Subjective Income Expectations and Income Risk

Xavier Ramos
Christian Schluter

January 2006

Subjective Income Expectations
and Income Risk

Xavier Ramos
Universitat Autònoma de Barcelona
and IZA Bonn

Christian Schluter
University of Southampton

ABSTRACT

Subjective Income Expectations and Income Risk[*]

In the context of income dynamics, we investigate whether aspects of agents' superior information relative to the econometrician's limited information are captured in subjective expectations data. It is natural, for instance, to assume that the econometrician cannot observe idiosyncratic shocks to both permanent and transitory components of income. In this case perceptions of risk differ between agents and the econometrician. Our tests are based on panel data elicited yearly from British households. We find evidence of superior information consistent with standard income modelling.

Corresponding author:

Xavi Ramos
Departament d'Economia Aplicada
Universitat Autònoma de Barcelona
Edifici B
08193 Bellaterra
Spain
Email: xavi.ramos@uab.es

---

1 Introduction

Survey data are often incomplete, failing to record some information relevant to the behaviour of the agent. The econometrician thus faces an informational disadvantage relative to the agents he observes. It is commonplace to assume that rational agents make forecasts based on a theoretical model which describes the true data generating process, using all relevant information. Agents condition on the full information set when taking expectations, but the disadvantaged econometrician can but condition on a limited information set. However, the survey might include subjective expectation reports of the agents. We examine whether aspects of agents' superior information relative to the econometrician's limited information are captured in subjective expectations data.

This issue is investigated in the context of income processes, where variants of error component models have become the standard modelling device. Below we refer to this class as the Canonical Models (CM) of income. For instance, it is natural to assume that the econometrician cannot observe idiosyncratic shocks to both permanent and transitory components of income. The Canonical Models focus on the variances and covariances of income changes, and thus on risk. The issue is to disentangle information from risk. As a motivating example consider the literature on precautionary savings. Here, the variance of income changes conditional on the full information set determines consumption and savings by the agent. If the econometrician does not observe all relevant information, his perception of uncertainty, as measured by the limited information variance, will exceed that of the agent. The same reasoning applies to the recent literature on the measurement of risk, where measures are often computed on cross-sectional data. In another setting, the econometrician could have an informational advantage, leading to a downward bias of the risk measure, if information collection about e.g. sector-specific unemployment shocks is too costly for the agent. We make these examples more precise below.

Our empirical implementation is based on the British Household Panel Survey (BHPS) which in each wave elicits subjective expectation reports from respondents. The expectation data are categorical, which is a common form found in many surveys (e.g. indices of consumer expectation, sentiments, or confidence). In particular, our expectations data about income changes is in the same format as the Dutch data analysed by Das and van Soest (1997, 1999, 2001), and Das, Dominitz and van Soest (1999). These authors, however, focus on the modelling of the expectations data and on tests of the rational expectations hypothesis. We follow Das, Dominitz and van Soest (1999) and Manski (1990), and assume the categorical reports satisfy a 'best-case' mean-report hypothesis. Not all expectation data in the literature are categorical. Dominitz (1998, 2001) and Dominitz and Manski (1997) use probabilistic expectations collected in Survey of Economic Expectations (SEE), while Guiso et al. (1992, 1996, 2002) and Pistaferri (2001) use the Bank of Italy's Survey of Household Income and Wealth (SHIW). Dominitz (2001) considers the problem of estimating subjective expectations conditional on observed attributes. Guiso et al. (1992, 1996) use the expectation data to study portfolio allocation and subjective uncertainty, and test for precautionary savings. Guiso et al. (2002) estimate empirical profiles of

income uncertainty over the working career taking into account that the distribution of future income depends on institutional features of the labour market. Pistaferri (2001) tests the permanent income hypothesis with quadratic preferences, i.e. the model which rules out precautionary savings. The coarseness of our categorical data might be considered a disadvantage compared to the probabilistic data, and this data limitation should be borne in mind. On the other hand, questions seeking to elicit probabilistic data are more difficult to understand by respondents especially in telephone interviews, and the panel nature of both SHIW and SEE is limited. For instance only 126 of 622 respondents in the SEE have responded in the spring 1994 follow up to the spring 1993 interview (Dominitz, 1998), thus limiting the comparison between expectations and actual realisations. We have at our disposal a large and long panel with a rich set of covariates. We find evidence of superior information of agents consistent with the Canonical Model of income.

The plan of the paper is as follows. Section 2 exposits the Canonical Model of income, defines the information sets, and proposes the tests of the influence of the expectation reports. The data are described in Section 3. The empirical results of our tests are reported in Section 4. Section 5 discusses the findings and concludes. Appendix A describes the specification tests used to arrive at the MA(1) formulation for the transitory income component, Appendix B summarises the data.

2   The "Canonical Income Model"

It is common to formulate the data generating process (DGP) for the log of income of person i at time t, y_{it}, as a decomposition into permanent and transitory components. The permanent component typically takes a Mincerian form, and the law of motion for its unobservable part is given by a random walk. Unobserved ability or heterogeneity gives rise to a fixed effect. The specification of the transitory component is given as an ARMA(p,q) process, but for expositional simplicity we adopt a MA(1) law of motion. Indeed, standard specification tests, detailed in Appendix A, lead us to adopt the MA(1) specification for our data. Given the prevalence of this type of income model in the literature, we refer to it as the "Canonical Model" (CM) of income.[1] A hallmark of this model is the properties that current innovations are "surprises", and that past innovations have lasting effects.

To be precise, log income y_{it} is modelled as a linear function of time-varying and time invariant covariates Z_{it}, and a dynamic error component structure u_{it}:

$$
\begin{aligned}
y_{it} &= \beta' Z_{it} + u_{it} & (1)\\
u_{it} &= \alpha_i + p_{it} + \varepsilon_{it}\\
p_{it} &= p_{it-1} + \eta_{it} \text{ with } \eta_{it} \sim iid\left(0, \sigma_\eta^2\right)\\
\varepsilon_{it} &= \nu_{it} - \delta\nu_{it-1} \text{ with } \nu_{it} \sim iid\left(0, \sigma_\nu^2\right),\ \varepsilon_{i0} \sim iid\left(0, \sigma_{\varepsilon_0}^2\right)
\end{aligned}
$$

---

[1]Variants of the model are, with reference to U.S. data, estimated in Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989), Gottschalk and Moffitt (1995), Baker (1997), and Carroll and Samwick (1997). A recent application to British data is Ramos (2003).

In this standard specification of the error-component structure α_i represents a fixed effect, p_{it} a permanent income component which follows a random walk with stationary innovations, and ε_{it} the transitory income component which evolves according to a MA(1) process.[2] $\sigma_{\varepsilon_0}^2$ denotes the variance of the initial condition. We assume $\beta \neq 0$. Taking first differences eliminates the unobservable fixed effect and p_{it-1}, yielding the CM for income changes

$$
\begin{aligned}
\Delta y_{it} &= \beta' \Delta Z_{it} + r_{1it} \\
r_{1it} &= V_{it-1} + \nu_{it} + \eta_{it}
\end{aligned}
\tag{2}
$$

where Δ denotes the first difference operator, and r_{1it} denotes the first-stage residuals with $V_{it-1} = \delta v_{it-2} - (1 + \delta) v_{it-1}$.

The error-component parameters

$$
\delta, \sigma_\eta^2, \sigma_\nu^2
$$

are commonly of primary interest since they capture the uncertainty or riskiness of the income process. We proceed to make this more precise.

2.1 Information Sets and Risk Measurement

Denote the information set of agent i at time t by Ω_{it}^A. We assume that the agent knows all relevant information, $\Omega_{it}^A = \Omega_{it}^F$, and that the agent forms full-information expectations

$$
E_{it}^F \equiv E\left\{\Delta y_{it} | \Omega_{it}^F\right\}.
$$

The information set of the econometrician is denoted by Ω_{it}^E. We consider a "best-case" scenario for expositional simplicity by assuming that current and past observables ΔZ_i are contained in this set, but the econometrician cannot observe past individual innovations $v_{it-\tau} \notin \Omega_{it}^E$ (which are known to the agent). More generally, we refer to DGP-relevant variables which are included in the information set of the agent but are excluded from the information set of the econometrician as 'private information'. Therefore, the econometrician can but form limited-information expectations

$$
E_{it}^L \equiv E\left\{\Delta y_{it} | \Omega_{it}^E\right\}.
$$

Now consider the conditional variance of the income process: The leading example is the literature on precautionary savings, in which the conditional variance determines consumption and savings. The agent computes the full information variance

$$
Var\left\{\Delta y_{it} | \Omega_{it}^F\right\} = \sigma_\eta^2 + \sigma_\nu^2.
$$

The only source of randomness stems from the contemporaneous innovations to both permanent and transitory components of income. The effect of past innovations in

---

[2]For expositional simplicity we consider only the case where $Var\{\nu_{it}\}$ is time-invariant. Meghir and Pistaferri (2003) propose to impose an ARCH specification on the conditional variance.

transitory income are known to the agent. However, the econometrician will perceive risk differently since $V_{it-1} \notin \Omega_{it}^E$:

$$Var\left\{\Delta y_{it}|\Omega_{it}^E\right\} = 2\sigma_\nu^2\left(1 + \delta + \delta^2\right) + \sigma_\eta^2.$$

He fails to disentangle information known to agent, $Var\{V_{it}\}$, from his perception of risk, i.e. the variance of the first stage residuals. The measure of risk thus computed exceeds the agents' perception, $Var\left\{\Delta y_{it}|\Omega_{it}^E\right\} \geq Var\left\{\Delta y_{it}|\Omega_{it}^F\right\}$. Since outcomes other than consumption and savings might be of interest to economists, a separate literature on the measurement of income risk has recently emerged. Popular cross-sectional measure of income risk is the coefficient of variation (as in Guiso et al., 2002) or the variance of the income growth residuals (Burgess et al., 2000 and references therein), given in CM by $Var\{V_{it-1} + \nu_{it} + \eta_{it}\} = Var\left\{\Delta y_{it}|\Omega_{it}^E\right\}$.

The correct measure of risk as perceived by the agent, $Var\left\{\Delta y_{it}|\Omega_{it}^F\right\}$, can be recovered by the econometrician under the maintained assumptions from realisations data alone provided he can identify the four parameters δ, σ_η^2, σ_ν^2, and $\sigma_{\varepsilon_0}^2$. This can be achieved by minimising the distance between the empirical and theoretical covariance matrix of (first-stage) residuals $E\{r_{1,it}r_{1,it-s}\}$, as detailed in Appendix A, if he has at least five waves of panel data. If the information set of the econometrician excludes an additional set of DGP-relevant mean-zero covariates with cross-sectional variance σ_X^2, (such as an indicator for a planned reduction in hours worked) identification requires one additional wave. With fewer waves, the full information variance is not identified, and the limited information variance provides only an upper bound. In this case subjective expectations data may be used to (partially) filter out private information of agents. Our procedure, explained below, complements the results of Dominitz (2001), who recovers structural parameters from two waves of panel data by comparing estimates based on probabilistic expectations data and realisations data. Such probabilistic expectations data and one-period ahead realisations data make also possible computing the measure of subjective risk directly as the variance of the subjective prediction errors. However, such datasets are rather rare and not without their own interpretative problems.

2.2   Information Sets and Expectation Reports

The agent forms full-information expectations, the econometrician forms limited information expectations as past innovations are not known to him. However, the agent makes a categorical report $E_{it}^R = k$ based on his expectations, which is recorded in the survey and thus observable by the econometrician. The categories are ordered {worse off, no change, better off}. The number of categories is $K = 3$. Any cardenalisation is arbitrary, and for notational simplicity we set $k = 1$ to represent the first, $k = 2$ the second, and $k = 3$ the third category. The problem of interpreting categorical expectation data has a long history. We follow Manski (1990) and Das, Dominitz and van Soest (1999) and assume that

$$\text{agent } i \text{ reports category } k = E_{it}^R \iff E_{it}^F \in I_k$$

(labelled the "mean-report assumption" and the leading interpretation), where $I_k = (m_{k-1}, m_k]$. The interval boundaries are not observable, assumed to be person-invariant, and the intervals do not overlap. Since the expectation categories are ordered, it is natural to assume that $m_0 < m_1 < m_2 < m_3$. The authors also discuss other potential interpretations of categorical expectations data. The next section briefly considers alternative but naive methods of forming expectations.

Given the assumed DGP, we have

$$\begin{aligned} \Delta y_{it} &= E_{it}^F + noise_{it} \\ &= E_{it}^L + r_{1it} \end{aligned}$$

with $r_{1it} = V_{it-1} + noise_{it}$, and $noise_{it} = \nu_{it} + \eta_{it}$.

2.3 The Influence of Expectation Reports

Do the expectation reports E_{it}^R contain information not contained in the observables which can be used by the econometrician ? We consider two tests based on the average residuals conditional on the expectation report $E\left\{r_{1it}|E_{it}^R = k\right\}$.

The first test is a standard difference of means test for the different categories. If the expectation report does not contain extra information, then the conditional residuals should have the same mean. This test does not require a (arbitrary) cardenalisation of E_{it}^R. A related test consists in regressing the first-stage residuals on the expectation reports

$$r_{1it} = \gamma_0 + \gamma E_{it}^R + r_{2it}. \tag{3}$$

This approach does require a cardenalisation. It is important to note that the two tests do not require agents to hold rational expectations (indeed, the literature cited in the introduction focusses on tests for rational expectations). What we test is whether the expectation reports contain private information of the agents (which is unavailable to the econometrician). Rational expectations is one example of expectation formation in which this should be observed. However, the tests have discriminatory power against naive expectation formation models. In particular, if the expectation reports do not contain extra information, then the expected residuals should be the same for all groups (and equal to zero), and $\gamma = 0$. For instance, if the reports were random then $E\left\{r_1|E_{it}^R = k\right\} = E\left\{V|E_{it}^R = k\right\} = E\left\{V\right\} = 0$. If agents were myopic, then next period's forecasted income would be this period's income, and the expectation report would be "no change".

If the tests diagnose private information of agents, regression (3) can be used to partially correct measures of subjective risk if the error component parameters cannot be identified from realisation data (i.e. when the panel is not long enough). This is achieved by filtering out the function E_{it}^R of the private information from the first stage residuals r_1, and computing the variance of the second stage residuals (or filtering out observables ΔZ from the expectation reports and regressing the residuals from this regression against the first stage residuals).

2.3.1 Implications for the Coefficient γ

What is the expected sign of the coefficient γ under the maintained assumptions, i.e. (i) the true DGP is given by CM, which is known to the agents and the econometrician, (ii) they form rational expectations, and (iii) expectation reports satisfy the mean-report hypothesis ? We now show that in this case the expected residuals should be monotonically increasing and $\gamma > 0$.

Taking expectations of (3) yields

$$E\left\{r_{1it}|E_{it}^R = k\right\} = \gamma_0 + \gamma k + E\left\{r_{2it}|E_{it}^R = k\right\}.$$

But $E\left\{r_{1it}|E_{it}^R = k\right\} = E\left\{V_{it-1} + noise_{it}|E_{it}^R = k\right\} = E\left\{V_{it-1}|E_{it}^R = k\right\}$. Moreover, $k = E_{it}^R \iff E_{it}^F \in I_k$ so

$$\begin{aligned} E\left\{r_{1it}|E_{it}^R = k\right\} &= E\left\{V_{it-1}|\beta'\Delta Z_{it} + V_{it-1} \in I_k\right\} \\ &= E\left\{V_{it-1}|V_{it-1} \in \widetilde{I}_{ki}\right\} \end{aligned}$$

where $\widetilde{I}_{ki} = (m_{k-1} - \beta'\Delta Z_{it}, m_k - \beta'\Delta Z_{it}]$ which is now person-specific. Assume for the time being that $\widetilde{I}_{ki}$ is person-invariant, i.e. $\widetilde{I}_{ki} = \widetilde{I}_k$. Note that $E\left\{V_{it-1}\right\} = 0$, therefore $E\left\{V_{it-1}|V_{it-1} \in \widetilde{I}_k\right\} \neq 0$ unless the conditional expectation is zero on all three intervals. Since the categories are ordered with $m_0 < m_1 < m_2 < m_3$, and since $E\left\{V_{it-1}|V_{it-1} \in \widetilde{I}_k\right\} = \int 1\left(v \in \widetilde{I}_k\right) v f_{v|v \in \widetilde{I}_k}(v)\,dv$, it follows that for $V > 0$ we have $E\left\{V_{it-1}|V_{it-1} \in \widetilde{I}_{k=3}\right\} > 0$ and $E\left\{V_{it-1}|V_{it-1} \in \widetilde{I}_{k=1}\right\} < 0$. Therefore, we have a positive correlation $cor\left(r_1, E_{it}^R\right) > 0$ which implies a positive regression coefficient $\gamma > 0$. Now consider the case when $\widetilde{I}_{ki}$ varies across persons. The above argument about the positive correlation holds for each agent. Summing positive correlations across all agents still gives a positive correlation, so again $\gamma > 0$.

2.3.2 Inference and Robustness

We briefly raise two issues regarding the implementation of regression (3).

First, inference about γ requires taking into account that an estimable version of equation (3) includes generated regressors. More precisely, $r_{1it} = V_{it-1} + noise_{it}$ is unobservable and estimated by the fitted residuals $\widehat{r}_1$ from the first stage regression (2). In particular $r_{1it} = \widehat{r}_{1it} + (r_{1it} - \widehat{r}_{1it}) = \gamma_0 + \gamma E_{it}^R + r_{2it}$, so the estimable version of equation (3) is

$$\widehat{r}_{1it} = \gamma_0 + \gamma E_{it}^R + \widetilde{r}_{2it} \tag{4}$$

with $\widetilde{r}_{2it} = r_{2it} - (r_{1it} - \widehat{r}_{1it})$. The mean of the squared second stage residuals estimates the variance of $\widetilde{r}_2$ and not of r_2. This requires a correction, and we have used a bootstrap procedure.

Second, it might be desirable to use an estimation procedure for (3) which is robust to departures from the null, in particular departure from the mean-report assumption. It might be conceivable, for instance, that under some alternative E_{it}^R and r_{2it} are correlated via V_{it-1}. Provided that E_{it-2}^R is sufficiently correlated with E_{it}^R then this could serve as a suitable instrument since V_{it-3} is uncorrelated with V_{it-1}.

3   The Data

Our empirical analysis is based on the first ten waves of the British Household Panel Survey (BHPS), a popular longitudinal dataset built on similar principles as the Panel Study of Income Dynamics (PSID) in the U.S. or the German GSOEP.[3] The income concept is dictated by the survey question about subjective income expectations, and refers to total (real) household income (and not personal earnings). Our unit of analysis, however, is the person, since households can form and dissolve over time. We control for demographic characteristics directly rather than using an equivalisation procedure. We focus on households whose heads are "prime-aged" (between 25 and 59 years), and thus exclude "young" and "older" households because their life-cycle and in particular labour markets events can differ fundamentally (the "getting started" phenomenon, and living in retirement). Appendix B provides summary statistics of the sample covering demographics, educational attainment, and occupation. Expectation reports and income data are the subject of this section. Note that expectations are reported at time t while data about actual income changes are obtained from the $t + 1$ wave.

First, we consider the data used to define the expectation reports E_{it}^R. The survey question is: "Looking ahead, how do you think you will be financially a year from now, will you be: better off, worse off than you are now, or about the same?" We have $K = 3$ different categories, which we have ordered "worse off", "about the same", "better off". The income concept is not defined explicitly.

Next, we consider the income data. The survey does not elicit directly income levels of respondents. Rather, it contains two different sources of information about income: categorical reports about income changes on the same line as the expectation reports, and estimates of income generated by the data provider. After describing in more detail each source, we show that we can use the (log of) generated income as an instrument for y_{it}. In each wave of the survey respondents are asked to categorise their perception of actual income changes ("looking back ..."). Again, the income concept is not explicitly defined, but it is highly likely that the respondent will use the same metric as in the expectation question. We refer to this data as "categorical income reports". The data provider also estimates income.[4] It is this variable which we have used to measure y_{it}, and we refer to it as "actual incomes". Since the income concept of the data provider might not correspond to that of the respondent, it is important to compare the two. In Figure 1 we plot the time series of the mean of the "actual income" changes for the three categorical income reports. The time-averaged

---

[3]The BHPS is a longitudinal panel data set consisting of some 5500 households (approximately 10000 individuals) first interviewed in the autumn of 1991 (wave 1) followed and re-interviewed every year subsequently. The initial sample represents a response rate of about 69% (proxies included) of the effective sample size. Wave-on-wave attrition rates for the subsequent waves are low. For a detailed discussion of BHPS methodology and representativeness see Taylor (1995), and http://www.iser.essex.ac.uk/bhps/.

[4]More precisely, the income measure is the log of the pre-tax post-transfer real household income. It includes earnings from employment and self-employment, cash social security and social assistance benefits, and income from savings and investment. See Bardasi, Jenkins and Rigg (1999) for more detailed information on the BHPS income variable.

means of actual income change equal -2.3% (perception: "worse off"), 3.0% ("about same"), and 6.4% ("better off").[5] We conclude that the income concept of the data provider can be used to approximate that of the respondents.
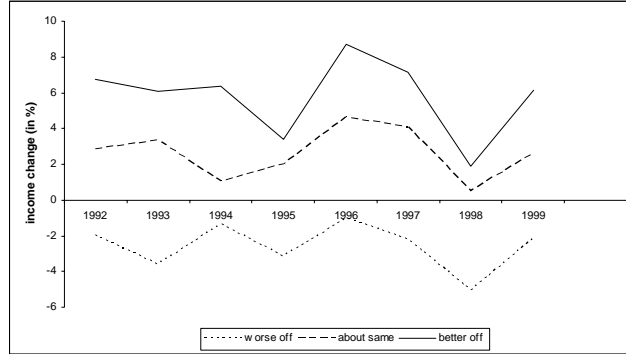

Figure 1: The mean of "actual income change" by categorical income reports.

Next, we address the quality of the expectations data, by comparing expectation and income reports, i.e. predictions and realisations. We provide a simple descriptive examination, which seeks to establish whether the expectations data are potentially informative, rather than being mere random or myopic reports. First, we compare the expectation reports E_{it}^R to the (next period's) categorical income reports. The results are reported in Table 1 . On average, expectations are fulfilled: 44 per cent of agents expecting to be worse off report having become worse off next period, 51 per cent expect and report being the same, and 49 per cent of those expecting to be better off report having become better off. Previous studies partition their sample by educational attainment (Meghir and Pistaferri, 2004, and Hubbard et al.,1995). It is therefore of interest to distinguish realisations by educational status, in our case in terms of school-leaving qualifications. Interestingly, the table reveals that the more highly educated individuals do not hold, on average, more accurate expectations. This also corresponds to results reported in Dominitz (1998) based on probabilistic expectations data for the U.S. We therefore do not partition the sample by educational attainment in the subsequent analysis.

Finally, we relate the expectation reports, via the "categorical income reports", to the "actual income changes" measured by Δy_{it}. The results are reported in Table 2. Reassuringly the ordering is consistent, as agents expecting to be better off experience income growth larger than agents expecting no change, who in turn experience a larger income growth than agents expecting to be worse off. For the sake of completeness, the appendix also contains a descriptive analysis of who holds what expectations based on ordered logit regressions.

_____

[5]Other ways of estimating the relationship between perceived and actual income changes provide very similar results. A random effects regression of actual income change on subjective expectations and time dummies (to capture macro shocks) provides the following estimates for the means of actual income change: -1.6% (worse off), 3% (about same), and 5.2% (better off).

expectation reports	"categorical income reports"			
	"worse off" [%]	"about same" [%]	"better off" [%]	Total [%]
Entire sample				
"worse than now"	44.26	33.99	21.75	100
"about the same"	22.16	50.98	26.86	100
"better than now"	18.03	32.86	49.11	100
University or higher degree				
"worse than now"	37.71	33.14	29.14	100
"about the same"	20.49	44.72	34.79	100
"better than now"	16.83	24.32	58.85	100
A Levels or higher vocational degree				
"worse than now"	47.52	30.20	22.28	100
"about the same"	23.25	47.32	29.44	100
"better than now"	18.59	31.26	50.15	100
O Levels or less				
"worse than now"	43.93	37.96	18.11	100
"about the same"	21.68	55.51	22.81	100
"better than now"	17.96	38.40	43.64	100

Table 1: Subjective expectations and "categorical income reports". Notes: The category 'O levels or less' refers to individuals who have left school aged 16 or younger, and 'A-levels' to school leavers who have graduated with a university entrance qualification (school grade 12). The population shares of the entire sample for the expectation report categories are 11.8 %, 56.9 %, and 31.3 %.

We proceed to discuss the empirical results based on BHPS data.

4 Empirical Results

The CM has been implemented in the customary way as follows. Observables are filtered out using OLS. The choice of regressors Z_{it} is conventional and includes all level covariates listed in column I of Table 8. Detailed results, not reported here for the sake of brevity, are available from the authors on request. Our set of regressors extends the usual ones by including covariates which are typically fixed for long periods if not time invariant, such as occupation. This provides an additional check on the CM, since the coefficients of these covariates should be zero. As it turns out, this is typically the case: in particular, the regressors relating to education and occupation are all insignificant. The estimated β conforms to results in the literature. Filtering out the observables yields the fitted residuals $\widehat{r}_{1it}$ which are used in the estimation of the error component structure. The specification tests detailed in the appendix lead us to adopt the MA(1) specification of the CM given in (1). In the appendix we also report and briefly discuss the estimates of δ, σ_η^2, and σ_ν^2, which turn out to be in line with those reported in the literature.

We turn to the implementation of our tests based on the expectation reports.

expectation reports	"categorical income reports"			
	"worse off"	"about same"	"better off"	Total
"worse than now"	-3.0	2.5	7.9	1.3
"about the same"	-1.4	2.7	5.3	2.5
"better than now"	-3.6	4.1	7.1	4.2
Total	-2.3	3.0	6.4	2.9

Table 2: Mean of "actual income changes" (in per cent) by subjective expectations and "categorical income reports."

4.1 The Empirical Influence of Expectation Reports

Column II of Table 3 reports the average fitted residuals by expectation group. These exhibit the positive monotonicity expected under the null hypothesis. The table also reports the difference of means tests, testing, for the sake of completeness, all means together and by pairs. Of course, only $K-1$ means need to be tested since the fitted residuals summed across all agents sum to zero by construction. The difference of means tests confirm that the mean residuals differ significantly by expectation group, and one sided t-tests confirm statistically the monotonicity.

expectation reports	mean fitted residuals	t-statistic of pairwise difference of means test	
E_{it}^R	$E\left\{\widehat{r}_{1it}\vert E_{it}^R\right\}$	"about the same"	"better off"
"worse off"	-.031	-4.565	-5.591
(p-value)		(0.000)	(0.000)
"about the same"	-.002		-3.843
(p-value)			(0.000)
"better off"	.016		
Anova F-test	21.03		
(p-value)	(0.000)		

Table 3: Mean of the residuals by expectation groups, and difference of means tests.

We turn to the estimation of equation (4) and the test of γ. Given the presence of a generated regressor, we have used a bootstrap procedure with 500 repetitions to generate standard errors. Table 4 reports point estimates and standard errors. As expected from the difference of means tests, the OLS estimates confirm the null hypothesis, exhibiting positive monotonically increasing dummies which are statistically significant. The F-test suggest explanatory power.

The next column reports the discussed robustness check when E_{it}^R is instrumented with E_{it-2}^R.[6] The instruments pass the Sargan test, and the first stage regression passes the F-test, the estimated γ is insignificant, but the IV regression does not pass the F-test. Hence this result should not be interpreted as evidence against the null hypothesis, and is not unexpected given the coarseness of the expectations data.

---

[6]Unreported results for extended instrument sets are qualitatively similar and available upon request.

expectation reports	$\widehat{\gamma}_{OLS}$	$\widehat{\gamma}_{IV}$
"about the same"	0.028	- 0.082
	(0.007)	(0.049)
"better off"	0.047	- 0.049
	(0.007)	(0.044)
constant	-0.031	0.063
	(0.006)	(0.042)
p-value	0.00	0.13

Table 4: Estimation of γ. Notes: SE in parentheses. p-value refers to the F-test.

5   Discussion and Conclusions

In the context of income dynamics, we have investigated whether aspects of agents' superior or 'private' information relative to the econometrician's limited information are captured in the categorical subjective expectations data elicited annually from British households. It is natural, for instance, to assume that the econometrician cannot observe past idiosyncratic shocks to both permanent and transitory components of income. Another example is the planned but, in the survey, unrecorded reduction in the agent's labour supply. Importantly, our tests do not require agents to hold rational expectations, but rational expectations is one expectation formation model in which expectations data should capture relevant private information. We first find evidence for private information, as the means of the first stage residuals vary by expectation group and γ differs from zero. At the same time, this evidence rejects naive methods of forming expectations. Second, the actual monotonicity of residuals and $\gamma > 0$ is consistent with the assumptions maintained in this paper.

If agents have superior information relative to the econometrician, the econometrician's perception of risk will exceed that of the agent. In particular, we find $Var\left\{\Delta y_{it}|\Omega_{it}^{E}\right\} = .12$ and, based on the Canonical Model, $Var\left\{\Delta y_{it}|\Omega_{it}^{F}\right\} = .09$, a difference of 33 per cent. This has important consequences for the literature on precautionary savings since here the variance of income changes conditional on the full information set determines consumption and savings by the agent. Based on the assumed DGP the correct risk measure can be computed from realisation data by the econometrician provided the panel is sufficiently long. If the individual error component parameters cannot be identified then the bias of the econometrician's risk measure can be reduced by filtering out the expectation reports from first stage residuals, or including the expectation reports as an additional regressor.

References

Abowd, J. and D. Card (1989), "On the covariance structure of earnings and hours changes", *Econometrica*, 57, 411-445.

Baker, M. (1997), "Growth-rate heterogeneity and the covariance structure of life-cycle earnings", *Journal of Labour Economics*, 15 (2), 338-75.

Baker, M. and G. Solon (2003), "Earnings dynamics and inequality among Canadian men, 1976-1992: evidence from longitudinal income tax records", *Journal of Labor Economics*, 21, pp. 289-321.

Bardasi, E., S. P. Jenkins and J. A. Rigg (1999), "Documentation for derived current and annual net household income variables, BHPS waves 1-10", Institute for Social and Economic Rerearch, University of Essex.

Burgess, S., K. Gardiner, S. P. Jenkins and C. Propper (2000), "Measuring income risk", CEPR Discussion Paper: 2512; also in ISER Working Paper no. 2000-15, University of Essex: Colchester.

Cappellari, L. (2004), "The dynamics and inequality of Italian men's earnings: long-term changes or transitory fluctuations?", *Journal of Human Resources*, 39(2), 475-499.

Carroll, C. D. and A. A. Samwick (1997), "The nature of precautionary wealth", *Journal of Monetary Economics,* 40, 41-71.

Das, M., Dominitz, J. and A. van Soest (1999) "Comparing predictions and outcomes: Theory and application to income changes", *Journal of the American Statistical Association*, 94, No. 445, 75-85.

Das, M. and A. van Soest (1997) "Expected and realized income changes: Evidence from the Dutch socio-economic panel", *Journal of Economic Behavior and Organization*, 32, 137-154.

Das, M. and A. van Soest (1999) "A panel data model for subjective information on household income growth", *Journal of Economic Behavior and Organization*, 40, 409-426.

Das, M. and A. van Soest (2001) "Expected versus realized income changes: A test of the rational expectations hypothesis", Tilburg University.

Dickens, R. (2000) "The evolution of individual male wages in Great Britain: 1975-95", *Economic Journal*, 110, 27-49.

Dominitz, J. (1998) "Earnings expectations, revisions, and realisations", *The Review of Economics and Statistics*, 80, 374-388.

Dominitz, J. (2001), "Estimation of income expectations models using expectations and realization data", *Journal of Econometrics*, 102, 165-195.

Dominitz, J. and C. F. Manski (1997), "Using Expectations Data To Study Subjective Income Expectations", *Journal of the American Statistical Association*, 92, No. 439, 855-867.

Gottschalk, P. and R. Moffitt (1995), "Trends in the covariance structure of earnings in the U.S.: 1969-1987", Working Paper, Brown University.

Guiso, L., Jappelli, T., and L. Pistaferri (2002), "An empirical analysis of earnings and employment risk", *Journal of Business and Economics Statistics,* 20(2),

241-253.

Guiso, L., Jappelli, T., and D. Terlizzese (1992), "Earnings uncertainty and precautionary saving", *Journal of Monetary Economics,* 30, 307-337.

Guiso, L., Jappelli, T., and D. Terlizzese (1996), "Income Risk, Borrowing Constraints, and Portfolio Choice", *American Economic Review,* 86, 158-172.

Hubbard, G., Skinner, J., and S. Zeldes (1995), "Precautionary saving and social insurance", *Journal of Political Economy,* 103, 360-399.

MaCurdy, T. (1982), "The use of time series processes to model the error structure of earnings in longitudinal data analysis", *Journal of Econometrics*, 18, 83-114.

Manski, C. (1990), "The use of intentions data to predict behaviour: a best-case analysis", *Journal of the American Statistical Association*, 85, 934-940.

Meghir, C. and L. Pistaferri (2004), "Income variance dynamics and heterogeneity", *Econometrica*, 72(1), 1-32.

Pistaferri, L. (2001), "Superior Information, Income Shocks, and the Permanent Income Hypothesis", *The Review of Economics and Statistics*, 83, 465-476.

Ramos, X. (2003) "The covariance structure of earnings in Great Britain, 1991-1999", *Economica*, 70, 353-374.

Taylor, M.F. (ed.) (1995), *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*, Colchester.

A Error Component Specifications

This appendix briefly details how the MA(1) specification of the CM has been arrived at, and reports the estimates of the error component parameters.

We started out fitting a low order ARMA specification for the transitory component ε_{it}. However, the estimates of this specification suffered from some common identification problems.[7] Thus, we assumed that ε_{it} follows an $MA(q)$ process, with q to be determined empirically by testing the suitability of the theoretical structure imposed on the estimated autocovariances $E\{\widehat{r_{1it}r_{1it-s}}\}$. Specifically, the MA(1) process given by (1) imposes zero-restrictions on third and higher order covariances. Table 5 reports the test results[8] of zero-restrictions for the null hypothesis that $E\{r_{1it}r_{1it-s}\} = 0$ (for $1 \leq s \leq 5$). Unexplained income growth rates seem to be correlated up to the second order: the test statistic does not reject the null of zero autocovariances at third or higher order (p-value $= 0.122$), but rejects it at second or higher order (p-value < 0.001). Therefore we use a $MA(1)$ process to model the transitory component.

	Order of covariances				
	1	2	3	4	5
degrees of freedom	36	28	21	15	10
χ^2	401.3	111.0	28.7	9.9	7.2
p-value	<0.001	<0.001	0.12	0.82	0.71

Table 5: Tests of zero-restrictions for MA(q) order.

Next, we test for the absence of a permanent shock. Following Meghir and Pistaferri (2004) we do so by testing whether the variance of the permanent shock $E\{\eta_{it}^2\}$ is zero. Given that the transitory component follows a $MA(1)$ process, the covariance structure of r_{1it} takes, under the null hypothesis of no permanent shock, the form

$$E\{r_{1it}r_{1it}\} = \begin{cases} E\{\nu_{it}^2\} + (1+\delta)^2 E\{\nu_{it-1}^2\} + \delta^2 E\{\nu_{it-2}^2\} & \text{for } s = 0 \\ (1+\delta) E\{\nu_{it-1}^2\} - \delta(1+\delta) E\{\nu_{it-2}^2\} & \text{for } s = 1 \\ \delta E\{\nu_{it-2}^2\} & \text{for } s = 2 \\ 0 & \text{for } s > 2. \end{cases} \quad (5)$$

Hence

$$E\{r_{1it}(r_{1it-2} + r_{1it-1} + r_{1it} + r_{1it+1} + r_{1it+2})\}$$

equals zero under the null. Under the alternative hypothesis, however, it equals the variance of the permanent shock $E\{\eta_{it}^2\}$. Since we do not observe the true shocks

---

[7]In particular, the estimated variance of the random walk σ_η^2 is negative. Such identification problems are not unusual in the literature, and have been encountered, inter alia, by Ramos (2003), Cappellari (2004), Dickens (2000), Baker and Solon (1999) and Baker (1997).

[8]See e.g. Abowd and Card (1989). The test statistic has the form $m_r' V_{rr}^{-1} m_r$, where m_r is the subvector comprising the elements of the covariance matrix restricted to zero and V_{rr} is the covariance matrix associated to the elements of m_r. The test statistic is distributed as a χ^2 with degrees of freedom equal to the dimension of m_r.

σ_η^2	0.0519
	(10.4)
$\sigma_{\varepsilon_0}^2$	0.0724
	(12.2)
σ_ν^2	0.0421
	(12.3)
δ	-0.2645
	(-8.7)
SSR	0.0051
$\chi^2 - test$	72.2

Table 6: Estimates of the error component structure. Notes: t-ratios in parenthesis.

to income, we use the predicted composite residuals to implement this test, which is a standard one-sided test. The null hypothesis of no permanent shock is strongly rejected by the data: The t-statistic is 3.45, which implies a p-value smaller than 0.01.

Table 6 reports the estimates of the error component parameters of the adopted Canonical Model, and given in (1). We briefly turn to the interpretation of the error component estimates. The estimates of Table 6 imply that the permanent component accounts for 43 per cent and the transitory component for 57 per cent of the variance of the income growth residuals. The $MA(1)$ process implies that there is no persistent serial correlation in the (unexplained) income growth rates. The negative estimate of the MA parameter δ accommodates the (unreported) sharp decline in the covariances at the first and second orders. Our results for the decomposition into permanent and transitory components are similar to those obtained in other similar studies, as are the estimates of the other error component parameters. It is not possible to compare directly our results with previous studies as these differ in terms of income concepts, and country focus. This notwithstanding, our parameter estimates are very plausible when compared to PSID-based studies (see, *inter alia*, Meghir and Pistaferri (2004), Baker (1997), Carroll and Samwick (1997), MaCurdy (1982)). The study most comparable to our set up is Carroll and Samwick (1997) for the U.S. They find a more unequal relative contribution of the permanent and transitory components to total income growth variance (33 and 77 per cent, respectively).

B   Data Appendix

Table 7 provides summary statistics of the sample. The educational dummies refer to school-leaving qualifications. The category 'O levels or less' refers to individuals who have left school aged 16 or younger, and 'A-levels' to school leavers who have graduated with a university entrance qualification (school grade 12). The variable "expectations error" is defined as the deviation between expectations and perceived income change. Hence, "underpredict a lot" refers to individuals holding "worse than now" expectations and reporting being "better off" in the categorical income change question, whereas "underpredict a little" refers to individuals holding "worse

than now" ("about the same") expectations and reporting being "about the same" ("better off") in the categorical income change question.

	Mean	SD
Age	40.66	9.20
Sex: Female	0.47	0.50
Education (increasing in levels)		
No Qualifications	0.14	0.35
Other Qualifications	0.09	0.28
O levels	0.21	0.41
A levels	0.12	0.32
Other Higher Qualifications	0.29	0.46
Health limits work: Yes	0.07	0.26
Marital Status: Married	0.80	0.40
Children		
1 child	0.18	0.39
More than 1 child	0.26	0.44
Decreased number of children	0.04	0.21
Increased number of children	0.03	0.18
Job Status		
Self-employed	0.12	0.33
Became self-employed	0.02	0.13
Became employee	0.02	0.12
Full-time or Part-time		
Part-time	0.19	0.39
Became Part-time	0.03	0.17
Became Full-Time	0.03	0.18
Contract		
Non-standard	0.05	0.22
Became non-standard	0.02	0.15
Became Permanent	0.03	0.17
Region		
North	0.58	0.49
Does change region	0.02	0.12

	Mean	SD
Stan.Occupation Class.		
Professional	0.12	0.33
Associate profess. technical	0.12	0.33
Clerical	0.17	0.38
Craft related	0.12	0.33
Protective services	0.09	0.29
Sales	0.06	0.23
Plant machine operatives	0.09	0.28
Other	0.06	0.24
Got a "worse" occupation	0.10	0.30
Got a "better" occupation	0.11	0.32
Time Dummies		
Wave 2	0.12	0.33
Wave 3	0.11	0.31
Wave 4	0.11	0.31
Wave 5	0.11	0.31
Wave 6	0.11	0.31
Wave 7	0.11	0.32
Wave 8	0.11	0.31
Wave 9	0.11	0.31
Log (income level)	9.75	0.50
Δ in income$_{t-1,t}$		
"worse off"	0.23	0.42
"better off"	0.33	0.47
Expectations Error$_{t-1,t}$		
"underpredict a lot"	0.02	0.14
"underpredict a little"	0.18	0.39
"overpredict a little"	0.23	0.42
"overpredict a lot"	0.06	0.23

Table 7: Summary statistics for the pooled sample. N=24,531

For the sake of completeness, Table 8 describes who holds what expectations. Reported are the estimates of ordered logits of the three expectations categories. The baseline specification, labelled I, includes a set of standard covariates measuring demographics and labour market characteristics (including dummies for: self-employment, part and full time worker, whether the job is permanent or temporary ("contract"), and the standard occupational classification). We have included occupational status as indicators of the potential complexity of the income generating process, and these turn out to be significant. Many level variables, in particular income levels and time-invariant covariates such as education, are not significant, whereas many "event" variables, referring to events between $t-1$ and t, are significant.

We also report an augmented specification, labelled II, which adds to I indicators for the perceived income change, and the past expectations error. The significance

of all these variables suggests a dynamic structure of how expectations are formed. Perceived income changes in the past twelve months seem to have an impact on expectation formation. People who suffered a subjective income fall (gain) are more (less) pessimistic about their future income than people who have experienced no income change. This phenomenon has also been observed for the Netherlands (Das and van Soest, 1999).

Individuals do not seem to revise their expectations in the light of past prediction errors. Individuals who underestimated their incomes in the past are more pessimistic, and people who overestimated them are more optimistic than those who saw their expectations fulfilled; i.e. the predictions of the under-predictors do not, on average, improve. Moreover, people who "underpredicted a lot" are more pessimistic than those who "underpredicted a little", while individuals who "overpredicted a lot" are less pessimistic than those who "overpredicted a little".

	I		II	
	Coef.	P-value	Coef.	P-value
Age	-0.087	0.000	-0.055	0.000
Age squared	0.001	0.014	0.0004	0.049
Sex: Female	-0.201	0.000	-0.170	0.000
Education	included	not significant	included	not significant
Health limits work: Yes	-0.145	0.025	-0.119	0.036
Marital Status: Married	0.035	0.495	-0.014	0.744
Children: 1 child	-0.123	0.012	-0.068	0.106
more than 1 child	0.020	0.703	0.043	0.306
decreased	0.023	0.725	0.065	0.331
increased	0.004	0.958	0.387	0.000
Job Status: self-empl.	0.326	0.000	0.236	0.000
Became self-employed	0.215	0.057	0.261	0.025
Became employee	0.378	0.001	0.273	0.024
Full-time or PT: Part-time	0.025	0.650	0.049	0.275
Became Part-time	0.062	0.449	0.117	0.175
Became Full-Time	0.116	0.120	0.002	0.984

	I		II	
	Coef.	P-value	Coef.	P-value
Contracts: Non-standard	0.176	0.055	0.137	0.095
Became non-standard	-0.109	0.366	-0.088	0.482
Became Permanent	0.241	0.001	0.129	0.105
Occupation Class.				
Professional	-0.147	0.044	-0.114	0.055
Associate profess. technical	-0.112	0.110	-0.077	0.185
Clerical	-0.117	0.081	-0.066	0.233
Craft related	-0.289	0.000	-0.189	0.001
Protective services	-0.168	0.027	-0.099	0.109
Sales	-0.041	0.605	-0.026	0.701
Plant machine operatives	-0.328	0.000	-0.239	0.000
Other	-0.320	0.000	-0.227	0.002
Got a "worse" occupation	0.210	0.000	0.157	0.001
Got a "better" occupation	0.149	0.001	0.103	0.023
Region: North	-0.045	0.249	-0.018	0.552
Changes region	0.497	0.000	0.530	0.000
Time Dummies	included	significant	included	significant
Log (income level)	included	not significant	included	not significant
Δ in income$_{t-1,t}$: "worse off"			-0.966	0.000
"better off"			1.547	0.000
Expectations Error$_{t-1,t}$ *				
"underpredict a lot"			-1.505	0.000
"underpredict a little"			-0.912	0.000
"overpredict a little"			1.002	0.000
"overpredict a lot"			2.198	0.000
Number of obs		25277		24531
P-value of χ^2		0		0
Log likelihood		-22705.2		-20687.3

Table 8: Ordered logits: Who holds what expectations ?