A latent class analysis towards stability and changes in breadwinning patterns among coupled households

Nakai, Miki and Pennoni, Fulvia

College of Social Sciences, Ritsumeikan University, Kyoto, Japan, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy

November 2018

A latent class analysis towards stability and changes in breadwinning patterns among coupled households

Miki Nakai* Fulvia Pennoni†

November 13, 2018

Abstract

We examine how couples'breadwinning patterns are classified and how they have changed over the past four decades during which we have seen increase in women's labor force participation. We explore how the latent variable of spousal breadwinning types is associated with the socioeconomic statuses. To this end, we consider a latent class model especially tailored for an underlying ordinal response derived by comparing two continuous variables. We develop method to estimate the model parameters accounting for the informative sampling design and missing responses. We estimate the measurement model parameters by means of a weighted likelihood function maximised through the Expectation-Maximization algorithm. In order to determine the suitable number of latent classes we rely on the Akaike Information Criterion. Then, fixing the obtained parameters we estimate the latent model parameters by adding the full set of covariates. We make predictions on the basis of the maximum a-posteriori probability. Using data from the Japanese surveys covering the period from 1985 to 2015 breadwinning patterns are examined. Our model discloses two latent classes, each of which represents distinct breadwinning types, namely "traditional couples" and "new couples". Interestingly, the two-class pattern persists across the four waves covering the past 40 years.

KEYWORDS: Akaike Information Criterion, Expectation-Maximization algorithm, Gender Inequality, Household Income Composition, Latent class model.

*College of Social Sciences, Ritsumeikan University, 56-1 Tojiin Kitamachi, Kyoto 603-8577 Japan, e-mail: mnakai@ss.ritsumei.ac.jp

†Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via degli Arciboldi,8, 20126 Mialno, phone:+390264485856, e-mail: fulvia.pennoni@unimib.it

Introduction

A clear division of paid and unpaid work along gender lines in households is found in most countries, but the trend in recent years in several advanced societies is away from the male breadwinner model and towards a dual-earner family. However, despite the increase in the paid employment of married women in Japan as well as Western industrialized societies, the notion of gendered division of labour, that is to say, men should be the family's primary economic provider and women should stay at home to focus primarily on domestic labour and care for family members, have been accepted as a tradition. While recently the number of households with wives entirely dependent on their spouses' income has dramatically declined, most women in dual-earner households still earn much less than their spouses and the households where wives earn equal to or more than their husbands have been very few.

The increasing rate of labor force participation of women not necessarily mean that economic inequality within couples is expected to have changed. Instead, dual-earner households does not liberate women from their traditional gender role, especially in the strong male breadwinner model like Japan. The Japanese tax and pension systems are legislated regarding the male breadwinner household as "standard household' and therefore the Japanese government policies have encouraged married women to stay at home and to be dependent on their husbands, or to work part-time or in non-regular jobs or have broken career patterns. Work and family reconciliation is difficult and this perpetuates women's dependence on a male breadwinner.

As gender inequalities in the division of labor within marriage/family are closely related to gender inequalities in other spheres of life, particularly in the labor market as well as political and other societal arenas, understanding what determines gendered intra-household inequality is essential for researchers who want also to understand other aspects of gender stratification as well as for politicians interested to adopt laws favouring an active women role also in the labour market. Many studies argued that women's economic dependency on men is an important attribute of stratification systems and essential force in the maintenance of gender inequality (e.g. Sorensen and McLanahan, 1987) and recent studies aims at exploring if the gender pay inequalities are due to the perception of both man and woman that they are fair (see Auspurg *et al.*, 2017). However, as women's educational attainment has been rising, more and more women might be part of the dual-career couple and other types of non-traditional couples in contemporary Japan.

Our aim is to examine how couples' breadwinning patterns such as, male-breadwinner or equal-income, are classified among Japanese couples according to a latent class model (Lazarsfeld and Henry, 1968) and how they have changed over the past four decades, during which we have seen increase in women's labor force participation and rising educational attainment. As well as, after identifying distinct latent classes of breadwinning types, we aim to estimate the latent part of the model concerning the associations of the latent variable with covariates in regards of their socioeconomic statuses to detect the main determinants of this classification.

In the present study, we concentrate on heterosexual married couples. In Japan, marriage has been considered as the only way to form a family until recently and marriage

must be officially registered at a local government. Lately there are some people who choose alternatives to marriage such as consensual unions and registered partnerships, but the way people live together in Japan has not significantly changed (1.1 percent live in a consensual union in 2015). It is not only because married couples receive many legal and economic benefits not available to unmarried couples, but also because marriage has been necessary for women's financial survival, social interaction and personal well-being until recently (Saito, 2000).

Some previous research (Raley, 2006) investigated what differentiates dual-income couples from husband sole provider couples, how these associations have changed over time, and what differentiates equal-provider couples from male-breadwinner couples among dual income households. For example, Nakai (2017)attempted to explore by means of a multinomial logit model the determinants which factors differentiate equal-provider couples, where wives share the breadwinning responsibility equally with their partners, from male-breadwinner couples among dual-income households. However, it focused only on dual-earner families and the respondents in sole-breadwinning families were omitted from the analysis. Furthermore, the breadwinning typology commonly used in those studies are usually operationalized based on a priori percentages of wife's income of the total household income. Breadwinning categories of "male breadwinner" or "equal" might arbitrary be constructed. A certain heterogeneity might exist among those with traditional or nontraditional breadwinner types.

We conduct the data analysis by employing a latent class model accounting for a derived ordinal response variable as well as for missing responses and considering the probabilities of the sampling units to get consistent estimates of the parameters of interest. In this way, we get reliable standard errors and correct for possible failures of the sample in covering the target population. The latent class model has been considered for the analysis of data arising in different contexts by many authors since it is a flexible model to account for the heterogeneity among responses provided by different individuals which cannot be explained by means of the observable covariates. This model can be made also tailored for an ordinal response variable by considering a proper parameterization when this response has been derived, like in the context at hand, by comparing values of two or more underlying continuous variables. It is a model-based approach that allow us to investigate the associations with the covariates as well as to dispose of data driven typologies of individuals (see, among others, Pennoni, 2014). Another advantage is that it is possible to elaborate the model parameters in many ways and to assess the tenability of the underlying hypothesis.

Maximum likelihood estimation of the model parameters is well established and it is carried out through the Expectation-Maximization algorithm (see, among others Bartolucci *et al.*, 2013). However, the use of weighted methods for the estimation of the parameters with missing responses is a more recent field of research. In the current proposal, instead of performing listwise deletion we rely on the missing at random assumption and we retain the missing responses for the derived outcome constructed on the responses to the survey, while the values of the missing covariates are imputed through multivariate imputation by chained equations (Van Buuren, 2018). In this way, we allow the allocations on the latent variables at individual level also those not providing a response. The

proposed model is applied to explore the coherent breadwinning arrangement classes and to estimate the effects of the covariates on the underlying latent variable.

The paper is organized as follows: in the next section, we describe the survey data covering the years from 1985 to 2015 on which the empirical analysis is based. Then, we describe the latent class model the maximum likelihood estimation of the model parameters and the computation of the posterior probabilities. Subsequently, we report the results and we make comparisons across waves. Finally, the last section concludes the paper.

Data, variable and measurements

Data are related to spouses within the households were obtained from the past three decades (1985, 1995, 2005) of Japanese cross-sectional data of the Social Stratification and Social Mobility (SSM) surveys, and the last decade (2015) of the Japanese Stratification and Social Psychology (SSP) survey. The respondents are interviewed and asked a wide range of questions such as respondents' socioeconomic background, education, and social consciousness, and the spouse's information when a respondent has a spouse. The surveys have been conducted with similar approach: face-to-face interviews with a special focus on social stratification and inequality in contemporary Japan. All the surveys selected national representative respondents through multiple-stage sampling. The subjects of these surveys were men and women, aged between 20 and 69 for the surveys in 1985, 1995, and 2005, and between 20 and 64 for the 2015 SSP survey.

Data were collected from 2473 men and 1474 women in 1985, 2490 men and 2867 women in 1995, 2660 men and 3082 women in 2005, and 1644 men and 1931 women in 2015. The response rates have been 67.9%, 66.0%, 44.1% and 43.0% in 1985, 1995, 2005, and 2015, respectively. The 2005 survey and the 2015 survey are provided with sampling weights for each survey respondent, which are obtained with respect to the following auxiliary variables: gender, age group, and region of residence. However, the weights are not available for the 1985 and 1995 data files to adjust for unit nonresponse.

We consider as key variable the couple's income provision-role type, obtained by the declared incomes of both survey respondent and spouse. The income variables include earned and investment incomes and all casual and additional incomes such as pensions and dividend on stock shares. The measurement of couple's income provision-role type is based on who may be a dominant provider and the share of contribution of wife's income for the household income. We refer to Nock (2001), and Raley *et al.* (2006) since they stress the relative importance of this concept. The derived response variable is shown in Table 1 and it has been obtained by the following five ordinal categories: (1) husband sole provider category, which consists of couples where only husband is employed; (2) husband provides majority category, which consists of couples where husbands' earnings represent 60 percent or more of the sum of the spouses' income; (3) equal providers category, which identifies couples where wife's earnings represent somewhere from 40 percent to 60 percent, meaning that each partner contributes between 40 and 60 per- cent of total household income; (4) wife provides majority category, which consists of couples where

wives' earnings represent 60 percent or more of the combined total income of the husband and wife; (5) wife sole provider category, which consists of couples where only wife is employed.

[Table 1 about here.]

Table 1 shows the frequencies for the response variable observed in 1985 (N = 2,803), 1995 (N=3,634), 2005 (N=4,386), and 2015 (N=2,497). Survey weights for each respondent are used to analyze the 2005 and 2015 data, but not for 1985 and 1995 data since they are not available. From Table 1 we notice how breadwinning patterns among married couples have changed over the past three decades. The proportion of the households with husbands as sole provider has declined from 42.9% in 1985 to 22.9% in 2015. However, despite the continuing rise in Japanese women's participation in the economy as well as in many Western societies, husbands until recently have been the sole or the primary breadwinner in 65% of the couples and equal-provider couples have been only 11.8% in 2015.

[Table 2 about here.]

The available covariates are illustrated in Table 2. They have been chosen because they may influence division of labour within households according to previous research. Educational level is considered as a proxy of human resources and we expect a relative importance of this variable on married couples' choice of breadwinning patterns. Couples' relative education-level between spouses has been considered to connect to values and culture which may influence division of labour within household. Whether the marriage is homogamous, hypergamous or hypogamous may be associated with patriarchal culture. More patriarchal households, which may be associated with female hypergamous couples, may prefer traditional breadwinning type, as well as their relative power within the marriage (see, among others, Simpson and England 1982). In addition, when available, we consider the variable related to locality of the respondents. The continuous covariates such as wife's age, husband's age, husband income, have been considered according to the observed quantiles illustrated in Table 2 for each survey year and included as dummy variables to adjust distributions throughout the thirty-year period.

[Table 3 about here.]

The age effects on the couple's provision-role type might not be linear and we can assume the impact of various life events and career stages. Therefore, we use dummy variables to detect a non-linear relationship of age group and the type of breadwinning. Descriptive statistics for the variables are presented in Table 3. For 2005 and 2015 data, the values are weighted frequencies, and for 1985 and 1995 data, the values are unweighted frequencies.

The proposed model

We propose a latent class model that is especially tailored for a derived response variable as the couple's income provision-role type. We account for missing responses and for survey weights for the representativeness of each unit in the population. The latter is especially important when the survey sampling design is informative. A latent class model is a special type of finite mixture model (McLaclan and Peel, 2000) and it includes an unobserved random variable with random parameters. Following a formulation proposed by Bartolucci et al., 2013 we account in this paper for missing responses assuming that they are conditionally independent given the latent variable and the observed covariates. For an illustration of the advantages of the use of latent variable models in social research see, among others, Chapters 1 and 2 of Bartolucci *et al.* (2013), and the book of Skrondal and Rabe-Hesketh (2004).

With reference to a random unit drawn from the population of interest let Y_{ij} be the observed derived variable for individual $i, i = 1, \ldots, n$ on the $j, j = 1, \ldots, r$ ordered category. Since this response is obtained by comparing two or more continuous variables we may assume that it is derived on an underlying unobserved latent variable denoted as U_i for subject $i, i = 1, \ldots, n$. We also assume that the latent variable has a distribution with k support points having finite discrete values, so that a semi-parametric model results because its distribution is left unspecified (Lindsay *et al.*, 1991). As in the standard latent class model formulation we assume that the observed responses are independent one another conditionally to the latent variable. The latter is named local independence assumption.

Two set of parameters are specified in this model: one is referred to the manifest part of the model and the other to the latent part. The first is given by the conditional probability of each response category given the latent variable

$$\phi_{j|u} = p(Y_j = y|U = u), \quad u = 1, \ldots, k, \; j = 1, \ldots, r.$$

The second is related to the probability to belong to each latent class. These probabilities may be influenced by individual covariates arranged in the vector $\boldsymbol{X}$ where $\boldsymbol{x}$ is a corresponding realization.

$$\log \frac{p(U = u|\boldsymbol{X} = \boldsymbol{x})}{p(U = 1|\boldsymbol{X} = \boldsymbol{x})} = \log \frac{\pi_{u|\boldsymbol{x}}}{\pi_{1|\boldsymbol{x}}} = \beta_{0u} + \boldsymbol{x}'\boldsymbol{\beta}_{1u}, \quad u = 2, \ldots, k, \tag{1}$$

where β_{0u} is an intercept specific of each latent class and $\boldsymbol{\beta}_{1u}$ is the vector of parameters defining the influence of the covariates on the distribution of the latent variable. Note that, in the parameterization above, the first latent class is taken as a reference category and it has a special role of the latent classes are ordered. In fact, we are interested mainly to understand the pattern with respect to the first level of the response variable and this choice makes the interpretation of the results easier. Other link functions can be chosen such as those presented in Agresti (2018).

To account for individual sampling weights denoted as w_i, $i = 1, \ldots, n$ such as that provided with survey related to the inverse of selection probability we propose to estimate

the model through a weighted log-likelihood. The latter is determined given a sample of n independent individuals for which we observe the responses $y_1, \ldots, y_n$ as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \ell_i(\boldsymbol{\theta}), \quad \ell_i(\boldsymbol{\theta}) = \log p(y_i, \ldots, y_n | \boldsymbol{x}_i),$$

where $\boldsymbol{\theta}$ denotes the overall vector of free parameters arranged in a suitable way and $\boldsymbol{x}_i$ denote the observed covariates configuration for individual i, $i = 1, \ldots, n$. The above quantity is maximized through the Expectation-Maximization (EM) algorithm (Baum *et. al.*, 1970, and Dempster *et al.*, 1977). It is based on the *complete data log-likelihood* and it represents the main tool to estimate the models with latent variables, for more details see Bartolucci *et al.* (2013) .

A two step approach is considered suitable to estimate the model parameters and to avoid bias in the estimated coefficients referred to the covariates. First, we estimate the model with the missing responses and sampling weights excluding the covariates by considering a model selection strategy to choose the best model. According to this procedure we choose the model with the suitable number of latent classes. We perform the model estimation several times to check for local maxima which may arise within the EM algorithm estimation. We rely on the AIC criterion (Akaike, 1973) to guide the selection of the suitable number of latent classes. The latter is a measure of the relative goodness of fit of a model, accounting simultaneously for the accuracy and complexity of the model since it is defined on the basis of the following index

$$\text{AIC}_k = -2\, \hat{\ell}_k(\boldsymbol{\theta}) + 2\#\text{par},$$

where $\hat{\ell}_k$ denotes the maximum of the log-likelihood of the model with k latent classes and #par denotes the number of free parameters in the model. Then, by fixing the parameters of the measurement model we estimate the remaining parameters by considering the full set of covariates. Standard errors for the parameter estimates are obtained according to the observed information matrix computed through numerical methods.

By updating the probabilities to belong to each latent class with the Bayes' theorem (Bayes, 1793) we compute the estimated *a-posteriori* probability to be assigned to each latent class which is determined as

$$\hat{q}_u = \frac{\prod_{j=1}^{r} \hat{\phi}_{j|u} \hat{\pi}_{u|\boldsymbol{x}}}{\sum_{u=1}^{k} \prod_{j=1}^{r} \hat{\phi}_{j|u} \hat{\pi}_{u|\boldsymbol{x}}}, \quad u = 1, \ldots, k, \quad j = 1, \ldots, r. \tag{2}$$

In this way, we dispose of a suitable allocation rule for each individual to be assigned to the latent class having the maximum *a-posteriori* probability and we profile respondents according to the latent class composition.

Results

In this section, we show the results of the model estimated on the data collected in 2015, 2005, 1995 and 1985. First, we show the results for the first step of the estimation

procedure illustrated in the previous section concerning the probability of the response given the latent variable. Second we show the results for the second step related to the estimation of the model with covariates.

Latent class model

Initially, for each wave we performed a multivariate imputation for the missing values reported for age and husband income by a using weighted mean matching method as explained in the previous section according with the sampling weights and with the other covariates as predictive variables (Van Buuren, 2018). The latent class model is estimated for each wave without covariates with a number of latent classes ranging from 1 to 4 by accounting for different initializations of the EM algorithm[1] The highest maximum log-likelihood at convergence, the number of free parameters and the corresponding AIC value are reported in Table 4 for the proposed model estimated on the four datasets.

[Table 4 about here.]

Two latent subpopulations are disentangled for each wave on the basis of the estimated probabilities of the manifest model reported in Table 5.

[Table 5 about here.]

According to the results we define the first latent class as that of Traditional couple (U_T) and the second one as that of New couple (U_N). The first one consist exclusively of individuals where the husband is the sole provider. It is characterized by a high degree of gender role specialization, strong gender based division of work where the husband specialize in market-work and wife in domestic work and caregiver.

The second latent class is a mixed type and it is comprised mainly of couples where the husband is not the unique provider. With respect to the estimated values for wave 2015 we notice that a small proportion of couples within this class is made by the husband as sole provider 18% and 58% are the couples where the husband provide majority. It is important to note that only 16% is the probability of equal providers and 7% are the couples where wife provides majority. It is interesting to note how the estimated probabilities of this class have been changed over time for each level of the response.

Latent class model with covariates

The estimates of the logit regression parameters of Equation (1) affecting the transition from the traditional couples (U_T) to the new couples (U_N) are reported in Table 6. The estimated intercepts are significant for each wave and they are positive indicating a general tendency towards the new type of family U_N especially in 2015.

[Table 5 about here.]

[1]The model is estimated by adapting the functions of the `R` package `LMest` (Bartolucci *et al. 2017. The complete `R` code is available from authors upon request.*

As for the results from the 2015 wave, we observe that having preschool children shows the highest estimated coefficient whose negative sign indicates that respondents having preschool children tend to belong to the cluster of traditional couples (U_T) with respect to new couples (U_N). We notice that the estimated odd ratio for those having preschool children is $\exp(\hat{\beta}_{11p}) = \exp(-2.485) = 0.08$ times the estimated odds for those not having preschool children. Moreover, the effects of preschool children are constantly very strong negative over time (in the 1985 wave, this question was not raised).

As for the effects of wife's education, for the 2005 two significant effects of women's tertiary education on the hypothesized positive effect are found. Compared with high school graduates, women who have completed four years of college or more were expected more likely to be new couple (U_N), because they are expected less likely to have traditional gender-role attitudes. However, significant effects of women's tertiary education are not found in other periods than 2005. Interestingly, we observe a negative effect of the lower level of education to transit to new types of family (U_N) in 2015. Having said that, the pattern of the effects of couple's relative educational level seems constant and significant through four waves: hypergamy, meaning that husband has higher education than his spouse, implicates less likely to be new couple, whereas hypogamy, meaning that wife has higher education than her spouse, implicates more likely to be new couple. The results suggest that women's higher education than husband has been linked to women's greater bargaining power within households. In addition, hypogamy may contribute to more gender-egalitarian attitudes.

Interestingly, we note that the effect of husband age has been positive on the transition to the new type of families (U_N) only in 1985 and only for middle ages from 39 to 53. On the other hand we highlight that in 1995 wife ages from 46 to above had a great impact on the transition to the new type of family (U_N). In 2005, living in the small towns shows a positive effects on this transition. The results can be understood in the light of the situation in Japanese society: the labor force participation rate of women is generally lower in urban areas than in smaller or rural cities in Japan. In large cities, commuting time to work tends to be longer than in rural areas. Moreover, in more urbanised regions, working hours are longer than that in rural areas. Working parents who have young children often rely on grandparents for childcare, but it is difficult for working mothers to ask parents to look after their children in big cities because many grandparents live in another area of the city.

Interestingly, husband's high incomes determine a lower probability towards new family (U_N), meaning that couples with husband having relatively high income, for example over 6 million yen (approximately 44,600 euro) in 2015 and 2005 waves, are less likely to be new families (U_N) than traditional families(U_T). The pattern of income effects are more clear cut after 2000s, although having upper middle level husband's income in 1985 and 1995 are related to belong to traditional households. We note that the median incomes of Japanese households were 4,620,000 in 2005 and 4,270,000 in 2015.

Estimated posterior probabilities

The spouses's allocation to each latent class is performed through the estimated maximum *a-posteriori* probability, determined as in Equation (2). In the last row of Table 6 we show the estimated percentages to be predicted as traditional couple (U_T) for each wave. We notice that the percentage of couples predicted in the traditional family structure is 11.21% in 2015. For these subpopulations we show the corresponding covariates configuration for each year in Table 7 and we compare them with the description made in Table 3.

[Table 7 about here.]

For example, we notice that in each wave we have a higher frequency for the first three categories of the husband's age with respect to Table 3 meaning that the traditional families are mainly prevalent when the husband is younger. We also notice that categories from 5 to 8 corresponding to wife's age from 44 to 58 (see Table 2) are less represented especially in 2015, meaning that especially recently, some wives are induced to get back to work during their middle ages. We expected that the younger couples support gender egalitarian values more and this would be reflected in gender equality in couples earnings structures. However, we found an association contrary to our expectation between age and the probability of being in U_N: households with wife in younger birth cohorts are more likely to be traditional families U_T. It is still not normative for young married women to share equal financial responsibilities within household. This is partly because of the chronic shortages of regular childcare arrangements.

We observe that from the last wave none of the couples is predicted in the traditional family if the husband's income is less then 1,750,000 yen a year (approximately 13,000 euro) and that 61% of the couples predicted as traditional family has preschool children. We observe that in this latent class is mainly characterized by households living in big cities and in relative young cohort, mainly with an husband income more than 4,250,000 yen, where wives have high school education and the husband's education is higher than that of the wife. They mainly have one or more children and in most cases preschool children.

Discussion

We propose a latent class model to account for the observed derived ordinal variable of breadwinner patterns. In the model framework of finite mixture models we suppose the existence of an underlying discrete latent variable and we model the effect of covariates on latent subgroups of couples. We account for informative sampling and for missing responses by means of a weighted maximum likelihood estimation method. We show the results concerning how the breadwinners are changed over time in the Japanese society.

From the analysis of the 1985, 1995, 2005 and 2015 survey data, the proposed model helps to disclose two latent classes which we label "Traditional couple" and "New Couple" among Japanese coupled households. About 20-30% of the households are predicted in the latent class where only husbands undertake the financial responsibility for the family

and wives bear domestic responsibility. The percentages show an increasing trend up to 2005 when 2015 we found a relevant increase in the proportion of the non-traditional families. One of the reasons why the new families have been more represented in the past decade may be that being a conventional single-income household is becoming more difficult due to income stagnation and even decline. Over the past two decades, average male annual income fell steadily also due to recent economic crisis. Therefore, in cases where husbands do not have relatively profitable and economically sound job, wife tend to have shared breadwinner roles.

Having young children has been constantly related to the greater likelihood that the respondents are in couples where husband is a sole provider up until now. It is partly because of gender stereotype on wife's employment and on the family roles in Japan that women have to stay at home, cook, clean, and care for children when having little babies. Another challenge for the families with preschoolers is the chronic shortages of regular childcare arrangements. Childcare has become an important policy issues in Japan as well as in many countries, but still there is a large number of children on waiting lists for childcare centers. Other childcare services such as home-based group care by family day care providers or babysitters who watch the children in either their own homes or the children's home are not very popular in Japan and they are expensive. Moreover, when the employed women become pregnant and hope to take some maternity leave, some employers place pressure on women to quit. This leads to the coining of the phrase "maternity harassment". These conditions have precluded mothers of preschoolers from staying the workforce.

The effects of husband's income have been significant and negative over time. Having a high-income husband implies to have more chance to build a traditional family. If the levels of husband's income are enough to maintain a family and run their home, wives tend to be dependent on him and stay at home. When husband's earnings are not enough to take care of all family financial obligations, wives tend to enter the labor force to supplement the husband's income. Under the current Japanese social security system, dependent spouses of regular employees are automatically entitled to the basic pension without making any payments. Critics say that this entitlement deters full-time housewives from participating in labor market. The result of this paper can be understood in the light of that argument.

The effect of relative educational level between couples has been constant until recently. The hypergamous couples where husband educated more than his wife highten a couple's likelihood of being a traditional, non-egalitarian marriage practices. This suggests that power imbalances within relationships may be related to the couple's choice of income provision-role type. These asymmetric gender relation within families seem to be associated with a patriarchal culture. More patriarchal couples may prefer traditional marriage practices. Having tertiary education does not show a higher probability to be in a latent cluster identified as the newer types of family, where wives contribute more or less to their household income until recently. This leads to low economic returns to investment in education for women in Japan.

A worthwhile direction for future research would be to make a few comparative analysis between Japan and Western societies, on which some existing studies particularly

have focusd (e.g. Crompton and Lyonett, 2006; Korpi *et al.*, 2013). Comparative analysis of transition of breadwinning patterns beyond Western societies where breadwinning and caregiving expectations are institutionalized differently would contribute to improve further understanding of the women's opportunities in the various institutional context.

References

Agresti, A. (2018). *An Introduction to Categorical Data Analysis*, 3rd ed. Wiley.

Auspurg K., Hinz T. and Sauer C. (2017). Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments, *American Sociological Review*, 82: 179–210.

Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. In BN Petrov, C F (Eds.), *Second International Symposium on Information Theory*, 267–281. Budapest: Akademiai Kiado.

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Boca Raton: Chapman and Hall/CRC press.

Bartolucci, F., Pandolfi, S. and Pennoni, F. (2017). `LMest`: An `R` Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software*, 81: 1–38.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41: 164–171.

Bayes T., Price M. (1763). An Essay towards solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London*, 53: 370-418.

Crompton, R. and Lyonette, C. (2006) Work-life "balance" in Europe. *Acta Sociologica*, 49: 379–393.

Dempster, A. P., Laird, N. M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39: 1–38.

Korpi W., Ferrarini T. and Englund S. (2013). Women's Opportunities under Different Family Policy Constellations: Gender, Class, and Inequality Tradeoffs in Western Countries Re-examined. *Social Politics*, 20: 1–40.

Lazarsfeld P. F. and Henry N. W. (1968). *Latent Structure Analysis*. Houghton Mifflin, Boston.

Lindsay, B., Clogg, C.C. and Grego J. (1991). Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis. *Journal of the American Statistical Association*, 86: 96–107.

McLachlan G. J., and Peel D. (2000). *Finite Mixture Models*. Wiley Series: New York.

Nock, S. L. (2001). The marriages of equally dependent spouses. *Journal of Family Issues*, 22: 755–775.

Nakai, M. (2017). Changes in couples' breadwinning patterns and wife's economic role in Japan. In: Greselin, F., Mola, F., and Zenga, M. (Eds.), *Proceedings of the conference of the CLAssification and Data Analysis Group (CLADAG) 2017*, Universitas Studiorum, Mantova, 1-6.

Pennoni, F. (2014). *Issues on the Estimation of Latent Variable and Latent Class Models.* Scholars'Press, Saarbücken.

Raley, S. B., Mattingly, M. J. and Bianchi, S. M. (2006). How dual are dual-income couples? Documenting change from 1970 to 2001. *Journal of Marriage and Family*, 68:11–28.

Saito, M. (2000). *Modern Girls*. Magazine House, Tokyo.

Simpson, I. H. and England, P.(1982). Conjugal work roles and marital solidarity. In: Aldous, J. (Ed.), *Two Paychecks: Life in Dual-earner Families*, Sage, Beverly Hills, CA, 147-172.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton: Chapman and Hall/CRC press.

Sorensen, A. and McLanahan, S. (1987). Married women economic dependency, 1940-1980. *American Journal of Sociology*, 93: 659-687.

Van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Second Edition. Chapman & Hall/CRC. Boca Raton, FL.

Table 1: *Observed and missing frequencies of the couple's income provision-role type weighted with the survey weights for years 2015 and 2005 and unweighted for years 1995 and 1985: (1) "husband sole provider", (2) "husband provides majority", (3) "equal providers", (4) "wife provides majority", (5) "wife sole provider".*

Income provision-role categories (%)	1	2	3	4	5	Missing
2015	22.9	42.2	11.8	5.3	0.6	17.2
2005	33.6	28.1	6.9	1.6	1.0	28.7
1995	38.2	37.8	6.2	1.2	0.7	16.0
1985	42.9	36.7	6.7	1.2	0.1	12.4

Table 2: *Classes for the continuous variables and labels for the categories of the other covariates. Husband's income in ten thousand yen a year. The presence of a preschooler is coded as a binary variable with respect to children's age 0-6. Couples' relative education is measured by two categories: 1 husband has higher education than wife, 2 wife has higher education than husband. Wife's education: 1 less than high school, 2 high school, 3 two-years college degree, 4 four-year college degree or higher. Size of the place of living: 1 major cities; 2 > 200,000; 3 (100,000, 200,000]; 4 ≤ 100,000; 5 small towns and villages.*

	Categories										
Covariates	0	1	2	3	4	5	6	6	8	9	10
preschool children		1									
relative education		1	2								
wife'education		1	2	3	4						
size of the place		1	2	3	4	5					
n. children	0	1	2	≥ 3							
2015											
husband'age		≤ 34	(34,39]	(39,43]	(43,46]	(46,50]	(50,54]	(54,58]	(58,61]	(61,64]	> 64
wife'age		≤ 32	(32,37]	(37,40]	(40,44]	(44,47]	(47,51]	(51,55]	(55,58]	(58,61]	> 61
husband's income		≤ 175	(175,275]	(275,325]	(325,375]	(375,425]	(425,500]	(500,600]	(600,700]	(700,900]	>900
2005											
husband'age		≤ 34	(34,40]	(40,45]	(45,50]	(50,54]	(54,57]	(57,61]	(61,64]	(64,68]	> 68
wife'age		≤ 32	(32,37]	(37,42]	(42,47]	(47,51]	(51,55]	(55,58]	(58,61]	(61,65]	> 65
husband's income		≤ 175	(175,225]	(225,275]	(275,375]	(375,425]	(425,500]	(500,600]	(600,700]	(700,900]	>900
1995											
husband'age		≤ 33	(33,38]	(38,42]	(42,45]	(45,48]	(48,52]	(52,56]	(56,60]	(60,65]	> 65
wife'age		≤ 31	(31,35]	(35,40]	(40,43]	(43,46]	(46,49]	(49,53]	(53,57]	(57,62]	> 62
husband's income		≤ 150	(150,250]	(250,350]	(350,400]	(400,500]	(500,600]	(600,700]	(700,800]	(800,900	>900
1985											
husband'age		≤ 32	(32,36]	(36,39]	(39,42]	(42,46]	(46,49]	(49,53]	(53,57]	(57,62]	> 62
wife'age		≤ 29	(29,33]	(33,36]	(36,39]	(39,44]	(44,48]	(48,51]	(51,56]	(56,61]	> 61
husband's income		≤ 130	(130,190]	(190,250]	(250,310]	(310,350]	(350,430]	(430,480]	(480,550]	(550,650]	>650

Table 3: *Descriptive statistics for the available covariates by survey year. The frequencies are weighted with survey weights for years 2015 and 2005. The labels of each variable are define in the footnote 1.*

Covariates (%)	0	1	2	3	4	5	6	7	8	9	10
2015											
husband'age		13.6	12.1	11.9	9.1	10.5	10.8	10.6	8.6	8.2	5.2
wife'age		12.6	12.1	10.6	11.5	9.4	10.9	9.9	7.5	8.6	6.9
husband's income		10.1	11.3	9.7	8.1	9.6	13.3	10.1	8.1	10.1	9.5
relative education:		32.6	16.5								
wife'edu.		4.1	56.5	24.4	18.3						
preschool		25.2									
n. children	87.4	24.3	43.6	18.0							
size of the place of living		28.9	24.8	17.2	20.5	8.7					
2005											
husband'age		15.0	13.3	10.3	10.7	9.0	9.6	10.1	7.0	9.1	5.9
wife'age		14.8	12.2	10.8	12.0	9.1	10.8	9.3	6.5	8.6	6.0
husband's income		10.9	6.9	7.8	16.3	11.4	12.1	11.1	6.8	8.3	8.5
relative education:		28.3	14.3								
wife'edu.		12.9	57.0	16.6	11.3						
preschool		20.3									
n. children	90.1	16.6	48.3	24.6							
size of the place of living		20 .0	22.3	18.7	25.3	13.8					
1995											
husband'age		11.2	9.3	10	10	10.2	10.3	9.1	10.3	10.7	8.9
wife'age		11.8	8.4	12.5	9	10.3	8.6	9.7	10	11.1	8.6
husband's income		4.9	7.2	13.5	14.1	14.5	12.3	10.1	7.0	6.4	10.0
relative education:		29.1	12.6								
wife'edu.		23.3	55.8	13.2	7.2						
preschool		18.3									
n. children	92.3	15.0	51.6	25.7							
1985											
husband'age		11.1	11	9.4	9.1	11.4	8.8	10.8	9.7	10.3	8.6
wife'age		11.3	9.7	10.5	9.3	11.4	8.8	9.3	11.9	8.9	8.7
husband's income		7.8	8.1	13.1	10.5	9.5	11.1	6.7	12.6	8.5	12.1
relative education		28.0	12.3								
wife'edu.		34.3	52.3	7.4	5.0						
n. children	98.0	6.0	16.4	11.3							

Table 4: *Maximum log-likelihood, number of parameters, and AIC index for the model with two latent classes for each wave.*

Survey year	$\hat{\ell}$	#par	AIC
2015	-2,456.7	11	4,935.4
2005	-3,421.2	11	6,864.4
1995	-3,653.5	11	7,329.1
1985	-2,737.5	11	5,497.1

Table 5: *Estimated conditional probabilities ($\hat{\phi}_{j|u}$) under the selected model of the responses given the latent classes (1 for latent class of traditional families U_T and 2 for the new types of families U_N) for each wave.*

	Income provision-role categories	1	2	3	4	5
	Latent classes					
2015	1	1.000	0.000	0.000	0.000	0.000
	2	0.179	0.578	0.161	0.074	0.008
2005	1	1.000	0.000	0.000	0.000	0.000
	2	0.214	0.586	0.145	0.033	0.022
1995	1	1.000	0.000	0.000	0.000	0.000
	2	0.312	0.568	0.093	0.018	0.010
1985	1	1.000	0.000	0.000	0.000	0.000
	2	0.359	0.526	0.096	0.018	0.002

Table 6: *Estimates of the regression parameters for each wave. (Income in ten thousand yen; significance at 10%(†), 5%*,1%**) and percentage of predicted couples in traditional type of families.*

		1985	1995	2005	2015
	$\hat{\beta}_0$	1.129**	2.506**	1.377*	3.777**
husband's age:	1	-0.081	-0.180	0.106	0.349
	3	0.147	-0.324	0.846**	0.613
	4	1.324**	-0.373	0.441	0.776
	5	1.234**	-0.668	0.434	-0.599
	6	1.223**	1.018	-0.314	-0.331
	7	1.730**	-2.512**	-0.630	0.599
	8	0.612	-2.571**	-0.480	-1.299
	9	0.030	-3.784**	-0.305	-1.412
	10	-0.124	-4.280†	0.167	0.428
wife's age:	1	-0.709†	-0.462	-0.451	-1.039**
	3	0.414	0.452	-0.272	-0.437
	4	0.238	0.729	0.334	-0.877
	5	0.413	2.723	0.133	-0.498
	6	0.198	2.738**	0.221	0.271
	7	-0.508	2.945**	-0.176	-0.411
	8	-0.338	2.133**	-0.819	1.284
	9	-0.920**	1.998**	-1.077	-1.045
	10	-0.995**	1.841**	-1.473*	-1.425
wife's education:	1	0.550**	0.309	0.034	-1.085**
	3	-0.399	-0.011	-0.131	0.115
	4	-0.407	0.515	0.050*	0.471
relative education:	1	-0.459**	-0.476**	-0.609**	0.328
	2	0.785†	1.051**	0.323	0.887*
preschool children:	1	-	-2.114**	-0.970**	-2.485**
children:	0	0.639	-0.621	0.133	-0.464
	1	-0.216	-0.559**	-0.331**	-0.479*
	3	0.444	0.707**	0.149	0.149
size of the place of living:	1	-	-	0.048	-0.164
	3	-	-	0.129	-0.379
	4	-	-	0.223	0.236
	5	-	-	0.438*	0.853
husband's income:	1	-0.019	0.256	0.380	0.772
	2	0.877	0.300	0.110	0.952
	4	0.291	-0.253	0.184	0.333
	5	0.255	-0.061	0.131	0.072
	6	-0.241	-0.673**	0.061	-0.547
	7	0.030	-0.434**	-0.272	-0.427
	8	-0.684*	-1.038	-0.657**	-1.479**
	9	-0.810*	-0.078	-0.671**	-1.221**
	10	-0.975**	-0.538	-0.632**	-1.265**
Percentage predicted in U_T:		21.5%	22.2%	31.0%	11.21%

Table 7: *Frequencies for couples predicted in latent class U_T of traditional families. For years 2015 and 2005 the frequencies are weighted with survey weights.*

		1985	1995	2005	2015
husband's age	1	19.1	27.9	18.9	25.5
	2	19.3	15.1	17.0	19.5
	3	15.9	8.7	5.6	10.1
	4	4.0	3.5	4.3	3.2
	5	1.0	1.4	2.6	9.9
	6	3.2	0.1	8.3	4.7
	7	2.2	1.1	13.2	1.0
	8	6.6	2.4	10.3	1.0
	9	13.6	18.2	13.1	14.7
	10	10.1	21.6	6.7	1.5
wife's age	1	17.3	30.3	20.7	29.4
	2	6.1	16.1	12.4	11.6
	3	5.8	8.4	9.6	11.9
	4	4.3	2.6	3.3	11.9
	5	5.1	0.0	4.6	5.5
	6	4.5	0.1	7.2	1.1
	7	11.1	0.1	10.1	5.8
	8	12.6	6.2	10.5	0.0
	9	14.5	17.1	12.7	11.3
	10	18.5	19.0	8.8	11.6
husband's income	1	4.3	4.6	7.6	0.0
	2	0.8	5.1	7.4	5.3
	3	9.5	16.9	8.3	10.2
	4	7.6	20.1	16.4	5.8
	5	6.1	14.4	9.6	10.0
	6	10.1	14.0	9.3	18.1
	7	4.7	6.9	12.0	11.0
	8	19.9	6.8	8.3	12.5
	9	13.3	3.2	10.2	11.9
	10	23.6	7.9	10.9	15.2
wife's education	1	15.6	20.2	14.5	8.4
	2	61.1	55.6	53.8	58.1
	3	13.1	17.4	20.4	20.4
	4	9.0	6.0	10.8	20.9
relative education	1	46.0	44.8	43.3	34.4
	2	4.7	5.3	6.4	10.0
preschool children	1	-	51.0	35.5	61.1
number of children	0	6.2	6.2	4.9	4.9
	1	7.3	28.7	24.7	40.9
	2	16.1	51.5	50.7	38.9
	3	4.7	13.6	19.7	14.9
size of the place of living	1	-	-	22.7	33.1
	2	-	-	26.7	27.0
	3	-	-	19.8	23.4
	4	-	-	21.7	13.6
	5	-	-	9.1	2.9