS is generally equal or smaller than the profile based on the NSFIE. The departure between them is greater for households whose head is between 35 and 47 years old. Probably, the departure is caused by the large income variance in the KHPS compared with the NSFIE for such households as we can observe in Figure 4.

[13] The KHPS in 2004 contains annual income in 2003.
[14] Based on the NSFIE in 1984, 1989, 1994, and 1999, Abe and Yamada (2006) found that size of permanent income shocks depend on age, which can explain the unique shape of the Japanese age-consumption variance profile.

5. Measurement Errors in Food Data

In a recent paper, Gibson (2002) found that recall food data have a significant amount of measurement error that is correlated with household size. Gibson (2002) compared two data sets, one with recall food data and the other with diary food data. Both data were obtained in Papua New Guinea. He examined the relationship between the food share and the family size of the two data sets and found that the recall food share had a larger negative correlation with the household size than the diary food share, which suggests that the measurement error in recall food data has a negative correlation with the family size.

With the KHPS and the NSFIE, I regressed ln(Food expense per capita) on ln(Household size) and ln(Total expense per capita). The results are summarized in Table 2. Because the age-food profiles in Figure 3 have different shapes from the age-total consumption profiles in Figure 2, I also report the results with age effects in columns (2) and (4) in Table 2. According to Gibson (2002), the measurement error has a negative correlation with the household size, suggesting that the coefficient for the ln(Household size) is much smaller in the KHPS sample than in the NSFIE sample.[15] In contrast to Gibson (2002), however, the estimated coefficients are almost identical between the KHPS and the NSFIE. Table 3 reports several other estimation results following Gibson (2002) and Deaton and Paxson (1998).[16] Table 3 shows no changes in the basic result. The coefficients for ln (Household size) in the KHPS and in the NSFIE are very close to each other, which implies that the measurement error in the KHPS is not a serious issue as long as the NSFIE is accurate.

6. The Covariance Structure and Estimations of Income and Consumption Processes

In this section, I decompose the income and consumption variances into (1) household fixed effects, (2) permanent shocks, (3) transitive shocks, and (4) covariance between the fixed effects, permanent shocks, and transitive shocks. To eliminate the effects of observable components, first, I regress the natural logarithms of the earnings by the male head and the log of household total expenditure on dummy variables for the year, the number of children, the number of family members, the final educational attainment by the household head, the size index of city, and the city area.[17] Hereafter, the residuals of the regressions are used as the income and consumption data.

[15] Gibson's estimates of the coefficient are -0.0026 for diary sample, and -0.1262 for recall sample.
[16] Unfortunately, the NSFIE does not contain information on education background.
[17] I eliminate households whose annual working time is shorter than 1000 hours.

Following previous literature such as Hall and Mishkin (1982) and Abowd and Card (1989), I assume following income and consumption processes:

$$y_{it} = y^*_{it} + x_i + \omega_{it}, \tag{7}$$

$$y^*_{it} = y^*_{it-1} + \varepsilon_{it}, \tag{8}$$

where y_{it} is log income or log consumption, y^*_{it} is the permanent component of income or consumption of household i at age t, x_i is the household fixed effects, ω_{it}, and ε_{it} are i.i.d. shocks. I estimate the income process and the consumption process separately. Although I assume both ω_{it} and ε_{it} are i.i.d. and independent from each other, I allow the fixed effects to be correlated with the shocks.[18]

Because the above consumption and income process are non-stationary, in order to calculate the variances, I assume that households enter the economy when they are 25 years old with initial endowments. Then, we can calculate the variance of the permanent component, y^*_{it} as follow:

$$Var\left(y^*_{it}\right) = \sum_{s=1}^{t} Var\left(\varepsilon_{is}\right) + Var\left(y^*_{i0}\right) = tVar\left(\varepsilon_{is}\right) + Var\left(y^*_{i0}\right). \tag{9}$$

Because we cannot identify x_i from y^*_{i0}, I assume $y^*_{i0} = 0$ for all i.

The differenced series of y does not contain the fixed effects and the covariance terms. Therefore, we can obtain:

$$\Delta y_{it} = \varepsilon_{it} + \omega_{it} - \omega_{it-1}. \tag{10}$$

After tedious calculation, we can obtain the following moment condition:

$$Var\left(y_{it}\right) = Var\left(\omega_{it}\right) + tVar\left(\varepsilon_{it}\right) + Var\left(x_i\right) + Cov\left(x_i, \omega_{it}\right) + tCov\left(x_i, \varepsilon_{it}\right), \tag{11}$$

$$Var\left(\Delta y_{it}\right) = Var\left(\varepsilon_{it}\right) + 2Var\left(\omega_{it}\right), \tag{12}$$

[18] Many previous studies assume that the transitory shocks, or the measurement error are serially correlated. I do not consider serial correlation in the transitory shocks for two reasons. The first is the lack of time series length in the KHPS. The KHPS contains only four years. The second reason is that in Abe and Inakura (2007b), I found that the correlation between the expenditure growth and its two year lag is almost zero, which implies that serial correlation is weak.

$$Cov(\Delta y_{it}, \Delta y_{it-1}) = -Var(\omega_{it}), \tag{13}$$

$$Cov(y_{it}, y_{it-1}) = (t-1)Var(\varepsilon_{it}) + Var(x_i) + 2Cov(x_i, \omega_{it}) + 2(t-1)Cov(x_i, \varepsilon_{it}), \tag{14}$$

$$Cov(y_{it}, \Delta y_{it}) = Var(\varepsilon_{it}) + Var(\omega_i) + Cov(x_i, \varepsilon_{it}). \tag{15}$$

To estimate $Var(\varepsilon_{it})$, $Var(\omega_{it})$, $Var(x_{it})$, $Cov(x_i, \omega_{it})$, and $Cov(x_i, \varepsilon_{it})$, I adopt the technique developed by Abowd and Card (1989). First, I calculate several covariance matrices, and obtain the empirical moments appearing on the left hand side in equations (11) to (15). Then, equal weighted minimum distant etimation (EQWMD) is conducted to obtain the estimates of $Var(\varepsilon_{it})$, $Var(\omega_{it})$, $Var(x_{it})$, $Cov(x_i, \omega_{it})$, and $Cov(x_i, \varepsilon_{it})$.[19]

Because the left hand side in equations (12) and (14) consist of the growth rate of income or consumption only, we can estimate the parameters appearing in equations (12) and (14) without using the other moment conditions. Therefore, I also conduct EQWMD for (12) and (14).

Table 4 reports the results. From column (1), we can observe that major part of the cross sectional variances in household income are the household fixed effects. The variance of the transitory shocks to income is three times larger than the variance of the permanent shock. The covariance between the fixed effects and permanent transitory shocks are statistically significant. This result suggests that when we use cross-sectional data, we have to deal with the household level heterogeneity very carefully. The estimated variance and covariance reported in column (2) is very different from those in column (1). About half of the cross-sectional variation in expenditure can be explained by the fixed effects, while the most part of the remaining variation are because of the transitive shocks. The variations coming from the permanent shocks are very small, although they are statistically significant. The results shown in columns (3) and (4) confirm the relative importance between the permanent and transitive shocks in both income and consumption.

Based on the result in column (2) in Table 4, we can interpret the characteristics of the consumption variance shown in Figure 6, similar variance sizes in the consumption level and its growth rate, as follows. Half of the cross-sectional variation in the consumption level comes from the household fixed effects, which disappear when we take the first difference of consumption.

[19] I do not use the optimal weight because the variance-covariance matrix used in the estimation contains many missing elements. For example, the covariance between income at age 25 and age 30 is not available because the KHPS covers only four waves. The variance-covariance matrix must be of full rank to get the optimal weight, which is not the case in this paper. Altonji and Segal (1996) found that EQWMD gives more accurate estimates than the minimum distant estimator with the optimal weight when the sample size is not large.

The variance of the expenditure growth is $Var(\varepsilon_{it}) + 2Var(\omega_{it})$. Because the variance of the permanent shock, $Var(\varepsilon_{it})$, is very small, and the variance of the transitive shock, $Var(\omega_{it})$, is about the same size as the variance of the fixed effects, $Var(x_{it})$, the variance of the consumption level is very similar to the variance of consumption growth.

Finally, when we compare the permanent shocks to income with that to expenditure, the variance of the permanent shocks to expenditure is smaller than that to the income, suggesting that consumption is smoother than income in the long run. The PIH claims that the consumption variation should be equal to the variation in the permanent income. This claim is obviously inconsistent with the empirical results in this paper. However, the fact that the permanent variation in consumption is smaller than that of income implies that the modified consumption model which allows for temporary movements in expenditure might be able to explain the movement of the household level panel data.

7. Conclusion

In this paper, I have compared two household surveys, one with diary expenditure data, and one with recall expenditure data. The two data sets are very similar in terms of both first and second moments. Because the diary data are supposed to be very accurate, the recall data are unlikely to be contaminated by measurement error. I have shown that expenditure growth is as volatile as the expenditure level, while income growth is much more stable than the income level. This result is inconsistent with standard consumption models such as the permanent income-life cycle model. Decompositions of consumption and income variances into several components reveal that the permanent shock to consumption is smaller than that to income, which suggests that consumption smoothing might occur in the long run. The most variation in expenditure growth comes from the transitive shocks that are not caused by measurement error.

By introducing taste shocks to the standard permanent income model, we could build a model that is consistent with the findings in this paper. However, I suspect that the degree of the taste shocks has to be extremely large to replicate the observed volatility in expenditure. In addition, if uncertainty in preferences is the main cause of the volatility in expenditure, under incomplete capital markets, the household will have a strong precautionary savings motive than the existing buffer-stock saving model, because consumption smoothing is not occurring in the short run. To build a model that can replicate the high volatility in expenditure in the short run and smoothing in the long run will be the next task.

References

Abe, N. (2004) "Multi-sector Business Cycle Model and Aggregate Shocks: An Empirical Analysis", *The Japanese Economic Review*, Vol. 55, No. 1, March. pp. 101-118.

Abe, N. and N. Inakura (2007a) "The Covariance Structure of Earning and Hours Changes - The Case of Japanese Households (in Japanese)", *The Economic Review*, 58(1), pp. 15-30.

Abe, N. and N. Inakura (2007b) "Age Profiles of Household Consumption, Labor Hours, and Asset Accumulation in Japan (in Japanese)", *Understanding Inflation Dynamics of the Japanese Economy Working Paper Series*, No. 7.

Abe, N. and T. Yamada (2006) "Nonlinear Income Variance Profile and Consumption Inequality over the Life Cycle", *Bank of Japan Working Paper Series*, No. 06-E-01.

Abowd, J. M. and D. Card (1989) "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, Vol.57, No.2, pp. 411-445.

Ahmed, N. M, Brzozowski and T. Crossley (2006) "Measurement Errors in Recall Food Consumption Data", *IFS Working Papers*, W06/21, Institute for Fiscal Studies.

Altonji, J.G. and L.M. Segal (1996) "Small-Sample bias in GMM Estimation of Covariance Strucutres", *Journal of Business and Economic Statistics*, 14, pp. 353-66.

Altonji, J.G. and A. Siow (1987) "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data", *The Quarterly Journal of Economics*, Vol. 102(2), pp. 293-328.

Attanasio, O.P. (1999) "Consumption," in J.B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, Vol. 1, North-Holland.

Blundell, R.W. and I. P. Preston (1998) "Consumption Inequality and Income Uncertainty", *Quarterly Journal of Economics*,113, pp. 603-640.

Cagetti, M. (2003) "Wealth Accumulation over the Life-Cycle and Precautionary Savings", *Journal of Business and Economic Statistics*,21, pp. 339-353.

Carroll, C. D.(2001) "Death to the Log-Linearized Consumption Euler Equation!(And Very Poor Health to the Second-Order Approximation)", mimeo

Deaton, A. and C. Paxson (1998) "Economies of Scale, Household Size, and the Demand for Food", *Journal of Political Economy*, 106 (5), pp. 897-930.

Deaton, A. and C. Paxson (1994) "Intertemporal Choice and Inequality", *Journal of Political Economy*, Vol.102, pp. 437-467.

Gibson, J. (2002) "Why does the Engel Method work? Food demand, economies of size and household survey methods", *Oxford Bulletin of Economics and Statistics*, 64(4), pp 341-359

Gourinchas, P. O. and J. Parker (2002) "Consumption over the Life Cycle", *Econometrica*, Vol. 70, No.1, pp. 47-89

Hall, R. and F. Mishikin (1982) “The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households”, *Econometrica*, 50. pp. 461-82.

Hayashi, F., A. Albert and R. Ferris (1988) “Life Cycle and Bequest Savings: A Study of Japanese and U.S. Households Based on Data from the 1984 NSFIE and the 1983 Survey of Consumer Finances,” *Journal of the Japanese and International Economies*, Vol. 2, No. 4, December, pp. 450-491.

Kydland, F.E. and E.C. Prescott (1990) “Business Cycles: Real Facts and a Monetary Myth.” *Federal Reserve Bank of Minneapolis, Quarterly Review,* Spring, pp 3-18.

Ludvigson, S. and C.H. Paxson (2001) “Approximation Bias in Linearized Euler Equations, ” *The Review of Economics and Statistics*, MIT Press, Vol. 83(2), pp. 242-256, May.

Mork, K.A. and V .K. Smith (1989) “Testing the Life-Cycle Hypothesis with a Norwegian Household Panel”, *Journal of Business and Economic Statistics*, July, Vol.7, No.3.

Ohtake, F. and M. Saito (1998) “Population Aging and Consumption Inequality in Japan”, *Review of Income and Wealth*, Vol.44, pp.361-381.

Rodgers, W.L., C. Brown, and G. J. Duncan (1993) “Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages”, *Journal of the American Statistical Association*, December, pp. 1208-1218.

Storesletten, K., C. Telmer and A. Yaron (2004) “Consumption and Risk Sharing over the Life Cycle,” *Journal of Monetary Economics*, 51, pp. 609-633.

Zeldes, S. (1989) “Consumption and Liquidity Constraints: An Empirical Investigation,” *Journal of Political Econom*y, 97, pp. 305-346.

Appendix

A1. Comparisons with Family Income and Expenditure Survey

The NSFIE contains three-month average household expenditure between September and November, while the KHPS contains expenditure for December. Because household expenditure has strong seasonality, I compare the expenditures in the two data sets with the Family Income and Expenditure Survey (FIES) that is conducted every month.[20] Table A1 shows the average expenditures of households of two or more persons and workers.[21] All the figures are in thousands of yen, which is approximately 6.6 US dollars. The NSFIE is very close to the FIES except for Education & Recreation in which the FIES is twice as large as the NSFIE. On the other hand, the KHPS differs from the FIES in Furniture & Daily Use, and Education, Reading, & Recreation.

[20] The FIES is a governmental survey conducted by the Statistical Bureau. Similar to the NSFIE, the FIES contains diary expenditure data, although the sample size is much smaller, about 8,000 per month.

[21] The figures of the FIES in Table 1 are taken from the official web site of the Statistical Bureau.

Table 1: Comparisons of the Two Household Surveys

	NSFIE		KHPS		KHPS Difference ΔE		KHPS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Corr (ΔE, ΔE(-1))
Food (inc. eat out)	4.2293	0.3799	4.1385	0.4375	-0.0165	0.4415	-0.4897
Dining Out	2.2215	0.9739	2.5258	0.6994	0.0632	0.6925	-0.4355
Rent & Repair *	1.7376	2.1418	3.6689	0.7992	-0.0402	0.5338	-0.5103
Electricity, Gas, and Water	2.9220	0.3671	3.0981	0.4000	-0.0231	0.4229	-0.5135
Furniture, Daily Use	1.7866	0.8894	1.9373	0.7953	0.0094	0.9979	-0.5228
Clothes & Footware	2.1443	1.0709	2.5092	0.7950	0.0406	0.8770	-0.4766
Medical Expense	1.8558	1.1278	2.2477	0.8728	0.0174	1.0784	-0.4791
Transportation	3.0467	1.0878	3.6203	1.1285	0.0485	0.8533	-0.4002
Communication **	2.6690	0.6112	2.6118	0.6674	-0.0585	0.6372	-0.4395
Education	2.8836	0.6624	3.1866	0.8667	0.0218	0.8864	-0.4307
Social Intercourse	3.4114	1.1654	3.4793	0.7974	0.0186	0.7892	-0.4851
Total Expense	5.6967	0.4470	5.6328	0.4362	0.0169	0.4221	-0.4999
Earning ***	8.5272	0.4786	8.4647	0.5159	0.0095	0.2517	-0.3637

NSFIE: National Survey of Family Income and Expenditure

KHPS: Keio Household Panel Survey

NSFIE: Average of September-November, 2004. Married, Male Head, Working as Employees, aged between 25 and 65

KHPS: Every December in 2004-2007. Married, Male Head, Workers, aged between 25 and 65.

Samples which are below the lowest 1% or above the highest 1% are excluded.

All the variables are in natural logarithms (in 1000 yen)

*: Excluding Imputed Rents.

**: Including internet connection fee

***: Annual Earning by male head in the previous year

Table 2: Comparisons of the Food Demand Functions

Dependent Variable: ln (Per capita Food Consumption)

	NSFIE		KHPS	
	(1)	(2)	(3)	(4)
ln (Household size)	-0.27075	-0.32235	-0.27588	-0.31417
	(-37.88)	(-45.01)	(-13.21)	(-15.03)
ln (Total Expense per Capita)	0.4168	0.3716	0.4167	0.3793
	(96.27)	(86.71)	(22.27)	(20.69)
Age	.	0.05101	.	0.02207
	.	(20.88)	.	(2.44)
Age^2	.	-0.00048	.	-0.00014
	.	(-17.19)	.	(-1.31)
Constant	5.48511	4.76699	1.28072	0.78962
	(101.32)	(67.92)	(14.13)	(3.97)
N	22191	22191	1928	1928
Log Likelihood	-4906.34	-3975.77	-795.545	-726.626
RMSE	0.30187	0.28948	0.36585	0.35319
R^2	0.456308	0.500047	0.36711	0.410777

Results of the Ordinary Least Squares Estimation
The Differences in the parameter estimates in constants are due to the differences in the scale between NSFIE and KHPS.
t-values are in parentheses.

Table 3: Comparisons of the Food Demand Functions, OLS and IV Estimates

Dependent Variable: ln (Food Consumption Share)

	NSFIE(OLS)		KHPS(OLS)		NSFIE(IV)		KHPS(IV)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln (Household size)	-1.27075	-1.32235	-0.99859	-1.03692	-1.06213	-1.16045	-0.90484	-1.00907
	(-177.8)	(-184.64)	(-47.88)	(-49.74)	(-87.02)	(-91.55)	(-24.56)	(-26.41)
ln (Total Expense per Capita)	-0.58324	-0.62836	-0.57748	-0.61546	-0.33405	-0.42817	-0.33802	-0.50574
	(-134.73)	(-146.62)	(-30.91)	(-33.67)	(-27.35)	(-32.54)	(-5.66)	(-8.15)
Age	.	0.05101	.	0.02126	.	0.04656	.	0.02268
	.	(20.88)	.	(2.36)	.	(17.12)	.	(2.2)
Age^2	.	-0.00048	.	-0.00012	.	-0.00046	.	-0.00014
	.	(-17.19)	.	(-1.2)	.	(-14.95)	.	(-1.19)
Constant	5.48511	4.76699	0.64098	0.16464	2.38281	2.447	-0.46115	-0.35272
	(101.32)	(67.92)	(7.08)	(0.83)	(15.7)	(15.31)	(-1.63)	(-1.06)
N	22200	22200	1928	1928	22200	22200	1928	1928
Log Likelihood	-4910	-3980	-792.444	-720.933	.	.	.	.
RMSE	0.30187	0.28948	0.36526	0.35214	0.32023	0.30081	0.37023	0.34616
R^2	0.60988	0.64126	0.5582	0.58978	0.54573	0.59918	0.5209	0.58173

(1) to (4): OLS

(5) to (8): IV Estimates. Instruments for ln (Total Expense per Capita) is ln(Annual Income per Capita)

t-values are in parentheses.

Table 4: Decomposition of Income and Expenditure Processes

	(1)	(2)	(3)	(4)
	Income	Expenditure	Income	Expenditure
Variance of Permanent Shocks	0.0034	0.0006	0.0078	0.0026
	(168.12)	(16.36)	(4.97)	(2.33)
Variance of Transitive Shocks	0.0099	0.0623	0.0083	0.0609
	(39.71)	(624.87)	(23.94)	(311.94)
Variance of Household Fixed Effects	0.0681	0.0575	.	.
	(6.40)	(4.21)	.	.
Cov(Fixed Effects, Permanent Shocks)	0.0008	0.0005	.	.
	(91.79)	(35.87)	.	.
Cov(Fixed Effects, Transitive Shocks)	-0.0185	-0.0088	.	.
	(-4.25)	(-1.73)	.	.

(1), (2): Using all the moment conditions. (3), (4): Using the differenced series only.
Equal Weighted Minimum Distant Estimation
t-values are in parentheses.
The number of moment conditions: 167 for (1) and (2); 67 for (3) and (4).
Sample used are married male employees who are 27-60 years old.
Data: KHPS

Table A1: Comparisons with FIES: Means and Seasonality

	NSFIE	FIES	KHPS	FIES
Variable	September-November	September-November	December	December
Food (inc. eat out)	72.45	70.18	73.85	84.48
Dining Out	12.46	13.18	13.01	16.05
Rent & Repair *	21.55	20.84	18.15	22.91
Electricity, Gas, and Water	26.52	18.73	20.05	21.23
Furniture, Daily Use	6.27	9.68	9.95	14.08
Clothes & Footware	12.91	13.80	14.19	17.12
Medical Expense	12.74	11.81	12.13	12.22
Transportaion	19.78	31.40	40.68	33.40
Communication **	17.38	13.74	31.27	13.74
Education, Reading, & Recreation	19.76	52.07	16.52	55.09
Social Intercourse & Pocket Money	36.64	43.19	50.75	57.44
Total Expense	292.34	316.29	335.74	369.47

NSFIE: National Survey of Family Income and Expenditure

KHPS: Keio Household Panel Survey

FIES: Family Income and Expenditure Survey

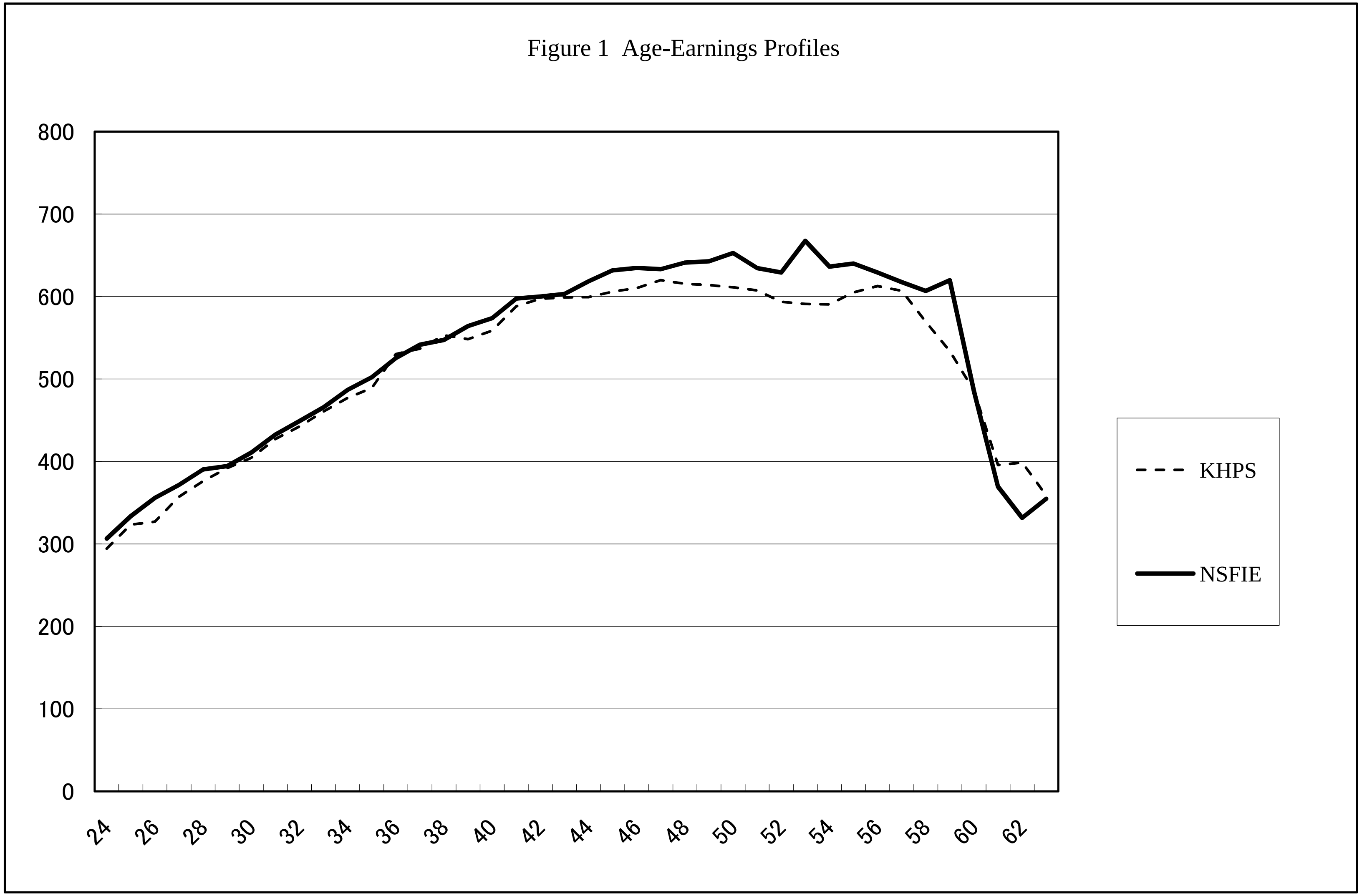
NSFIE:Average of September-November, 2004. Married, Male Head, Employees

KHPS: December 2004. Married, Male Head, Employees

FIES: 2004 September-December, Married Workers.

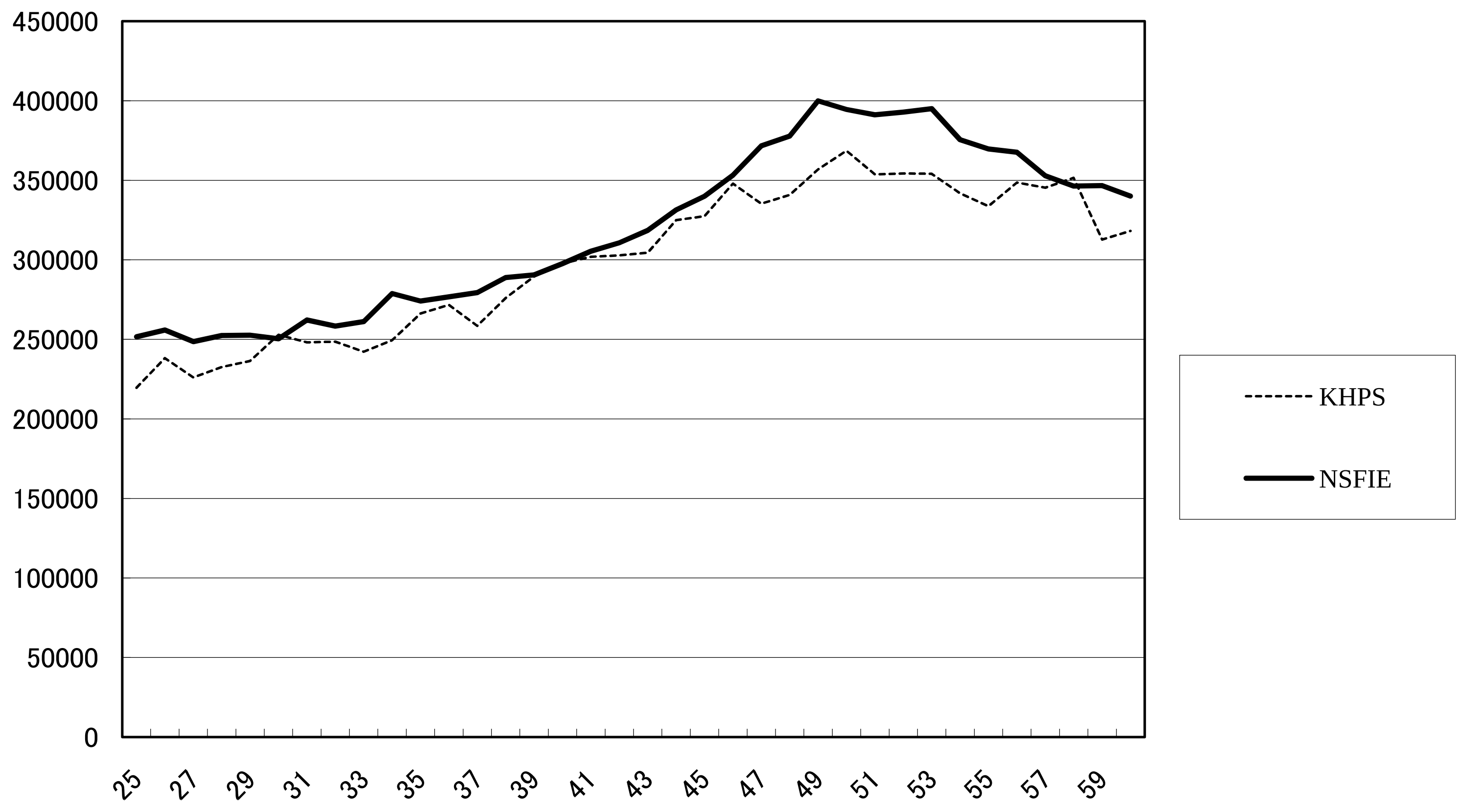
*: Excluding Imputed Rents.

**: Including internet connection fee



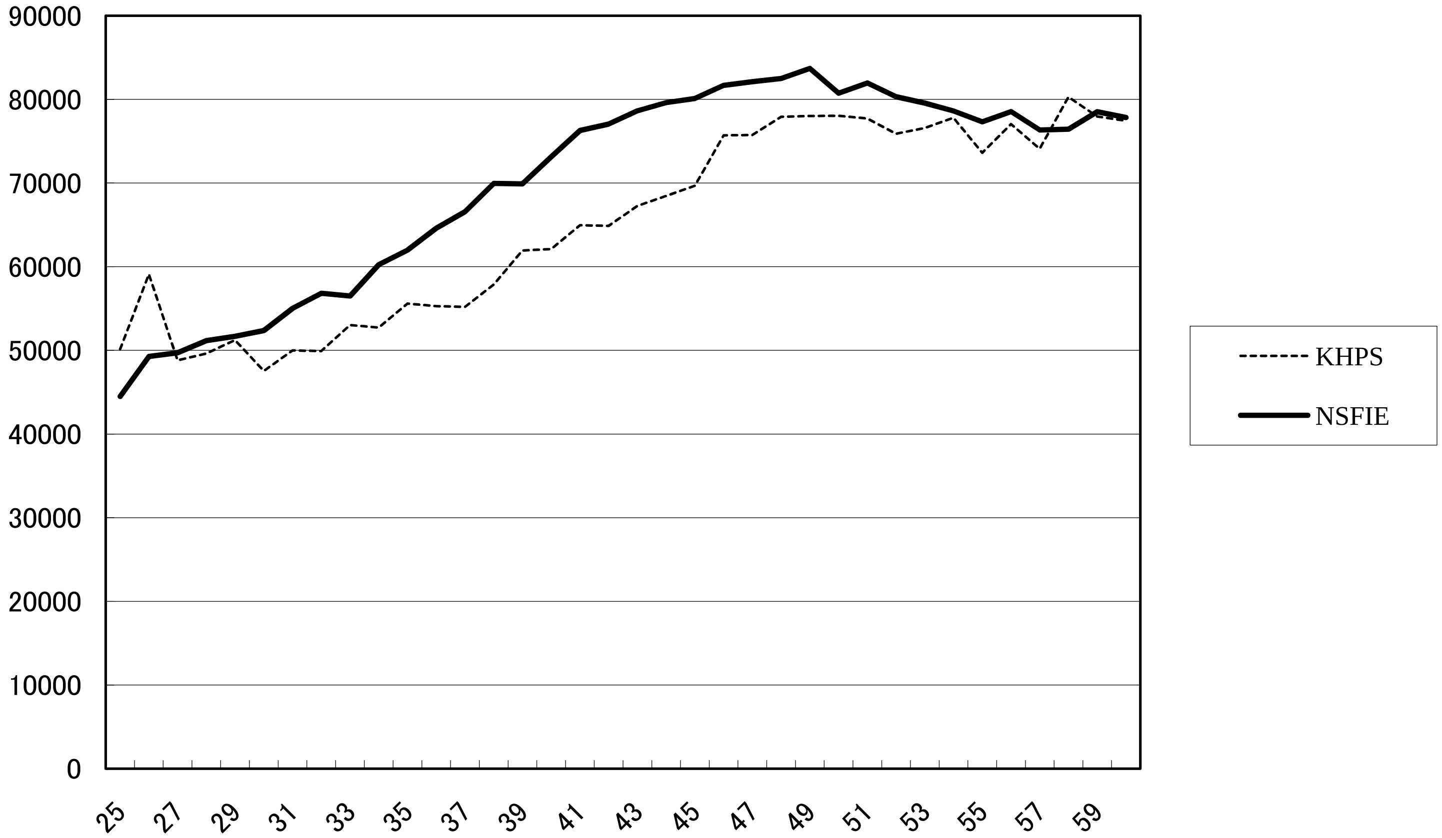
Annual Earnings by Male Head (1000yen), NHFIE:2003, KHPS: 2003-2006

Figure 2:Age- Household Total Expense Profile



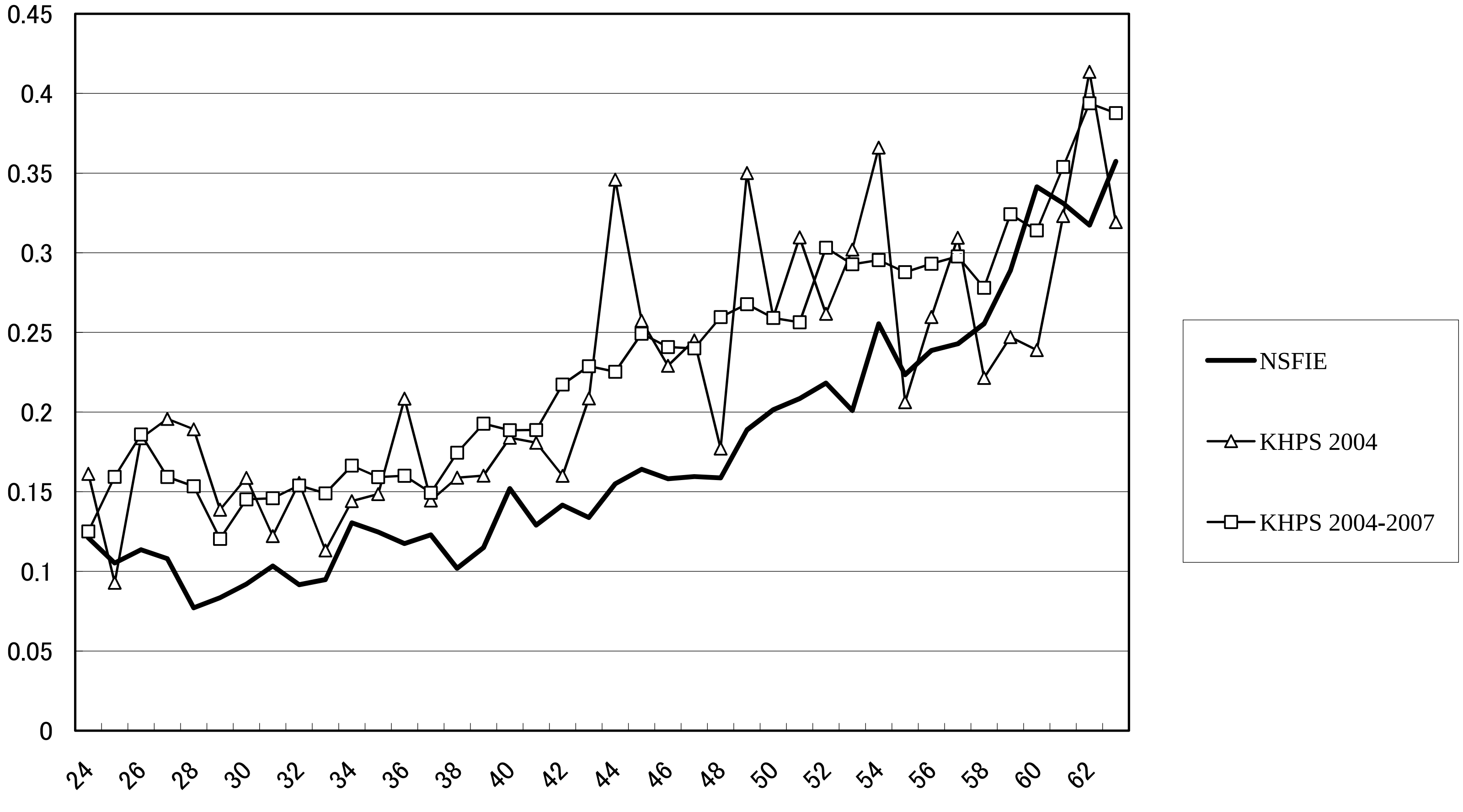
Monthly Expenditure (yen), NHFIE:2004 Sep-Nov, KHPS: 2003-2006 (December)

Figure 3: Age-FOOD Profile



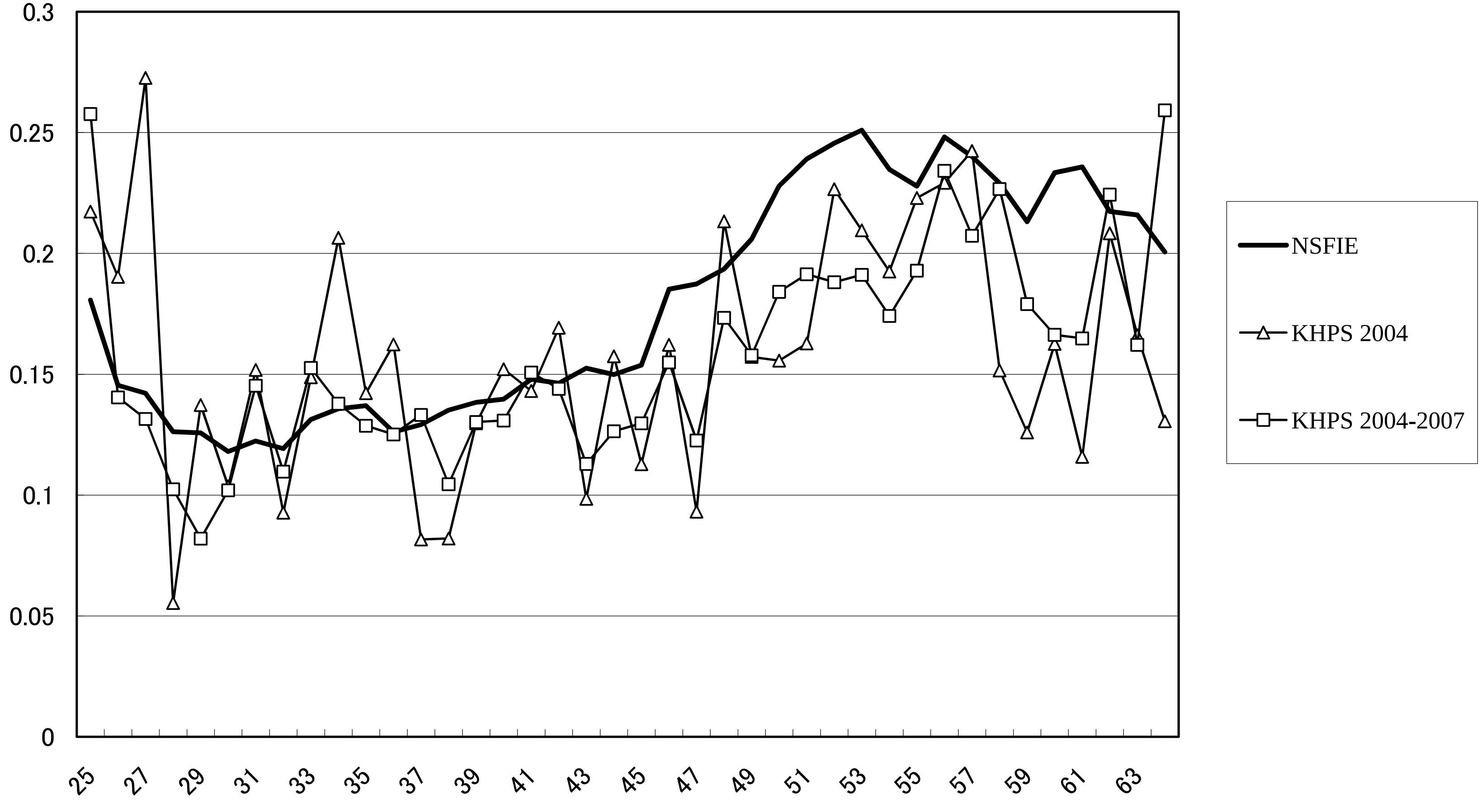
Monthly Food Expenditure (yen)

Figure 4: Age-Income Variance Profile



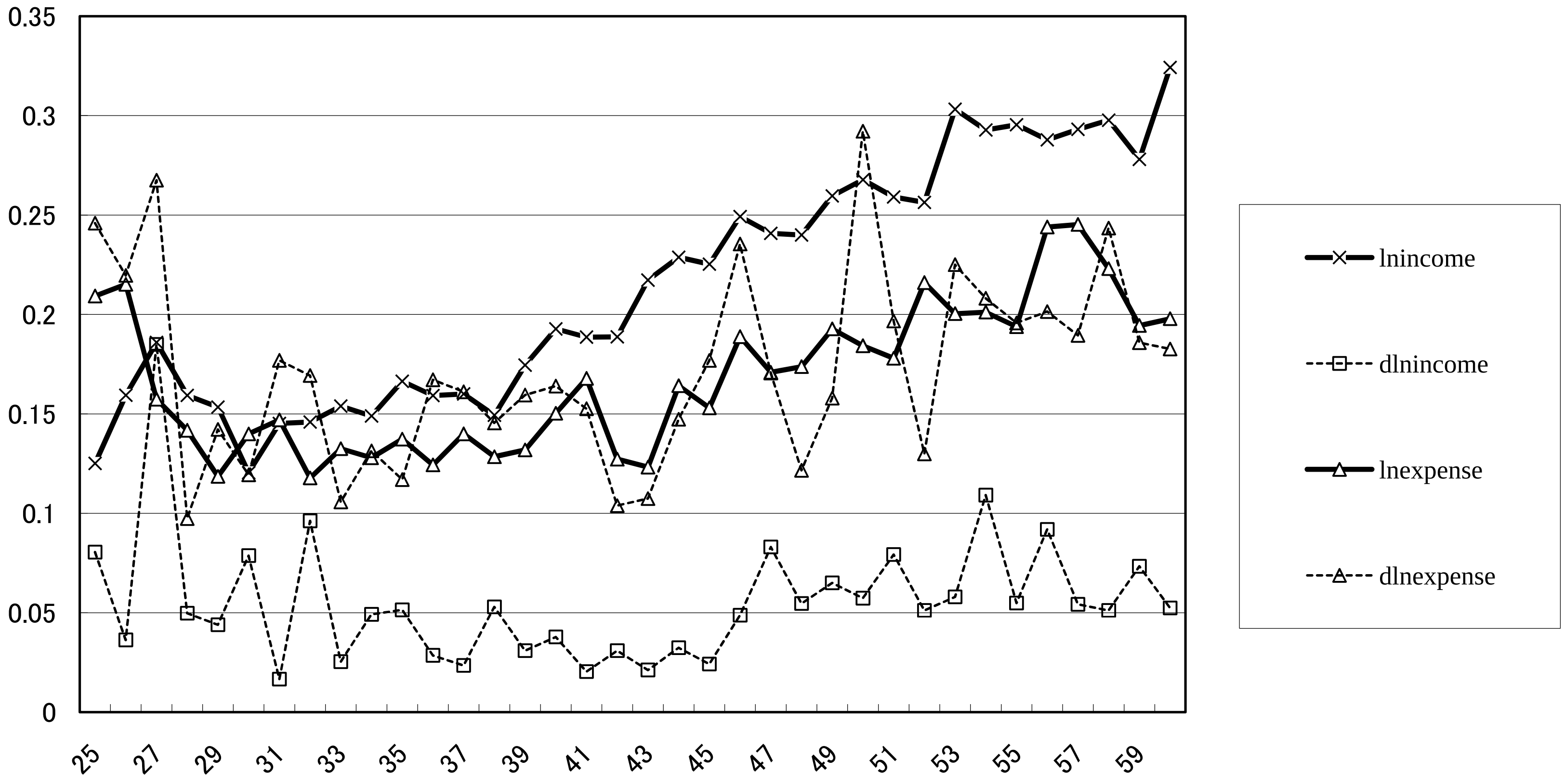
Annual earnings by male head in natural logarithms (1000yen)

Figure 5 : Age-Consumption Variance Profile



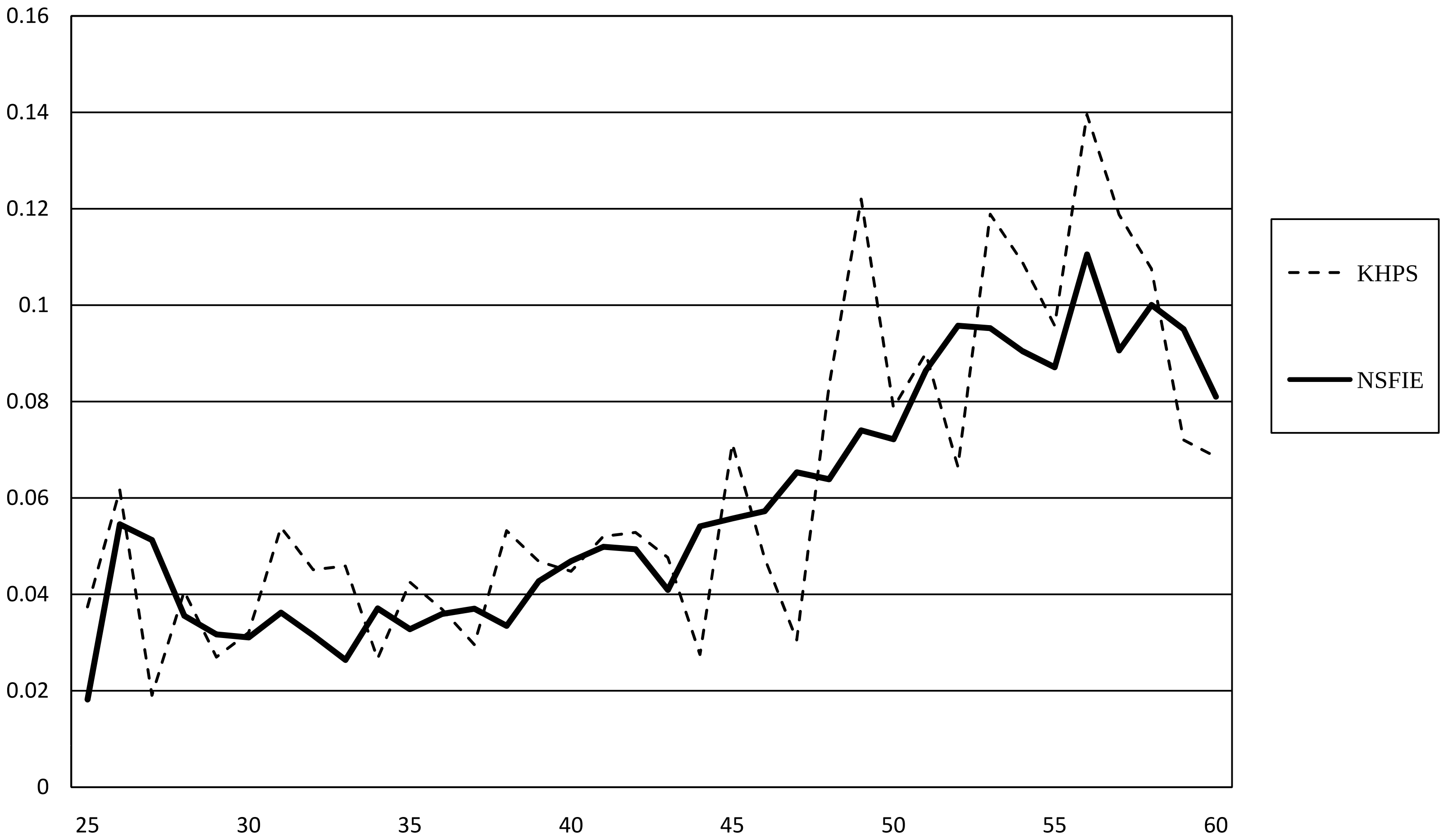
Monthly expenditure in natural logarithms (yen)

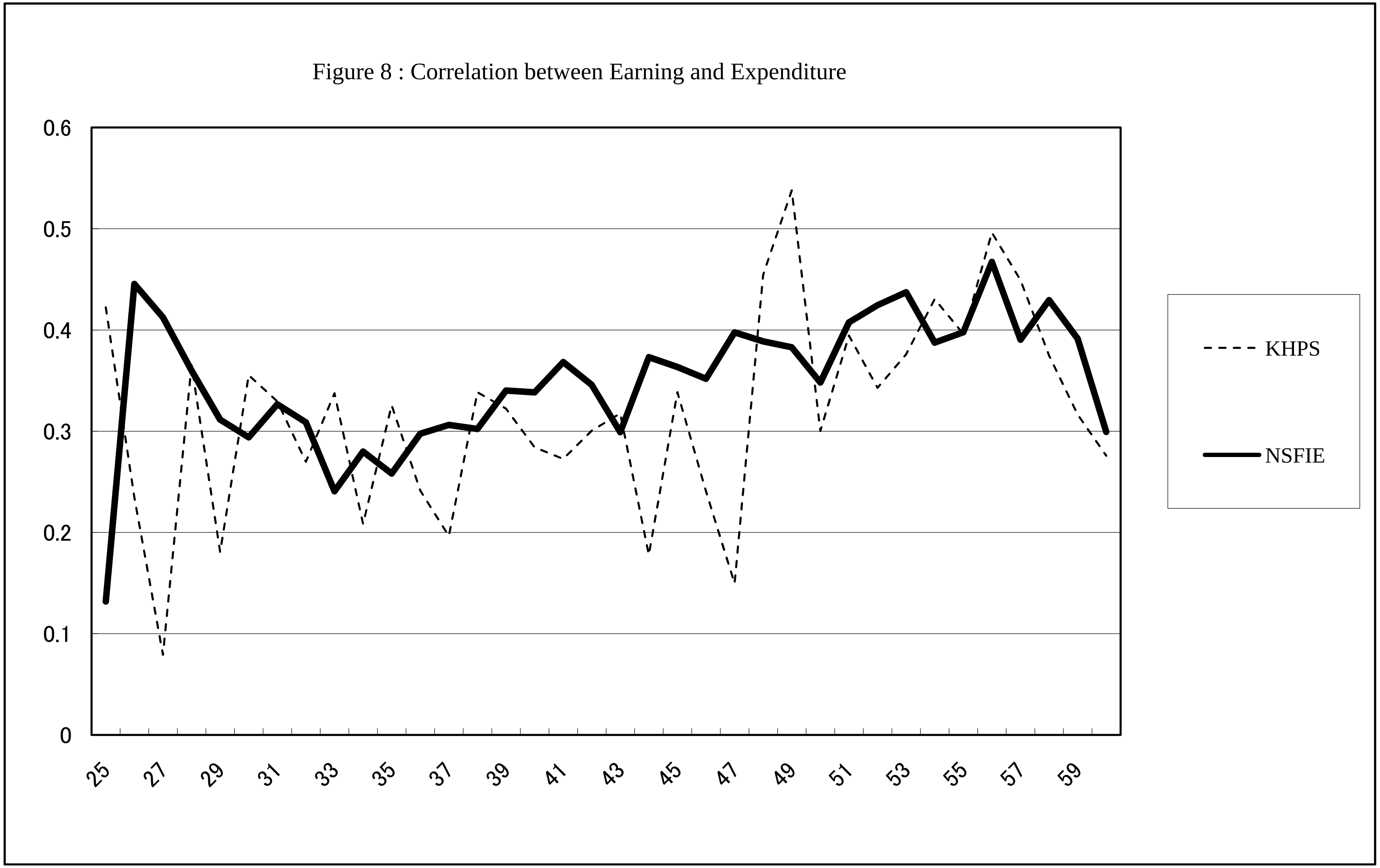
Figure 6: Age-Variance Profile



Annual earnings, monthly expenditures in natural logarithms and their first differences

Figure 7: Covariance Between Earning and Expenditure





Annual earnings in previous years and monthly expenditures