# Diskussionspapiere
# Discussion Papers

Discussion Paper No. 213

# Measuring Income Risk

by
Simon Burgess[1], Karen Gardiner[2],
Stephen P. Jenkins[3], and Carol Propper[4]

[1] University of Bristol, CASE and CEPR
[2] CASE, London School of Economics
[3] ISER, University of Essex and DIW Research Professor
[4] University of Bristol, CASE and CEPR

Berlin, May 2000

Deutsches Institut für Wirtschaftsforschung, Berlin
Königin-Luise-Str. 5, 14195 Berlin
Phone: +49-30-89789- 0
Fax: +49-30-89789- 200
Internet: http://www.diw.de
ISSN 1433-0210

miles_may00.doc

# MEASURING INCOME RISK


| Simon Burgess | Karen Gardiner | Stephen P. Jenkins | Carol Propper |
|---|---|---|---|
| University of Bristol, CASE and CEPR | CASE, London School of Economics | ISER, University of Essex and DIW-Berlin | University of Bristol, CASE and CEPR |


3 May 2000


*Abstract*

We provide a critique of the methods that have been used to derive measures of income risk and draw attention to the importance of demographic factors as a source of income risk. We also propose new measures of the contribution to total income risk of demographic and labour market factors. Empirical evidence supporting our arguments is provided using data from the British Household Survey.



*JEL codes:* D12, D31, D84



*Acknowledgements*

CASE receives core funding support from the ESRC; ISER receives core funding support from the ESRC and the University of Essex. We thank John Micklewright for helpful comments on an earlier draft.



*Addresses for correspondence*

Burgess and Propper: Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN (email simon.burgess@bristol.ac.uk; carol.propper@bristol.ac.uk.) Gardiner: Centre for Analysis of Social Exclusion, London School of Economics, Houghton Street, London WC2A 2AE (email k.gardiner@lse.ac.uk). Jenkins: Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ (email stephenj@essex.ac.uk)

## 1. INTRODUCTION

A person faces income risk whenever his or her future income stream deviates from its expected future path. Measurement of this income risk is of increasing interest to economists: they are concerned to analyse the impact of risk on behaviour, and also to summarise the amount of income risk for normative reasons. Under the former heading, there is, for example, a growing literature on 'precautionary saving' that considers the extent to which individuals facing greater income uncertainty consume less and save more.[1] Under the second heading, there is a close overlap between the concepts of income risk, income mobility and transitory income variation (more on this link below). Information about how much income risk there is, and how this has been changing over time, has been found useful for assessing the 1980s rise in earnings and income inequality in the USA and the UK.[2] Patterns of income risk are also of relevance to the analysis and design of social insurance schemes and they direct attention to the sources of income risk with which the welfare state should be concerned.[3] Clearly all these studies rely on getting good empirical measures of income risk. In this paper we show how to derive income risk measures taking proper account of demographic events, and provide new evidence about the contributions to an individual's income risk of demographic factors vis-à-vis labour market factors.

Our research was motivated by the observation that much of the literature provides an unduly narrow perspective on what the potential sources of income risk are, and therefore who experiences risk. Much research has focused on labour market risk and tended to ignore demographic risk. Some of this focus is implicit but is evident nonetheless from researchers' choice of samples to use to study income risk. For instance the empirical analysis in several leading papers in the precautionary savings literature has focused on prime-aged male household heads who did not separate (if married) or marry (if single) during the observation period. By construction, income risk amongst all other persons in the population – e.g. many women (spouses), the elderly, those experiencing family formation or dissolution – is simply not

[1] See e.g. Banks et al. (1994, 1999), Carroll (1994), Carroll and Samwick (1994, 1995), Dardanoni (1991), Guiso et al. (1992, 1996), Kazarosian (1997), Miles (1998), and Skinner (1988).
[2] See e.g. Blundell and Preston (1998), Dynarski and Gruber (1997), Gottschalk and Moffitt (1994), Jarvis and Jenkins (1998), and Moffitt and Gottschalk (1993).
[3] See e.g. Bird (1995), Bird and Hagstrom (1999), and Haveman and Wolfe (1985).

examined. This leads to an understatement of income risk in total and may also give a misleading picture of population patterns of income risk in terms of characteristics such as age or income. Another important effect of the conventional sample selection method is that it downplays the role of changes in the composition of an individual's household as a potential source of income risk in favour of labour market events. The persons included in the sample have not experienced major demographic events such as marriage, divorce, or death of a partner, all known to have strong associations with changes in income and poverty status (see e.g. Bane and Ellwood, 1986; Jenkins, 1998; Stevens, 1994, 1995) and important sources of income risk.

To underline the importance of demographic factors for income risk, consider the following example based on six waves of data from the British Household Panel Survey. We classify adults as living in either 'intact' households (those for which the person's household head and the head's marital status remain unchanged during the time the person is in the sample) or in 'non-intact' households (the remainder). There are two important findings. First, almost half of the sample (46 percent) live in a non-intact household. It is important to be able to generate income risk measures for such a substantial proportion of the population. Put another way, excluding adults from non-intact households is unlikely to produce unbiased estimates of population income risk. Thus, second, we find that the degree of income risk differs significantly between adults in non-intact and adults in intact households. Mean income risk is 0.18, while the mean for persons living in non-intact and intact households is 0.21 and 0.16 respectively.[4] This relationship also holds at the median, 25th, 75th and 90th percentiles of the distribution of income risk (though not at the 10th percentile), with the income risk differential between non-intact and intact households increasing as one moves up the income risk distribution. Clearly demographic factors matter.

The goals of this paper are twofold. First, we aim to show how to measure income risk in a manner that treats demographic events properly. Much research has derived measures using methods which incorrectly condition upon demographic events (even though these are themselves a source of income variation). Second we aim to measure the contribution of demographic factors to income risk and contrast it with the contribution of labour market factors (the conventional focus).

[4] The measure of income risk for each person is the variance of log household income.

At the heart of the paper are analyses of the determinants of household income risk. We show how analysts' choices of the factors that determine the income generation process affect estimates of an individual's income risk, and of its heterogeneity across the population. We also examine the importance of demographic factors to income risk. We find that the level and distribution of estimated risk across different groups in the population is sensitive to the selection of factors regarded as determinants of permanent as distinct from transitory determinants of income heterogeneity. We find that both labour market and demographic factors are associated with the household income variability experienced by individuals. While on average the importance in risk of demographic factors is less than that of labour market factors, for some sub-populations demographic factors are associated with a high proportion of their income variability. Individuals who live in intact households experience significantly less income variability than those who live in non-intact households. Younger individuals experience more demographic risk than older ones. Lower income individuals experience more risk attributable to labour market events than richer individuals.

The paper is organised as follows. Section 2 examines the specification of models that underpin the measurement of income risk. By definition, all measures of income risk require a model of how individuals form expectations about the path of future incomes, and two main approaches have been employed. The first approach supposes that individuals make their predictions by comparing their current income with the average income of all people with similar characteristics to themselves at that time. Each person's income risk is summarised by the personal income deviations from the average. The second approach supposes that each individual derives her expectations about future income from a projection based upon her fixed and therefore predictable characteristics. We argue that the first provides inappropriate measures of income risk. We then derive new measures of the contribution to total risk for an individual of any conditioning variable. Our empirical analysis is based on data from the first six waves of the British Household Panel Survey (BHPS). We describe this source, subsample selection criteria, and income definitions in Section 3. In section 4 we present the empirical results. Section 5 concludes.

## 2. CONCEPTUAL FRAMEWORK

### 2.1 Defining Income Risk

We define income risk as unpredictability of income,[5] not simply variability, and denote it $\pi$. An income stream with high variance that was nevertheless perfectly predictable would not be defined as risky. We therefore need to make assumptions about the basis on which individuals form predictions of their future income stream and, thus, implicitly about the process generating income. We assume that individuals use simple linear predictors. Which variables should be included in the set of variables used to form the predictions (the conditioning variables) is the key question we consider in this paper.

Suppose that the following specification describes the evolution of income for each individual:

$$y_{jt} = Z_{jt} + \beta X_{jt} + e_{jt} \qquad (1)$$

where $y_{jt}$ is the log income of person $j$ at date $t$, $Z$ and $X$ are variables that explain income, and $e_{jt}$ is the unpredictable component, distributed with mean zero and time-invariant variance $\sigma_j^2(e)$. $Z$ and $X$ may include characteristics of the individual and also the macro environment.[6]

If each person $j$ knows the specification in (1), and the values of $Z_{jt}$ and $X_{jt}$ are predictable, then an obvious measure of her income risk is $\pi_j^1 = \sigma_j^2(e)$.[7]

At first glance, an obvious way for an analyst to estimate $\pi_j^1$ would be to estimate specification (1) using regression analysis applied to data about $Z$ and $X$, and to compute the estimated residual error terms. The squared residual for each individual $j$, i.e. the estimate of $\sigma_j^2(e)$, would be the measure of income risk for $j$.

For this procedure to work, there must be suitable data available and it must be legitimate

---

[5] This definition is the standard one in the literature. See *inter alia* MaCurdy (1982), Haveman and Wolfe (1985), Bird (1995), Carroll and Samwick (1995, 1997). 'Income risk' is closely related to 'income mobility'. The main distinction between the concepts is that the former typically refers to income variability from an *ex ante* perspective, whereas the latter refers to variability from an *ex post* perspective. This distinction gets blurred in practice because analysts estimate both from data referring to outcomes rather than prospects.

[6] We explain why we need to distinguish between $Z$ and $X$ below.

[7] Other measures of dispersion could be used.

to condition on $Z$ and $X$. The best data source is clearly longitudinal information for a large sample of individuals describing the intertemporal sequence for each person of (log) income and $Z$ and $X$. Researchers have also estimated (1) using cross-sectional data. In this case each person's income risk is estimated by the personal income deviations from the average among similar persons at a point in time.[8] Contrast this 'cross-section' approach with the 'longitudinal' approach in which each person's income risk is estimated from the deviations from their own inter-temporal income average.[9]

The choice of conditioning variables ($Z$, $X$) for (1) is crucial, regardless of the sources of data, as we shall now explain with reference to the link between predictability and the time-varying character of variables.

## 2.2 Time-varying conditioning variables

Some characteristics of individuals change over time and this needs to be taken into account when computing income risk. Suppose that $Z_{jt}$ is time-invariant (and therefore predictable) and $X_{jt}$ is time-varying and also completely unpredictable. For concreteness, think of $Z$ as gender and $X$ as hours worked per year. Suppose that $X_{jt}$ is distributed with mean $\overline{X}_j$ and variance $\sigma_j^2(X)$.

In this case, $X$ cannot legitimately be included among the conditioning variables (cf. specification 1). Variability in $X$ contributes to variability in the residual error term. We may therefore define a second income risk measure for the case when there are unpredictable time-varying covariates:

$$\pi_j^2 = \sigma_j^2(e) + \beta^2\sigma_j^2(X) + 2\beta\,\mathrm{cov}(e, X). \tag{2}$$

Note two things about this expression. First, the unpredictability of $X$ increases income risk, unless cov$(e, X)$ is large and negative. Second, and more interestingly, the distribution of income risk across individuals implied by (1) – (wrongly) assuming predictability – may be very

[8] Some researchers (e.g. Haveman and Wolfe 1985) using cross-section data have not used regressions to derive income variances. Instead they have classified each individual into a group comprising those with the same values of $Z$ and $X$, and income risk is then the within-cell variance of log incomes. This is equivalent to derivation based on a regression using $Z$ and $X$ and all interactions.

[9] Carroll (1994) labels the methods 'forward-projection' and 'backward-projection' methods rather than 'cross-section' and 'longitudinal'. Haveman and Wolfe distinguish between the calculation of income 'inequality' and income 'uncertainty'.

different to that derived from (2). People may have very different distributions of $X$: individuals with high values of $\sigma_j^2(e)$ may have low values of $\sigma_j^2(X)$.

Allowing for the fact that some potential conditioning variables are time-varying has implications for the measurement of income risk. The key issue is the predictability of $X$. By including $X$ in the baseline income regression – call this the standard case – the researcher is implicitly assuming either that $X$ never changes, or that $X$ does not affect income, or that such forecasts are perfect. That is, if $E(X) = X$, where $E(.)$ is the expectations operator, we can legitimately condition on $X$ in the regression. Hence no risk derives from the variability of $X$ as all such movements are perfectly predictable, and therefore $\pi_j^1 = \sigma_j^2(e)$ is an unbiased measure of risk.

An alternative approach is to suppose that all an individual can do is to forecast $X$ from $Z$, albeit imperfectly. In this case the expected log income for person $j$ is:

$$Ey_{jt} = Z_j + \beta E(X_{jt}|Z_j) = f(Z_j). \tag{3}$$

Hence the unpredictable component of income is:

$$y_{jt} - Ey_{jt} = \beta[X_{jt} - E(X_{jt}|Z_j)] + u_{jt}. \tag{4}$$

Income risk is now given by $\sigma_j^2(u) + \sigma_j^2(\varepsilon) + 2\operatorname{cov}(u,\varepsilon)$, where $\varepsilon$ is the unpredictable component of $X$. To derive this measure, one would simply regress $y_{jt}$ on $Z_j$ and compute the squared residual for each person. Since $Z_j$ is time-invariant (by assumption), $E(X_{jt}|Z_j)] = \overline{X}_j$, and so this method yields $\pi_j^2$ as the income risk measure for person $j$.

It is likely that individuals would be able to forecast better than this because some time-varying variables may be predictable.[10] But these two measures, the one including time-varying variables ($\pi_j^1$) and the other excluding all time-varying variables ($\pi_j^2$) bound the true value of risk. Moreover a comparison of $\pi_j^1$ and $\pi_j^2$ across individuals can be used to analyse the importance of time-varying variables for different groups, as we show below.

[10] Not least by looking at more sophisticated specification of income dynamics (in terms of lagged $X$ or lagged $y$, see below).

### 2.3 Unobservable characteristics and longitudinal data

Some characteristics are observable to the researcher and some are not. Unobservable time-varying influences enter the error term and are unrecoverable, but unobservable time-invariant factors we can deal with.

Let us partition $Z_j$ into observable time-invariant factors, denoted $S_j$, and unobservable time-invariant factors, denoted $W_j$. We rewrite (1) as:

$$y_{jt} = S_j + W_j + \beta X_{jt} + e_{jt}. \tag{5}$$

Clearly, unobservable time-invariant factors are only recoverable with longitudinal data for individuals. With cross-section data, the $W_j$ simply form part of the error term and hence one of the sources of measured risk. This is clearly inappropriate: heterogeneity between people is different from longitudinal income risk for an individual precisely because of these unobservable differences.[11]

The issue of whether time-varying variables are predictable arises regardless of whether panel data or cross-section data are available. Hence we may define a panel data analogue of the earlier measure $\pi_j^2$ in which time-varying variables are assumed to be not predictable:

$$\pi_j^3 = [y_{jt} - E(y_{jt} \mid S_j, W_j)]^2 = (\beta[X_{jt} - E(X_{jt} \mid S_j, W_j)] + e_{jt})^2. \tag{6}$$

Income risk measure $\pi_j^3$ is the best measure of those that we consider in this paper. It conditions on unobservable individual effects and therefore does not erroneously include individual heterogeneity as part of time-series risk. It treats time-varying variables appropriately, allowing for the fact that they may be largely unpredictable.[12] For any given individual, measure $\pi_j^3$ will over-estimate income risk as it excludes other information (for example, lags) in the prediction of $X_{jt}$. However, assumptions made by researchers about the dynamic evolution of income are still likely to be worse than those of the individual, so any estimate of the form of $\pi_j^3$ will overestimate risk.

---

[11] Haveman and Wolfe (1985, pp. 299-300) also make this point.

[12] One can calculate $\pi_j^3$ from panel data either by running a fixed effects regression and squaring residuals (as we do below) or, equivalently, by calculating the longitudinal variance for each person.

### 2.4 The sources of income risk

A lot of the literature implicitly assumes, through sample restrictions or choice of conditioning variables, that the source of income risk is unpredicted variability in returns from the labour market.[13] But the majority of people live in households of more than one person for at least part of their lives, and assuming that there is income pooling within the household, a wider concept of income is relevant.

Household income is very different from the earnings of an individual within a household, and more difficult to model.[14] This is partly because labour supply is now a household decision, but the key addition is the dynamics of household composition. Changes in household composition are by no means rare in their incidence, nor trivial in their effects on income.[15] The second main theme of this paper is to focus attention on the importance of household composition as a source of income risk.

We therefore differentiate several sources of income risk. To be specific, partition the set of time-varying variables relevant to the income generation process into two categories, $X^m$ and $X^d$. (These have coefficients $\beta^m$ and $\beta^d$ in the analogue to specification (1).) For the moment, think of these as 'labour market' variables and 'demographic' variables (we discuss this partition further below). The former include factors such as employment status and macroeconomic conditions; the latter include characteristics such as marital status, and number of adults and children in the household.

We can use the framework developed earlier to evaluate the relative importance of these variables to income risk. We make the following argument with scalar variables $X^d$ and $X^m$; the generalisation to the vector case adds no further insight. Suppose the income generation process is given by:

$$y_{jt} = W_j + \beta^d X^d_{jt} + \beta^m X^m_{jt} + \varepsilon_{jt} \tag{7}$$

where as before $W_j$ is an individual-specific fixed effect and we assume that $E(\varepsilon_{jt}) = \mathrm{cov}(\varepsilon_{jt}, X^d_{jt}) = \mathrm{cov}(\varepsilon_{jt}, X^m_{jt}) = 0$.

---

[13] See *inter alia* Carroll and Samwick (1995, 1997) or Miles (1997).

[14] See Jenkins (1999) for a survey and Burgess and Propper (1998) for a structural model.

[15] For example, Jenkins (1999, Table 6) shows that 10 percent of wave 1 respondents to the BHPS had a different household head by wave 2 (more than one fifth by wave 6). On the movements into and out of poverty associated with demographic changes, see e.g. Bane and Ellwood (1986) for the US, or Jenkins (1999) for Britain. The income changes associated with marital splits are described by Burkhauser et al. (1990) and Jarvis and Jenkins (1999).

Assume first the individual cannot predict the movements in either $X^d_{jt}$ or $X^m_{jt}$, knowing only her time means $\overline{X}^d_j$ and $\overline{X}^m_j$. In this case, the squared deviations from her expected income are:

$$\begin{aligned}\left[y_{jt} - E(y_{jt})\right]^2 &= \left[\beta^d (X^d_{jt} - \overline{X}^d_j)\right]^2 + \left[\beta^m (X^m_{jt} - \overline{X}^m_j)\right]^2 \\ &\quad + 2\beta^d \beta^m (X^d_{jt} - \overline{X}^d_j)(X^m_{jt} - \overline{X}^m_j) + \varepsilon^2_{jt}\end{aligned} \tag{8}$$

Now assume instead that the individual knows $X^d_{jt}$ exactly, but remains ignorant of $X^m_{jt}$. Expectations of $y_{jt}$ are now taken conditional on $X^d_{jt}$. Squared deviations from her expected income are now:

$$\left[y_{jt} - E(y_{jt} | X^d_{jt})\right]^2 = \left[\beta^m (X^m_{jt} - E(X^m_{jt} | X^d_{jt}))\right]^2 + \varepsilon^2_{jt} \tag{9}$$

Note that terms in $(X^d_{jt} - \overline{X}^d_{jt})$ do not appear on the right hand side of (9) as $X^d$ is known. Furthermore, knowledge of $X^d$ may help in forecasting $X^m$; this also reduces income risk. These two components are, respectively, the direct and indirect effects on risk of knowledge of $X^d$.

We define the difference between (8) and (9) as the contribution to income risk of $X^d$, and we denote this by $\Delta^d_{jt}$.

$$\begin{aligned}\Delta^d_{jt} &= \left[\beta^d (X^d_{jt} - \overline{X}^d_j)\right]^2 + \left\{\left[\beta^m (X^m_{jt} - \overline{X}^m_j)\right]^2 - \left[\beta^m (X^m_{jt} - E(X^m_{jt} | X^d_{jt}))\right]^2\right\} \\ &\quad + 2\beta^d \beta^m (X^d_{jt} - \overline{X}^d_j)(X^m_{jt} - \overline{X}^m_j)\end{aligned} \tag{10}$$

where the second set of terms on the right hand side of (14) is the indirect gain from knowing $X^d$.

To simplify (10) we need to formulate the relationship between $X^d$ and $X^m$. We assume a simple linear form, $X^m_{jt} = \gamma + \delta X^d_{jt} + u_{jt}$ with $E(u_{jt}) = \mathrm{cov}(u_{jt}, X^d_{jt}) = 0$. Using the fact that $(X^m_{jt} - \overline{X}^m_j) = \delta(X^d_{jt} - \overline{X}^d_j) + u$ and taking expectations for $j$ of $\Delta^d_{jt}$ over $t$,

$$E(\Delta^d_{jt}) = \Delta^d_j = \beta^{d\,2} \sigma^{d\,2}_j + \beta^{m\,2} \delta^2 \sigma^{d\,2}_j + 2\beta^m \beta^d \delta \sigma^{d\,2}_j \ .$$

Simplifying,

$$\Delta^d_j = \sigma^{d\,2}_j (\beta^d + \beta^m \delta)^2 \tag{11}$$

This formulation is intuitively plausible. The contribution a variable makes to income risk depends on its variability ($\sigma^{d\,2}_j$) and the impact, direct and indirect, it has on income ($\beta^d + \beta^m \delta$). The former varies from individual to individual; the latter in our empirical work we have assumed constant over individuals, but it is clearly a straightforward extension to allow ( $\beta^d + \beta^m \delta$ ) to vary between sub-groups of the population.

Symmetrically, for the case where $X^m{}_{jt}$ is known but not $X^d{}_{jt}$ we have:

$$\Delta_j^m = \sigma_j^{m^2} (\beta^m + \beta^d \tilde{\delta})^2 \tag{12}$$

where $\tilde{\delta}$ is the coefficient resulting from a regression of $X^d$ on $X^m$.

In the empirical section below, we calculate both $\Delta_j^d$ and $\Delta_j^m$. We label the first demographic risk and the second labour market risk. We are interested in both the levels of these two measures for any one individual, and in the differences in their distribution across individuals (all of the measures are individual-specific).[16] In other words, we attempt to provide answers to two different questions: (i) how much of each person's risk is accounted for by demographic and labour market factors, and (ii) how important demographic and labour market risk is for different individuals (for example at different points in the life-cycle).[17]

Three points need to be made clear at this point. First, clearly it is not possible to produce a definitive unambiguous partition of all explanatory variables into the two categories. Different researchers may allocate factors differently (if only because of different views about whether a variable is fixed and predictable). What we are investigating is what is the reduction in risk when it is assumed one set of variables is known to the individual and the other set is not known, allowing for correlation between the sets of variables. Second, and related, the economics of individual and household decision making suggests that most decisions about either demographic and labour market matters depend on both sets of factors. For example, labour supply may depend on marital status and a decision to divorce may depend on earnings. But since the linear predictors that we (and the rest of the literature) adopt are simply reduced forms, the *process* by which income changes does not matter. We are simply asking what happens to income in the light of knowing or not knowing the variables that determine income. Third, this is a study of income risk and not simply of (adverse) events (cf. Bane and Ellwood, 1986) in that we do not allocate income changes to particular events. Rather, we relate income risk to variability in both

---

[16] Note that $\Delta_j^d$ and $\Delta_j^m$ are not constructed from $\pi_j^3$ and do not provide an additive decomposition of $\pi_j^3$.

[17] There are other ways of approaching this second question. For example, we could regress income risk upon demographic and labour market factors and examine the extent to which these factors accounted for the variation in income risk across individuals.

labour market and demographic variables.

### 2.5 Individual income and household income

It is a person's household income (rather than their employment income or wages) which is of most interest for income risk, as we noted above. We now turn to consider more explicitly the implications of making this distinction.

Household income is a function of the characteristics of the focal individual *j* and also of other members of her household, denoted *k*. We expand (5) accordingly:

$$y_{jt}^{h} = S_j + S_k + W_j + W_k + \beta X_{jt} + \beta X_{kt} + e_{jt}. \tag{13}$$

Similar logic could then be applied to this equation to derive analogues of $\pi_j^1$ to $\pi_j^3$.

But we cannot follow households over time, only individuals, because households split and re-form (Duncan and Hill, 1985). Individuals move between different households, and it is the individual that is the unit of analysis. Our argument is precisely that changes in a person's household composition mean that the existence and identity of any other members of the individual's household are not fixed features. Thus while some factors, for example the ethnic origin of one's spouse, are time-invariant for the spouse, the race of her partner is a potentially time-varying characteristic as far as the focal individual *j* is concerned.

We therefore return to the income risk measures set out earlier, noting that *all* the characteristics of any other individuals presently in the household of person *j* can be included as time-varying characteristics in the specification describing the income process for person *j* (i.e. included in $X_{jt}$). In this case the expectations of the form $E(Z_k|Z_j)$ in the specification of $\pi_j^i$ have economic meaning. They arise as the outcome of endogenous household formation whereby individuals form households on the basis of their characteristics (assortative mating being a leading example).

### 2.6 Dynamic specification issues

A number of papers in the literature on income risk focus much attention on the longitudinal covariance structure of income (or earnings) in the specification corresponding to (1). This raises the issue of whether to use lagged income or lagged $X_{jt}$ when forecasting current

$X_{jt}$. We have ignored this issue for a number of reasons.

First, the aim of this paper is to draw attention to two other issues. One is the distinction between approaches which condition on time-varying variables (the standard one) and approaches which condition only on time-invariant variables, and the other issue is the role of demographic change in generating income risk. For neither of these issues is a time series structure crucial for the specification of the income model. As we argued earlier, $\pi_j^1$ and $\pi_j^2$ bound the true value of risk. Second, autoregressive moving average (ARMA) models of the type typically employed are unlikely to fit well the discontinuous changes observed in *household* income (because, for example, demographic changes are associated with large discrete changes in income: see Jenkins, 1999). Third, the panel data we use are drawn from a short-length panel (the BHPS began in 1991), which makes the identification of ARMA models difficult. Fourth, part of the argument we make relies on mimicking the use of cross-section data, and clearly these have no lagged information to exploit.

**2.7 Previous literature on income risk**

Almost all of the literature on income risk starts from a quantitative model of income. (One exception is Guiso, Jappelli and Terlizzese (1993, 1996) who have access to qualitative survey data about people's subjective beliefs about their income in the following year.) Our particular interest is in the sets of conditioning variables and sample selection criteria which researchers have used. Table 1 illustrates the range of choices adopted in the literature. We briefly discuss a few of these articles relevant to the two themes of this paper, namely the selection of factors to condition on, and the role of demographic factors.

<Table 1 near here>

In terms of the decision on whether to use only an individual's fixed characteristics as conditioning variables, or a full set of factors, the range can be illustrated by comparing Miles (1997) on the one hand with Haveman and Wolfe (1985), Kazarosian (1997), Jarvis and Jenkins (1998) on the other. The conditioning variables in the latter group of papers include only gender, age, and (in Haveman and Wolfe's case), race and disability status also. Miles, by contrast, conditions on both time-varying and non-time-varying variables, including region, sex of the head of household, mean age of adults in the household, number of adults in the household, number of children, occupational status of the head, marital status of the head and number of working people in the household. So employment and demographic states are taken as fixed and predictable. His analysis was based on cross-sectional data from the UK Family Expenditure Survey. Observe that, even using the most appropriate set of conditioning variables, a cross-sectional approach cannot properly assess the degree of income risk since (by contrast with a longitudinal approach) one cannot control for the income differences which are unobserved (but predictable) individual fixed effects.

Demographic factors are typically dealt with in the literature by using demographic conditioning variables or by sample exclusions (see Table 1). An example of the former approach is the work of Dynarski and Gruber (1997) who condition on both levels and changes in household characteristics, thereby completely eliminating any possibility for demographic risk to contribute to their measure of income risk. A representative example of the latter approach is the set of papers by Carroll and Samwick (1995, 1997). Their empirical analysis using the US Panel Study of Income Dynamics (PSID) excludes households from the low-income subsample, households whose head is aged over 50 or below 26 years, households where the head changes over the seven year observation period, households where the marital status of head changes, and households where head is not in the labour force in 1981. Again demographic risk can hardly contribute to measured income risk. Miles (1997) excludes households whose head is unemployed or retired, and households containing adults other than a single person or couple.

Arguably such sample exclusions simply restrict the population to which the results generalise. But given the pervasiveness of household change, this approach produces a very partial picture of income risk. Significant sources of household income risk are systematically removed, potentially biasing the results for the remaining sample.[18] The risks of unemployment and household flux do not just relate to a separate minority of the population but are relevant for the majority. In much the same way as it is not correct to condition on time-varying characteristics, it is also not appropriate to exclude sections of the sample on the basis of time-varying factors such as labour market status or whether household composition remains unchanged. Of course, the sample exclusions on age grounds also mean that these studies do not

[18] Samwick (1994, p. 144) argues that ‘it would be inappropriate to treat changes in income associated with changes in the household as a reflection of income uncertainty’. This confuses the source of the income risk and the risk itself.

enable us to say anything about the income risk of large sections of the population.

### 2.8 Summary

To summarise so far, we have argued, first, that income risk arising from unpredicted variability in demographic factors has been neglected in previous research (and there is no *a priori* reason to assume that income unpredictability associated with demographic factors is less than that associated with labour market factors) and, second, that it is inappropriate to use time-varying (and hence unpredictable) variables as covariates in models used to derive measures of income risk. These variables, typically summarising demographic and labour market factors, should be treated as sources of income risk.

To illustrate these points we have defined different measures of income risk, $\pi_j^1$ through $\pi_j^3$, whose features are summarised in Table 2. (The far right-hand column is discussed in Section 4). In the remainder of the paper, we estimate examples of each of these income risk measures, and compare both their average levels and their distribution across the population. We also contrast the relative contributions of demographic and labour market risk to income risk.

<Table 2 near here>

## 3. DATA AND DEFINITIONS

Our arguments are illustrated using data drawn from the first six waves of the British Household Panel Survey (BHPS) covering 1991-96. The first wave of the BHPS was designed as a nationally representative sample of the population of Great Britain living in private households in 1991, and had an achieved sample size of some 5,500 households covering some 10,000 persons. See Taylor A (1996) and Taylor M (1998) for detailed information about the BHPS.

On-going representativeness of the (non-immigrant) population has been maintained by using a ‘following rule’ typical of household panel surveys: at the second and subsequent waves, all original wave 1 sample members (OSMs) are ‘followed’ (even if they move house, or if the household splits up), and there are annual interviews with all adult members of all households containing either an OSM, or an individual born to an OSM whether or not they were members of the original sample. New panel members who subsequently stop living with an OSM are, however, not followed and interviewed again. Thus, for example, if a non-OSM married an OSM at wave 2, and the partnership subsequently dissolved, the OSM is followed, but the non-OSM is not.

For the most part, our analysis sample includes all adults who have income information from at least four interviews. It includes children of OSMs who turn 16 in the course of the panel and who get interviewed in their own right. Thus, our sample is not balanced—the number of observations per individual ranges from four to six.

Our sample selection criteria are as follows. Of the 17,626 individuals who have a value for household income in at least one BHPS wave we have 77,067 observations of household incomes, an average of 4.4 income observations per individual. We then excluded the 16,782 household income observations for people who are aged under 16 (but income information for these individuals from later waves is used as long as they are aged 16 or above).[19] A further 13,326 income observations were excluded where any person in the household had missing income information, or because the sum of individual incomes over the household members did not tally with the household income variable. Another 8,323 income observations were dropped for individuals with fewer than four observations each. This results in a base sample of 38,636 income observations for 7,079 adults, an average of 5.6 income observations per person.

Separate subsamples from this base sample are used here to illustrate the differences between the longitudinal and cross sectional approaches to measuring income risk. The longitudinal analysis excluded a further 7 adults who had missing values for sample weights. The resulting longitudinal sample contains 7,072 individuals and 38,636 income observations. The cross sectional analysis was based on the 6,180 adults with an income observation in wave 1.

Some of our calculations have been weighted using the BHPS sample weights. Weights have been applied to the estimates of averages of income risk for the population, where it is important to adjust for differential non-response. Cross-sectional enumerated individual weights for wave 1 were used for the cross-sectional analysis. However there was not an appropriate set of weights for our longitudinal sample: the BHPS provides longitudinal weights only for individuals who remain continuously in the panel from wave 1. (Hence, for example, there would

[19] We have not made any age selections apart from this, which contrasts with previous literature. This is intentional—we can then compare income risk for the groups conventionally excluded from such analysis with those who are included (see below).

be no longitudinal weight for an individual who was present in only waves two to five.) For the longitudinal analysis the data each for each person was weighted using the cross-sectional enumerated individual weight for the most recent wave for which they are present in the sample.

We work with definitions of household income which have been commonly used in related research. We analyse the log of household income, where household income comprises the sum over all household members for the month prior to interview of all sources of income, and has been adjusted for inflation.[20] Income includes earnings from employment and self-employment, cash social security and social assistance benefits (including state retirement pensions and housing benefits), private transfers (such as child support receipts) and income from savings and investments including private and occupational pensions. In UK jargon this is a 'gross income' definition; in US jargon it is a 'pre-tax post-transfer income' definition.

Income has not been adjusted for differences in household size and composition ('needs') using an equivalence scale. This was because we wish to examine income risk *per se*. Measures of income risk based on equivalised income are affected by changes in money income (the numerator) *or* changes in the number of equivalent adults (the denominator) *or* both. We argue that it is more transparent to start with measures of risk based on unequivalised income and then go on to look at the factors which are associated with that risk, including demographic change. If instead we had calculated risk of equivalised income, we would have already to some extent adjusted for demographic factors.

That said, equivalent income is used in one part of our analysis, namely to define a classification of sample individuals into quintile income groups to use for breakdowns of the derived income risk measures. The idea was to capture some notion of 'permanent living standard'. Equivalised household income for each person was calculated by averaging the wave-specific equivalent incomes for each person over the waves for which s/he was present in the sample. This was done prior to the exclusion from the sample of information for individuals with fewer than four income observations, to make the income classes more representative of the population as a whole. The equivalence scale used was the semi-official McClements Before Housing Costs scale (Department of Social Security, 1998).

[20] A household is defined to be one person living alone, or a group of persons who either share living

## 4. RESULTS

Our results are in two parts. First, we present a comparison of the implications for income risk measures of using different conditioning sets. We examine how this choice affects the predicted distribution of income risk according to characteristics across the sample, with a particular focus on the results produced by our preferred income risk measure, $\pi_j^3$. Second, we decompose this measure to investigate the relative contribution of demographic and labour market factors.

### 4.1 The impact of choice of conditioning variables

The purpose of this section is to examine whether the conceptual distinctions concerning income risk measurement, drawn in Section 2, will lead to significant differences in empirical estimates of income risk when applied to real data. In particular, we look at the impact of choice of conditioning variables. Above we have proposed that it is only appropriate to control for fixed characteristics, not varying, but this should include both observed and unobserved (since both are known to the individual).

What is considered time-varying and what is not by the individual cannot be detected by the researcher, and for the researcher what may be considered fixed and what not depends on the length of the observation window provided by the available data source. In cross-sectional data (i.e. where the window is of length one) factors which are time varying cannot be distinguished from those that are not. In our analysis, the distinction between 'fixed' and 'time varying' is rather based on some general notion of which characteristics of an individual and their household are likely to be predictable, so, for example, the individual's education and age are taken as fixed but the occupation of adults in the household is not. On the second substantive issue, it is only using panel data, with repeated observations on the same individuals, that we have any way to determine differences between individuals which may relate to unobserved fixed characteristics.

To systematically examine the effect of employing different sets of control variables, we constructed three measures of income risk, each one being the variance of the residuals from regressing every individual's household income on a different set of covariates. Two estimates are derived from the data used as a cross-section. The first, $\pi^1$, utilises the full set of controls

accommodation or one meal a day, and who have the address as their only or main residence.

listed in the far right-hand column of Table 2; the second, $\pi^2$, controls only for age, sex, region and education of the individual. Hence, the comparison here is between two sets of regressors which are both restricted to observables but where one has both fixed and time-varying ($\pi^1$) but the other has only invariant ($\pi^2$) characteristics. The third estimate, $\pi^3$, conditions on the observable and unobservable fixed characteristics of the individual using panel data to estimate a fixed effects regression. By comparing the results for $\pi^1$ with those for $\pi^2$ we can examine the effect on the empirical estimates of (we would argue wrongly) including controls for time varying factors. The comparison of $\pi^2$ and $\pi^3$ indicates whether there is any impact empirically from omitting controls (again, we would argue wrongly) for the unobserved characteristics of the individual. Hence, out of the three measures, $\pi^3$ represents our ideal measure in terms of the set of conditioning variables used.[21]

Table 3 presents the average values (medians and interquartile ranges) of the estimates of $\pi^1$, $\pi^2$, and $\pi^3$, for all individuals in the sample and also for selected subgroups. The estimates of income risk derived from a model that conditions only on fixed and observed effects ($\pi^2$) is significantly larger than the other two estimates. This is as expected: by construction the fewer the conditioning variables that are used, the higher the variability of the estimated residual. Over the whole sample, average estimated income risk is similar for $\pi^3$ and $\pi^1$, roughly a third of the size of the estimates for $\pi^2$. So the effect of omitting controls for unobservable fixed effects is to overestimate the level of risk but then the result of adding extra controls for time varying characteristics is to underestimate risk. The similarity of $\pi^3$ and $\pi^1$ is co-incidental: there is no reason to believe that this will occur as a rule.

<Table 3 near here>

It is clear that there are considerable differences in income risk across population groups, here broken down on the basis of sex, age group and household income quintile group (see Section 3 for the definition of the income groups). For example, the levels of $\pi^1$ and $\pi^3$ are similar over the whole sample but if we compare the figures for, say, the rich male group the $\pi^3$ median is about a third of the $\pi^1$ median. For this same group the $\pi^2$ median is about seven times the $\pi^3$ figure but for the group of well-off women aged 31 to 50, the equivalent ratio is

[21] Regression results to derive these three measures of risk are available from the authors.

about eleven.  This suggests that the impact of conditioning for invariant unobservables (in $\pi^3$ but not $\pi^2$) is not uniform over the whole population.

The degree of income risk for all men and all women does not differ greatly for any of the three measures.  To some extent this is not surprising since multi-adult households are most often households in which adult men live with adult women. Although personal income variability may differ for men and women, *household* income variability need not.

Table 4 summarises the heterogeneity in income risk across the population in greater detail by regressing each risk measure on the age, income group and sex of each individual.[22]  In order to be able to compare coefficients across columns, the estimates of income risk for each person are normalised by the sample average (i.e. $\pi_j^1$ divided by the mean for all individuals, and similarly for $\pi^2$ and $\pi^3$). Since Table 3 has already illustrated how *absolute* average levels can vary across the three measures, the purpose of Table 4 is to bring out how the *relative* associations between risk estimates and these selected characteristics differ.

<Table 4 near here>

The table provides further evidence that the investigator's decision about which factors are predictable and which are not will result in different patterns of estimated risk over the population. According to measure $\pi^2$, derived by conditioning on fixed observable characteristics only, women's income risk appears to be similar to men's, other things equal (the estimated coefficient on the dummy variable for females is not significantly different from zero).  But if instead one uses a measure of income risk which also conditions on time varying characteristics such as those for other household members (measure $\pi^1$), or unobservable differences calculated from longitudinal data ($\pi^3$), then women's income risk is estimated to be less than that of men. Clearly, the association between gender and income risk depends upon which type of variables are included in the conditioning set.

The pattern of variation in income risk with age is quite similar for the two cross section data-based measures ($\pi^1$and $\pi^2$) which both omit the unobservable fixed effects: other things equal, individuals aged 51-65 years are the group with the highest income risk, whereas individuals over 65 experience lower income risk than all other age groups.  The pattern of

[22] These regressions are simply a descriptive device to illustrate the relationship between income risk estimates and individual characteristics, holding other variables constant.

results by age for measure $\pi^3$ present quite a different picture, bringing out the effect of taking account of unobservable fixed effects. Whilst the $\pi^3$ estimates also show the oldest age group to have the lowest income risk, it is now the youngest age group who have the highest income risk; the middle age brackets have relatively less.

Table 4 also indicates significant differences in estimates of income risk by household income group. The cross-sectional measures suggest that that the relationship between household income and income risk is roughly U-shaped: values for the middle quintile groups are less than those for the lowest and highest quintile groups. According to the longitudinal measure $\pi^3$, the pattern of variation in income risk by income group is somewhat different: risk declines with income. Also, in general, the extent of heterogeneity across income groups is smaller for measure $\pi^3$ than for $\pi^1$ or $\pi^2$.

Even comparing the two cross-sectional measures $\pi^1$ or $\pi^2$, there are significant differences in the patterns of income risk. To capture this we calculated the *difference* between the (normalised) values of $\pi^1$ and $\pi^2$ for each individual and regressed this difference on age, sex and income group (results not shown) and found that the two estimates differ significantly by all three characteristics. Measure $\pi^1$ provides estimates of income risk which are higher for older age groups, lower for females, and higher for the top income groups than measure $\pi^2$.

The results indicate that assumptions about the predictability of both time-varying and unobservable fixed characteristics have implications not only for the estimates of the amount of income risk, but also the distribution of risk across different groups in the population. They illustrate the empirical import of the conceptual points presented in Section 2: that a measure of income risk should not control for time-varying factors (thereby treating them as predictable), but should take account of unobserved individual fixed effects.

Given our case for $\pi^3$ being closest to an ideal measure, it is worth summarising the findings about predicted risk. The results for $\pi^3$ suggest that riskiness of household income is slightly lower for women compared to men.[23] With other things held constant, risk is much lower for those of post retirement age and highest among the under 30's than other age groups.

[23] This finding comes from Table 4 which estimates the associations between income risk and characteristics, holding other things constant. This is distinct from the subgroup averages in Table 3 which show that the median risk for all women is higher than for all men. The reason for this apparent discrepancy is that the Table 3 figure does not control for any age and income differences between the men and women in the sample.

There appears to be a negative relationship between household income quintile group and risk, so that estimated risk tends to be lower for individuals in richer households.

### 4.2 The contribution of demographic and labour market factors to total income risk

We now turn to the contribution of demographic and labour market factors to income risk. For each individual we examine how much of the variability in their household income can be related to changes in demographic characteristics (what we have termed demographic risk) and how much to changes in the labour market characteristics of their household (labour market risk). We compute individual measures of $\Delta_j^d$ and $\Delta_j^m$, defined using (11) and (12) respectively. In calculation we substitute the scalars $X^d$ and $X^m$ used in section 2.4 above with vectors, where the demographic and labour market factors are defined as in Table 2.[24] This gives the level of risk for each individual that can be attributed to demographic and labour market factors.

It is also of interest to examine the ratio of demographic and labour market risk to total risk. By expressing each of these amounts as a proportion of total risk, $\pi^3$, for each adult we can define a "demographic risk ratio" and a "labour market risk ratio". As the measures of demographic risk and labour market risk are not exact additive decompositions of $\pi^3$ our estimates of this ratio can be greater than 1.

Table 5 summarises the estimates of income risk according to these five measures: $\pi^3$, demographic risk, $\Delta^d_j$, labour market risk, $\Delta^m_j$ and the two ratios, showing the median values for all persons and among our selected subgroups. Taking all individuals together, median demographic risk is about half that for labour market risk. When the absolute amounts of each type of risk are expressed as a share of the total, the median values for the two ratios show that demographics account for around 7% of total risk, compared to 12% for labour market factors. This result indicates the importance of demographic factors as a source of income risk.

<Table 5 near here>

Although women's income risk is of a similar magnitude to men's, Table 5 shows that men have both more demographic and labour market risk than women (at the median). Table 5 also gives estimates for subgroups to give a feel for how much average levels of risk of the different types can vary across the population. Among men and women, the poorest fifth have two to three times the level of demographic risk of the richest fifth. For the groups shown here, the differences in median amounts of labour market risk are even more dramatic. For example, the poor female group has more than 16 times the level of labour market risk of the richer group of women. The results for the ratios indicate whether the variation in amounts of demographic and labour market risk across groups accords with differences in amounts of total risk. In general, some of the variation across groups indicated by the levels of risk is removed, so higher levels of demographic and labour market risk are partly about having more total risk.

Table 6 explores this further and summarises the variation in risk by regressing demographic risk, labour market risk and the two ratios (all measures normalised to allow comparison across columns) against age, sex, and income group. Demographic risk falls as age increases. This means that, as for total risk, the level of demographic risk is greatest for the youngest age group and smallest for the oldest, but the gap is even greater and the youngest age group stands out as having markedly more demographic risk than all other age groups. Men and women appear to experience similar amounts of demographic risk, other things held constant. The pattern of demographic risk by household income group is very distinct from the total risk picture. Demographic risk is highest for the second poorest fifth, then declines with income and so is lowest for the top and bottom income classes (where the difference between them is small and not significant).

Labour market risk is also largest for the youngest age group and smallest for the oldest age group. But individuals approaching retirement, those aged 51-65, have income risk of similar magnitude to the under-30s. What is distinct from the age distribution of demographic risk, where the largest gap was between the youngest and the others, is that labour market risk most noticeably differs for the oldest group versus the rest. This is as we might expect. The level of labour market risk is not significantly different between males and females but this may well be a function of the focus on riskiness of *household* income. The way labour market risk varies by income quintile group is similar to total risk - risk declines as household income improves - but with an even stronger gradient.

<Table 6 near here>

Table 6 also presents results for the demographic risk ratio and labour market risk ratio,

---

[24] The estimates of the vectors $\beta^m$, $\beta^d$, and $\delta$ used to compute these terms are available from the authors.

and illustrates how the *share* of total risk which is accounted for by demographic and labour market factors is associated with the individual's age, gender and household income quintile group.[25] It is interesting to compare the findings from the levels and ratios. Looking first at the demographic risk ratio, the pattern by age is not greatly affected by controlling for total risk although the variation across age groups is reduced. Gender remains insignificantly associated with the demographic risk ratio. However, the distribution of demographic risk by income group differs considerably between the ratio and the absolute level. The results for the ratio show that the *relative* amount of demographic risk actually rises with household income so that for those in the richest fifth, relatively more of their risk is related to demographic factors than for other income classes.

The results for labour market risk are quite different for the ratio compared to the absolute level. The youngest group has a smaller ratio than all other ages i.e. relatively less of their total risk is linked to labour market characteristics. The labour market risk ratio is significantly higher for the 51-65 age groups. Age is the most important characteristic in terms of variation in the labour market risk ratio. Gender is now statistically significant for the share of labour market risk, with women having relatively less of their risk related to labour market factors than men. Interestingly, the ratio of labour market risk to total risk does not vary significantly across income groups.

To summarise the risk ratio findings, the role of demographic factors falls with age but rises with household income. In the case of labour market factors, the related share of risk is smallest for the youngest age group and largest for the 51 to 65 year olds; the association with income group is insignificant.

In terms of the relative importance of demographic risk compared to the traditional focus of labour market risk, Table 5 has shown that the overall median value for demographic risk is about half that for labour market risk, and that for each of the selected subgroups demographic risk is below labour market risk. However, this hides the greater importance of demographic risk for a significant proportion of the sample and other aspects of the distributions of the two types of risk. These issues are explored further in Table 7 which groups observations according to the relative amount of the two types of risk. The first row of the table indicates that over a third of

[25] Note that the shares do not sum to 1.

the individuals in the sample have a higher level of demographic risk than labour market risk. Around half the observations experience relatively greater labour market risk and about one seventh of adults have equal amounts of the two types of risk (due to these individuals having zero values for both risks).[26] So while the aggregate picture may suggest that demographic factors are a less important source of risk than labour market factors, heterogeneity within the sample means that this does not apply to all individuals.

< Insert Table 7 near here>

The lower part of Table 7 gives the age, sex and income group breakdowns for the same three groups, and the whole sample for comparison. Those with greater demographic risk than labour market risk are more likely to be in the two youngest age groups and, to a lesser extent, the richest two income classes. It is interesting to see how this contrasts with the results in Table 6. This showed that younger people have the highest levels of both types of risk, compared to other age groups. But there is relatively less variation by age in the amount of labour market risk so this is consistent with the finding that significant numbers of young people experience relatively more demographic risk. In the case of the high income groups, the reasoning is similar: Table 6 points to both types of risk being lowest for the richest individuals but the variation by income is larger for labour market risk. Hence it follows that some of the well-off may end up with more demographic risk.

What seems to be distinctive about the individuals with more labour market risk than demographic risk compared with the whole sample, is that they are aged 51-65. This matches the picture from Table 6 for this age group who are found to have less demographic risk than all but the oldest group but the highest levels of labour market risk, along with those aged 30 and under.

As argued earlier, demographic risk is closely associated with household formation and dissolution. We should therefore expect to see a difference between the contribution of demographic factors to total income risk for people from households where household composition has changed than for people whose household composition has not changed. We define an intact household to be one in which the head of the household and their marital status remained the same over the sample window. Table 8 presents the estimates of income risk for

[26] This group are older and poorer than the rest of the sample. To the extent that older individuals are more likely to exit the household survey when they experience demographic change, the data here may be an upwardly biased estimate of the extent of lack of demographic or labour market change in the older age group.

individuals in intact and non-intact households. Average labour market risk is the same across the two household types, but for those in non-intact households median demographic risk is five times the level for members of intact households (0.010 compared to 0.002). We know that those adults from non-intact households have higher total risk, on average, so it is useful to see how the *share* of total risk related to demographics differs by household type. The median value for the demographic risk ratio is 0.128 for people from non-intact households, almost four times the figure for intact household members. Hence, not only do individuals from non-intact households experience higher demographic risk, this also represents a substantially greater share of their total risk than for other people.

<Table 8 near here>

We would like to know whether this difference in average risk between the household types disappears after control for observed characteristics such as age and income, so we regressed $\pi^3$, the levels of demographic and labour market risk, and the two ratio measures, against age, sex, income groups, and whether or not a person is from an intact household. The analysis is the same as that presented in Table 6, except for the addition of an extra regressor, intact/non-intact household membership. The size and significance of the coefficients on the repeated regressors are not notably affected by this addition and hence in Table 9 we only present the calculated coefficients for the extra household type variable. The results show that, holding other things constant, levels of total risk, demographic risk and labour market risk are all significantly higher for individuals in non-intact households compared to those from intact households.

<Table 9 near here>

## 5. CONCLUSIONS

Measures of income risk are in increasing use in several fields of economics. We believe that it is time for a reassessment of how these measures should be derived. In this paper we have argued that measures of risk which are derived from cross-sectional data are inappropriate. Cross-sectional data does not allow the researcher to differentiate between income risk and heterogeneity. Deriving estimates from cross sectional data by conditioning on a wide range of labour market and demographic characteristics will not identify individual fixed effects.

Furthermore the tendency to condition upon demographic effects, common in the literature on income variability, incorrectly identifies these characteristics as fixed. But for individuals who experience household change, such characteristics are not fixed.

We estimate income risk using both cross sectional and panel data for the UK and find that the two types of data produce quite different estimates of the distribution of income risk. For example, the panel data estimates indicate that risk declines as income rises whilst the cross sectional estimates that condition on a wide set of labour market and demographic characteristics suggest the relationship with income is non-linear. Panel estimates indicate that risk falls with age, whilst cross-sectional estimates indicate the relationship is non-linear. Measures of the latter type are often used to as covariates in models of consumption and savings decisions. Incorrect identification of individual fixed effects from cross sectional differences in demographic characteristics, and inappropriate decisions about what elements of income the individual is likely to be able to predict, will produce misleading estimates of risk and therefore poor estimates of consumption and savings decisions.

In contrast with much of the literature, we have argued that demographic factors are best treated as a source of income risk rather than conditioning variables used to derive a particular measure. We derive estimates of the amount of risk associated with demographic and labour market factors (allowing for correlation between the two sets of factors) and find that demographic risk is associated with a considerable proportion of total income risk. Further, for some groups the contribution of demographic factors to risk is more important than that of labour market factors. The latter set has been the conventional focus: we argue that it is time to re-assess this and to treat demographic change as a legitimate source of household income risk.

## REFERENCES

Bane MJ, Ellwood DT (1986) Slipping into and out of poverty: the dynamics of spells. Journal of Human Resources 21: 1-23

Banks J, Blundell R, Brugiavini A (1994) Income uncertainty and consumption growth in the UK. Unpublished paper, Institute for Fiscal Studies, London

Banks J, Blundell R, Brugiavini A (1999) Risk pooling, precautionary saving and consumption growth. Working Paper W99/19, Institute for Fiscal Studies, London

Bird EJ (1995) An exploratory comparison of income risk in Germany and the United States. Review of Income and Wealth 41: 405-426

Bird EJ, Hagstrom PA (1999) The wealth effects of income insurance. Review of Income and Wealth 45: 339-352

Blundell R, Preston I (1998) Consumption inequality and income uncertainty. Quarterly Journal of Economics 113: 603-640

Burgess S, Propper C (1998) An economic model of household income dynamics, with an application to poverty dynamics among American women. CEPR Discussion Paper No. 1830, Centre for Economic Policy Research, London

Burkhauser RV, Duncan GJ, Hauser R, Berntsen R (1990) Economic burdens of marital disruptions: a comparison of the United States and the Federal Republic of Germany. Review of Income and Wealth 36: 319-333.

Carroll CD (1994) How does future income affect current consumption? Quarterly Journal of Economics 109: 111-148

Carroll CD, Samwick AA (1995) How important is precautionary saving? NBER Working Paper No. 1594, National Bureau of Economic Research, Cambridge MA

Carroll CD, Samwick AA (1997) The nature of precautionary wealth. Journal of Monetary Economics 40: 41-71

Dardanoni V (1991) Precautionary savings under income uncertainty. Applied Economics 23: 153-160

Department of Social Security (1998) Households below average income 1979-1996/7. Corporate Document Services, London

Duncan GJ, Hill MS (1985) Conceptions of longitudinal households: fertile or futile. Journal of Economic and Social Measurement 13: 361-375.

Dynarski S, Gruber J (1997) Can families smooth variable earnings? Brookings Papers on Economic Activity 1:1997: 229-303

Gottschalk P, Moffitt R (1994) The growth of earnings instability in the U.S. labor market. Brookings Papers on Economic Activity 2:1994: 217-272

Guiso L, Jappelli T, Terlizzese D (1992) Earnings uncertainty and precautionary saving. Journal of Monetary Economics 30: 307-338.

Guiso L, Jappelli T, Terlizzese D (1996) Income risk, borrowing constraints, and portfolio choice. American Economic Review 86: 158-172

Haveman RH, Wolfe BL (1985) Income, inequality and uncertainty: differences between the disabled and nondisabled. In: David MH and Smeeding TM (eds) Horizontal inequity, uncertainty, and economic well-being, University of Chicago Press, Chicago

Jarvis S, Jenkins SP (1998) How much income mobility is there in Britain? Economic Journal 108: 428-443.

Jarvis S, Jenkins SP (1999) Marital splits and income changes: evidence from the British Household Panel Survey. Population Studies 53: 237-254

Jenkins SP (1999) Modeling household income dynamics. ESRC Research Centre on Micro-Social Change Working Paper 99-1. Journal of Population Economics, forthcoming.

Kazarosian M (1997) Precautionary saving: a panel data analysis. Review of Income and Statistics 79: 241-247.

MaCurdy TF (1982) The use of time series processes to model the error structure of earnings in a longitudinal data analysis. Journal of Econometrics 18: 83-114.

Miles D (1997), A household level study of the determinants of incomes and consumption. Economic Journal: 107, 1-25

Moffitt R, Gottschalk P (1993) Trends in the covariance structure of earnings in the United States: 1969-1987. Institute for Research on Poverty Discussion Paper, University of Wisconsin, Madison WI.

Skinner J (1988) Risky income, lifecycle consumption, and precautionary saving. Journal of Monetary Economics 22: 237-255

Stevens A H (1994) The dynamics of poverty spells: updating Bane and Ellwood. American Economic Review (Papers and Proceedings) 84, 34-37

Stevens A H (1999) Climbing out of poverty, falling back in: measuring the persistence of poverty over multiple spells. Journal of Human Resources 34: 557-588.

Taylor A (1994) Appendix: sample characteristics, attrition and weighting. In: Buck N, Gershuny J, Rose D, and Scott J (eds) Changing Households: the British Household Panel Survey 1990-1992. ESRC Research Centre on Micro-Social Change, University of Essex, Colchester

Taylor MF (ed) (1998) British Household Panel Survey user manual: introduction, technical reports and appendices. ESRC Research Centre on Micro-Social Change, University of Essex, Colchester

**Table 1**
**Conditioning variables and sample selection criteria used in previous literature: selected examples**

| Paper | Method | Conditioning variables | Sample selection criteria and data set |
|---|---|---|---|
| Bird (1995) | (a) Cross-section | Age, sex, education, ethnic group (lagged income and predicted 'event' variables included in some runs) | All persons in households with non-missing income; PSID and GSOEP, 1983-1986. |
| | (b) Longitudinal | None | Ditto |
| Banks et al. (1999) | Longitudinal | Lagged income, regional and seasonal variables, changes in number employed in household, and changes in the number of adults and children | Pseudo-panel data drawn from pooled Family Expenditure Survey data for 1968-1992. Household heads born 1923-1950 (seven birth cohorts), excluding the self-employed |
| Carroll (1994) | (a) Cross-section | Household head's age, education, occupational group, interaction of age and occupation. | Household heads aged 25-65 with no change in household composition during survey year; US CEX 1960/61 |
| | (b) Longitudinal | Ditto (initial year values). Some runs also include marital status and number of children (initial year values) | Household heads in 1968 aged 25-65, and remaining household head throughout observation period; PSID, 1969-1985 |
| Carroll and Samwick (1995) | Longitudinal | Household head's age, gender, marital status, race education, occupation, industry, time trend, number of children in household. | Household heads in initial year and throughout observation period remaining household head with same marital status, aged 26-50 and in an intact households; PSID, 1981-1987 excluding poverty subsample |

| | | | |
|---|---|---|---|
| Carroll and Samwick (1997) | Longitudinal | Household head's age, education, occupation, industry, time trend, household demographic variables | Household heads in initial year and remaining household head throughout observation period and in intact households; PSID, 1981-1987, income in any year not less than 20% of own average |
| Dardanoni (1991) | Cross-section (cell) | Economic status, occupation and industry of household head. | Household heads not retired or unemployed or in households with other adults working; 1984 Family Expenditure Survey. |
| Dynarski and Gruber (1997) | Longitudinal | Age, education, marital status, family size, number of children, family composition | Male household heads aged 20-59 who are not full-time students; PSID data for 1970-91 |
| Haveman and Wolfe (1985) | (a) Cross-section (cell)<br>(b) Longitudinal | Disability status, age, education, race<br>Ditto, plus time trend | Men aged 51-62 years in 1969; PSID 1969-1981<br>Men aged 51-62 years in 1969; PSID 1969-1981 |
| Jarvis and Jenkins (1998) | Longitudinal | Age, sex, year of interview | British Household Panel Survey 1991-1994 |
| Kazarosian (1997) | Longitudinal | Age, occupation | Men aged 45-59 in 1966, and younger than 65 throughout observation period; NLS Older Men Cohort survey, 1966-1981. |
| Miles (1997) | Cross-section | Household head's sex, marital status, age left school, age squared, number of workers in household, household investment income, region, head's occupation and labour market status, interaction of occupation and age group | Single-family households, excluding those with retired or unemployed household head, or average adult age 55+ years; Family Expenditure Survey for 1968, 1977, 1983, 1986, 1990. |

Notes. ‘Cell’: conditioning variables used to define all possible subgroups, and deviation from cell mean used as estimate (equivalent to regression with full interactions). Subsample selection criteria also included various other conditions: e.g. rejection of cases with missing data, or high- and low-income trimming.

**Table 2**
**Three income risk measures: summary**

| Income risk measure | Assumptions about predictability of time-varying variables | Type of data estimated from | Conditioning variables used in empirical analysis (Section 4) |
|---|---|---|---|
| $\pi^1$ | 'Perfect' predictability | Cross-section | Demographic variables[a] plus labour market variables[b] |
| $\pi^2$ | Not predictable | Cross-section | Individual's age, age-squared, sex, education, region. |
| $\pi^3$ | Not predictable | Longitudinal | Individual's age, age-squared, sex, education, region and unobserved individual fixed effect [c] |

Notes. [a]: demographic variables are individual's region, household size, number of male adults in household, number of female adults in household, number of children in household in various age groups, number of adults in household in various age groups, education of adults. [b]: labour market variables are the work status and occupation of each adult in the household (including individual), summarised in terms of the proportion of the number of adults in each of various categories. [c]: using fixed effects regression.

**Table 3**
**Distribution of income risk for all persons and for selected subgroups: median and interquartile range**

| | Income risk measure | | |
|---|---|---|---|
| | $\pi^1$ | $\pi^2$ | $\pi^3$ |
| All persons | 0.050 | 0.170 | 0.049 |
| | [0.150] | [0.429] | [0.131] |
| Men | 0.051 | 0.143 | 0.048 |
| | [0.151] | [0.386] | [0.129] |
| Women | 0.049 | 0.190 | 0.050 |
| | [0.146] | [0.462] | [0.132] |
| Men, aged 31-50, poorest quintile group | 0.103 | 0.730 | 0.105 |
| | [0.285] | [1.309] | [0.174] |
| Men, aged 31-50, richest quintile group | 0.078 | 0.177 | 0.027 |
| | [0.212] | [0.312] | [0.066] |
| Women , aged 31-50, poorest quintile group | 0.088 | 0.682 | 0.097 |
| | [0.237] | [0.878] | [0.228] |
| Women, aged 31-50, richest quintile group | 0.078 | 0.299 | 0.028 |
| | [0.238] | [0.551] | [0.086] |

Notes. Figures for medians with interquartile ranges in square brackets. See text and Table 2 for definitions of income risk measures and their derivation.

**Table 4**
**How normalised income risk varies with personal characteristics, by income risk measure**

| Individual characteristics | Income risk measure (normalised by mean value) | | |
|---|---|---|---|
| | $\pi^1$ | $\pi^2$ | $\pi^3$ |
| Age 31-50 | 0.1230 | -0.0865 | -0.7979** |
| | (0.1065) | (0.0816) | (0.1115) |
| Age 51-65 | 0.4333** | 0.2456** | -0.7595** |
| | (0.1687) | (0.1096) | (0.1344) |
| Age 66+ | -0.1388 | -0.5568** | -1.4270** |
| | (0.1263) | (0.1060) | (0.1452) |
| Female | -0.1712** | -0.0312 | -0.1962** |
| | (0.0764) | (0.0541) | (0.0752) |
| Household income quintile group 2 | -0.5408** | -1.1366** | -0.1383 |
| | (0.1908) | (0.1361) | (0.1722) |
| Household income quintile group 3 | -0.4699** | -1.2997** | -0.4407** |
| | (0.2156) | (0.1511) | (0.1898) |
| Household income quintile group 4 | -0.5254** | -1.1652** | -0.7418** |
| | (0.2104) | (0.1565) | (0.1828) |
| Richest household income quintile group | 0.0555 | -0.8061** | -0.7775** |
| | (0.2467) | (0.1724) | (0.1891) |
| Constant | 1.3163** | 2.0224** | 2.2596** |
| | (0.2127) | (0.1631) | (0.2370) |

Notes. Table shows regression coefficients (with robust standard errors in parentheses) from a regression of normalised income risk on personal characteristics. (The assumption of independence of errors is relaxed for members of the same household.) The value of each individual's measure of risk has been normalised by dividing it by the mean value for the relevant measure. Reference categories: man aged 16-30, poorest household income quintile group. **: $p < 0.05$. *: $p < 0.10$

**Table 5**
**Average income risk for all persons and for selected subgroups**

| | Total income risk $\pi^3$ | Demographic risk $\Delta^d$ | Demographic risk ratio[a] $\Delta^d/\pi^3$ | Labour market risk $\Delta^m$ | Labour market risk ratio[b] $\Delta^m/\pi^3$ |
|---|---|---|---|---|---|
| All persons | 0.049 | 0.003 | 0.069 | 0.007 | 0.118 |
| | [0.131] | [0.034] | [0.585] | [0.033] | [0.672] |
| Men | 0.048 | 0.004 | 0.089 | 0.008 | 0.133 |
| | [0.129] | [0.035] | [0.643] | [0.033] | [0.697] |
| Women | 0.050 | 0.002 | 0.054 | 0.005 | 0.104 |
| | [0.132] | [0.030] | [0.537] | [0.033] | [0.649] |
| Men aged 31-50, poorest quintile group | 0.105 | 0.010 | 0.162 | 0.138 | 0.708 |
| | [0.174] | [0.038] | [0.547] | [0.216] | [1.811] |
| Men aged 31-50, richest quintile group | 0.027 | 0.004 | 0.184 | 0.008 | 0.177 |
| | [0.066] | [0.027] | [1.251] | [0.015] | [0.827] |
| Women aged 31-50, poorest quintile group | 0.097 | 0.011 | 0.154 | 0.132 | 0.447 |
| | [0.228] | [0.064] | [0.530] | [0.197] | [1.958] |
| Women, aged 31-50, richest quintile group | 0.028 | 0.003 | 0.122 | 0.008 | 0.180 |
| | [0.086] | [0.036] | [0.884] | [0.018] | [0.897] |

Notes. Figures for medians with interquartile ranges in square brackets. See text and Table 2 for further details. [a,b]: Calculated for each individual, and then averaged within each subgroup.

**Table 6**
**Regression summarising heterogeneity in normalised demographic and labour market risk**

| Individual characteristics | Income risk measure (normalised by mean value) | | | | |
|---|---|---|---|---|---|
| | Total income risk $\pi^3$ | Demographic risk $\Delta^d$ | Demographic risk ratio $\Delta^d/\pi^3$ | Labour market risk $\Delta^m$ | Labour market risk ratio $\Delta^m/\pi^3$ |
| Age 31-50 | -0.7979** | -1.5024** | 0.2055 | -0.4498** | 0.3440** |
| | (0.1115) | (0.1105) | (0.1941) | (0.0584) | (0.1034) |
| Age 51-65 | -0.7595** | -1.8915** | -0.5510** | -0.0028 | 0.7418** |
| | (0.1344) | (0.1190) | (0.2683) | (0.0960) | (0.1693) |
| Age 66+ | -1.4270** | -2.0770** | -0.6311** | -1.4034** | 0.4629 |
| | (0.1452) | (0.1231) | (0.1320) | (0.0835) | (0.4174) |
| Female | -0.1962** | 0.0439 | -0.0304 | 0.0210 | -0.1834** |
| | (0.0752) | (0.0585) | (0.1334) | (0.0326) | (0.0882) |
| Household income quintile group 2 | -0.1383 | 0.5267** | 0.2965** | 0.1433 | -0.0508 |
| | (0.1722) | (0.1044) | (0.0706) | (0.0988) | (0.2066) |
| Household income quintile group 3 | -0.4407** | 0.3454** | 0.9947** | -0.5213** | 0.1199 |
| | (0.1898) | (0.1094) | (0.2424) | (0.0996) | (0.2683) |
| Household income quintile group 4 | -0.7418** | 0.2457** | 0.8833** | -0.9165** | 0.0928 |
| | (0.1828) | (0.1024) | (0.2013) | (0.0949) | (0.3824) |
| Household income quintile group 5 | -0.7775** | 0.0610 | 1.0505** | -1.1357** | -0.0151 |
| | (0.1891) | (0.1039) | (0.2623) | (0.0947) | (0.2518) |
| Constant | 2.2596** | 2.0701** | 0.5834** | 1.9377** | 0.7148** |
| | (0.2370) | (0.1301) | (0.1636) | (0.0968) | (0.2628) |

Notes. Table shows regression coefficients (with robust standard errors in parentheses) from a regression of risk measures (see text for definitions) on personal characteristics. (The assumption of independendence of errors is relaxed for members of the same household.) The value of each individual's measure of risk has been normalised by dividing it by the mean value for the relevant measure. Reference categories: man aged 16-30, poorest household income quintile group. **: $p < 0.05$. *: $p < 0.10$

**Table 7**
**Is demographic risk greater or less than labour market risk?**

| Characteristics | Number in subgroup as percentage of all individuals | Percentage of individuals with | | |
|---|---|---|---|---|
| | | Demographic risk > labour market risk | labour market risk > demographic risk | Demographic risk = labour market risk ( = zero) |
| Column percentages | | | | |
| Age | | | | |
| 30 & under | 27 | 37 | 25 | 5 |
| 31-50 | 34 | 43 | 33 | 12 |
| 51-65 | 19 | 12 | 25 | 18 |
| 66+ | 20 | 8 | 17 | 66 |
| Sex | | | | |
| Male | 46 | 49 | 45 | 36 |
| Female | 54 | 51 | 55 | 64 |
| Household average income quintile group[a] | | | | |
| group 1 | 18 | 9 | 17 | 43 |
| group 2 | 21 | 16 | 23 | 27 |
| group 3 | 20 | 22 | 21 | 11 |
| group 4 | 21 | 27 | 19 | 8 |
| group 5 | 21 | 25 | 20 | 11 |
| Row percentages | | | | |
| Overall Percentage | 100 | 38 | 48 | 14 |

Notes. [a] The share of the sample belonging to each of the household income quintile groups is not exactly one fifth because the quintile groups were created prior to the exclusion of adults with less than four income observations (see Section 3).

**Table 8**
**Average income risk, by whether household is intact or not**

| Individual characteristics | Total income risk $\pi^3$ | Demographic risk $\Delta^d$ | Demographic risk ratio $\Delta^d/\pi^3$ | Labour market risk $\Delta^m$ | Labour market risk ratio $\Delta^m/\pi^3$ |
|---|---|---|---|---|---|
| All | 0.049 | 0.003 | 0.069 | 0.007 | 0.118 |
| | [0.131] | [0.034] | [0.585] | [0.033] | [0.672] |
| In intact household | 0.041 | 0.002 | 0.043 | 0.007 | 0.129 |
| throughout observation period | [0.107] | [0.022] | [0.503] | [0.030] | [0.703] |
| In non-intact household | 0.070 | 0.010 | 0.128 | 0.007 | 0.102 |
| | [0.170] | [0.077] | [0.732] | [0.045] | [0.608] |

Notes. Figures for medians with interquartile ranges in square brackets.

**Table 9**
**Coefficient on 'member of non-intact household' in regressions to summarise normalised risk measures, by individual characteristics**

| Dependent variable in regression | Coefficient on 'member of non-intact household' |
|---|---|
| Income risk $\pi^3$ | 0.2935** |
| | (0.0881) |
| Demographic risk $\Delta^d$ | 0.4491** |
| | (0.0887) |
| Demographic risk ratio $\Delta^d/\pi^3$ | 0.0702 |
| | (0.1523) |
| Labour market risk $\Delta^m$ | 0.2807** |
| | (0.0568) |
| Labour market risk ratio $\Delta^m/\pi^3$ | -0.1853 |
| | (0.1248) |

Notes. Table shows the regression coefficient (with robust standard error in parentheses) on the variable indicating membership of a non-intact household. (Assumption of independence of errors is relaxed for members of the same household.) This is from a regression of the dependent variable on personal characteristics: age group, gender, household income quintile group as well as household type. For all risk measures the value of each individual's measure of risk has been normalised by dividing it by the mean value for the relevant measure. Reference categories: man aged 16-30, poorest household income quintile group and member of intact household. **: $p < 0.05$. *: $p < 0.10$