Measuring High-Frequency Income Risk from Low-Frequency Data

Paul Klein  Irina A. Telyukova
Simon Fraser University  University of California, San Diego

November 21, 2011

Abstract

In this paper, we estimate a monthly income process using standard household-level longitudinal annual income data, in order to understand the nature of income risk faced by households at high frequency, and to provide an input for models that wish to study household decision-making at higher frequency than available data. We find that at both monthly and annual frequencies, idiosyncratic earnings shocks have a highly but not perfectly persistent component. At the monthly frequency, transitory shocks account for most of the variance; at the annual frequency, it is the persistent component that accounts for most of the variance. We then use our estimated earnings processes to investigate whether frequency of decision-making per se has significant implications for risk-sharing in a standard incomplete-market model. We find that an annual and monthly such model yield very similar results in terms of risk-sharing at annual frequency.

1 Introduction

In the literature on household consumption-saving decisions under exogenously incomplete markets, it is typically assumed that households face some form of idiosyncratic risk. For any applications involving working-age households, a widely studied form of such risk is income uncertainty.
In order to study implications of idiosyncratic income uncertainty, researchers typically assume some process in income that may involve permanent, persistent and/or transitory components. In order to calibrate the models, researchers need to measure these components in the data. There is a large and active literature on estimating income uncertainty in the data; a few recent examples are Guvenen (2007), Guvenen and Smith (2010) and Heathcote et al. (2010). For the estimation of persistent processes, the econometrician needs longitudinal household-level data on income, which leads researchers to use, in most cases, survey data such as the Panel Study on Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX). Alternatively, as in Daly et al. (2011), administrative (register) data are used.

The limitation of all these datasets is that they are annual at best, and sometimes biennial, like the PSID in recent years. This means that the literature typically relies on these once-a-year observations of income to estimate income risk; models then typically use the same period length as in the data.\footnote{An exception is Erosa et al. (2011) who use an indirect inference approach to calibrate the four-monthly wage process on the basis of annual data.} This of course restricts model households to make decisions at an annual frequency. For some decisions, this is an acceptable approximation, but one can think of many other aspects of economic behavior for which we may prefer to have decision-making at quarterly, and even monthly, frequency. Understanding portfolio allocation, especially with respect to liquid assets, studying decisions to revolve or repay secured or unsecured debt, characterizing demand for money are some issues for which a high-frequency model would be preferred, even necessary. An example is Telyukova (2011), who addresses the question of co-existence in household portfolios of expensive credit card debt and low-return checking and savings accounts. In such a model, annual decision-making would be uninformative, as it would obscure the decision to revolve credit card debt each month, or to repay a portion of it using currently available liquid assets.

In this paper, we provide an estimate of a yearly and a monthly earnings process from annual PSID data. Based on an extension of Gervais and Klein (2010), we posit a monthly process underlying the observed annual income process; in both cases, we assume that income has a
permanent component, a persistent stochastic component, and a transitory component. We estimate the monthly process based on annual data, using GMM and moments of the autocovariance function. We find that in the estimated monthly model, the dominant component of income uncertainty is the transitory one and the persistent component is less important; for the annual model, the reverse is the case. Meanwhile, though the persistent component is quite persistent, with an annual autocovariance of about 0.95, it is not a random walk. Our approach to estimating the annual model can be thought of as a contribution to the ongoing debate about how best to estimate wage and earnings processes and the implications of each method for the relative importance of persistent and transitory shocks (see Domeij and Floden (2010) and Daly et al. (2011) for recent contributions) but this is incidental; our main innovation concerns the estimation of the monthly model using annual data.

In addition to providing estimates of monthly income risk, which we believe to be of interest in themselves, we also test whether frequency of decision making matters for risk-sharing implications of a standard consumption-saving model. We do this by computing an infinite-horizon version of the Huggett (1993) model at annual and monthly frequency. We find that while households in a monthly model may self-insure more completely in some cases, which then brings the model closer to the data than the annual case, the differences are small, and sensitive to parameters such as the borrowing limit. In a sense, it is reassuring to know that from the perspective of understanding risk-sharing broadly, the frequency of decision making is a matter of limited quantitative significance. Thus we conclude that the interest in high-frequency income risk should be driven by specific questions that require the modeling of frequent decision-making, where annual frequency would be insufficient for understanding the issues of interest.

In addition to the literature that estimates income uncertainty in the data, our work is related to the literature on risk-sharing in incomplete-market models. Some examples are Krueger and Perri (2004) and Kaplan and Violante (2009), who do this in calibrated models of household decision-making, and Blundell et al. (2008), who use econometric techniques to measure the degree of consumption risk sharing.
The rest of the paper is organized as follows. Section 2 describes the data we use in our estimation, the estimation procedure and results. We then apply these annual and monthly estimates in the context of the Huggett-style model, which we describe, calibrate and compute in section 3. Section 4 concludes.

2 Estimation of the Earnings Process

2.1 Data

In order to estimate the earnings process, we rely on the Panel Study of Income Dynamics (PSID). We employ the data from 1968 to 1997, which is the period during which data are available annually; after 1997, PSID becomes biennial. Our sample consists of individuals between the ages of 21 and 62. We consider different subsamples: all men and women, only men, and only male heads of households. For our purposes, all of these samples yield similar results. We also drop those with annual earnings below $2000 in 1968 dollars.

2.2 Procedure

The main challenge in estimating the earnings process is of course that we have annual data but want to estimate a monthly process. But before we tackle that issue, we have to choose a specification for the time series process of earnings that applies generally, regardless of the length of the time period.

Our choice of specification is designed to capture the key statistical properties in the micro data on earnings. Following many other authors, we extract the idiosyncratic component of log earnings by regressing log-earnings on a cubic in age, dummies for education, gender, race, marital status and birth cohort, and retaining the residuals. The question is how to model this
Our specification is as follows. The residual $y_{i,t}$ of monthly or annual log earnings is assumed to have three distinct components according to

$$y_{i,t} = \alpha_i + z_{i,t} + x_{i,t}$$

(1)

where we call $\alpha_i$ the *permanent* component (since it doesn’t change as a household ages), where the *persistent* component $z_{i,t}$ satisfies

$$z_{i,t-1} = \rho z_{i,t-1} + \varepsilon_{i,t}$$

where $\varepsilon_{i,t}$ is i.i.d. and where $x_{i,t}$ is also i.i.d. In contrast to Guvenen (2007), we do not allow for heterogeneous predictable age-earnings profiles.

Since in our model people have infinite lives, it makes sense to ignore the age dimension of observations, and we do not allow the variances of the shocks $\varepsilon_{i,t}$ and $x_{i,t}$ to depend on age. For reasons of parsimony, we also do not allow them to depend on cohort or on calendar time. Thus the only parameters that need to be estimated are $\sigma^2_\varepsilon$, $\rho$, $\sigma^2_x$ and $\sigma^2_\alpha$. This is done by GMM where the moments are simply the autocovariances $\Gamma_k = E[y_{i,t}y_{i,t+k}]$ where the kth covariance is computed as the average over all possible products $y_{i,t}y_{i,t+k}$ for which data is available and regardless of age.

The choice of specification described in Equation (1) is based on some striking features of the autocovariance function of the residuals, displayed in Figure 1. What we see there is that $\Gamma_k$ falls steeply as $k$ goes from 0 to 1 and then very gradually, with a near-constant rate of decay, as $k$ increases further. This is evidence in favor of the view that purely transitory shocks (or possibly measurement error) accounts for a large fraction of the total variance of earnings. It also points in the direction of idiosyncratic earnings having a component that is persistent but not quite a random walk.

These conclusions—the persistent component does not have a unit root and the transitory component is important—contrast somewhat with the influential work of Storesletten et al.
Figure 1: The autocovariance function of log earnings residuals.

(2004), who argue that the persistent component is a random walk and that it accounts for most of the variance. A representative recent statement of a widely shared view can be found in Kaplan and Violante (2009), who state that “it is well known that income changes are best described by a combination of very persistent and very transitory shocks”. This statement should perhaps not be taken literally. Clearly the change in income (or earnings) is not well represented as a sum of a very persistent and a very transitory component. Presumably what the authors want to say is that (log) earnings are well represented as the sum of a random walk and white noise, as in Blundell and Preston (1998). However, as Figure 1 shows, the autocovariance function of earnings is not consistent with such a representation. When $k \geq 1$, $\Gamma_k$ tends to decline more or less geometrically and at a non-negligible rate.

Moreover, it is not obvious (though it is possible) that $\Gamma_k$ tends to zero as it would if the earnings residual consisted of just a transitory and a persistent (but unit root) component. In fact $\Gamma_{k+1}/\Gamma_k$ starts out at about 0.92 at $k = 1$ and then increases somewhat towards unity.

As discussed in Domeij and Floden (2010) and Daly et al. (2011), given the specification in Blundell and Preston (1998), the estimation results are dramatically different depending on whether the estimation is done in levels or differences. Here we bypass that issue by using a different specification.
as \( k \) increases so that \( \Gamma_k \) appears to tend a strictly positive limit. More data would of course be required for any firm conclusion about this, but in the absence of any strong evidence that \( \Gamma_k \) tends to zero as \( k \to \infty \), it makes sense to allow for a truly permanent individual-specific component as well.

Given the specification (1) it is straightforward to derive the theoretical autocovariance function. It is

\[
\Gamma_k = \sigma_\alpha^2 + I_{\{k=0\}} \cdot \sigma_x^2 + \frac{\rho^k}{1 - \rho^2} \sigma_\varepsilon^2
\]

where of course \( I_{\{k=0\}} \) is an indicator function that equals one if \( k = 0 \) and zero otherwise. The GMM estimation chooses \( \rho, \sigma_\alpha^2, \sigma_\varepsilon^2 \) and \( \sigma_x^2 \) so as to minimize the (unweighted) distance between the first 20 theoretical autocovariances and their empirical counterparts. The estimation results are summarized in Table 1 and discussed in Section 2.4 below. In particular, notice that \( \rho \approx 0.95 \), a number quite far from unity.
2.3 A monthly Earnings Process

We now discuss how to estimate a monthly model. The statistical specification is still given by Equation (1). If we had monthly data, we could proceed exactly as above. But we do not. Nevertheless, a monthly model has implications for the annual autocovariance function and we can use these implications to estimate the parameters of the monthly model. This approach is similar to but extends that of Gervais and Klein (2010).

Computing the theoretical moments presents something of a challenge. They can of course be computed by simulation, following Lee and Ingram (1991). However, this is very wasteful in terms of CPU time. On the other hand, exact analytical formulas are not available because of Jensen’s inequality. To see this, denote (residual) log earnings in month $s$ by $v_{i,s}$. We then define annual earnings via

$$y_{i,a} = \ln \left( \frac{1}{12} \sum_{s=0}^{11} \exp \{ v_{i,12a+s} \} \right)$$

Thus our statistical model implies that

$$y_{i,a} = \ln \left( \frac{1}{12} \sum_{s=0}^{11} \exp \left\{ \alpha_i + x_{i,s} + \sum_{k=0}^{s} \rho^{s-k} \varepsilon_{i,k} \right\} \right).$$

Now if it weren’t for the ln followed by exp, analytical formulas would be available. If we remove these functions, we obtain approximate results, and the errors are small for the autocovariances with the partial exception of the 0th autocovariance (the variance). The approximating formulas are

$$\Gamma_0 = \frac{1}{12^2} \left[ \frac{(1 - \rho^{12})^2}{(1 - \rho)^2(1 - \rho^2)} + \frac{1}{(1 - \rho)^2} \left( 11 + \frac{\rho^2(1 - \rho^{22})}{1 - \rho^2} - 2 \rho \frac{1 - \rho^{11}}{1 - \rho} \right) \right] \sigma_\varepsilon^2 + \frac{1}{12} \sigma_x^2 + \sigma_\alpha^2$$

and

$$\Gamma_k = \frac{1}{12^2} \cdot \frac{\rho^{12k}}{(1 - \rho)^3} \left[ (1 - \rho^{12})(\rho^{12} + \rho^{-12} - 2) + \frac{(1 - \rho^{12})^2}{1 + \rho} \right] \sigma_\varepsilon^2 + \sigma_\alpha^2,$$

for $k > 0$.

To get a sense of how far these formulas go wrong, consider a case where $\rho = 0.9953$, $\sigma_\varepsilon^2 = 0.0015$ and $\sigma_x^2 = 0.7382$ (these are our estimated values) and let’s consider only one value of $\alpha^i$ so that
that component (which is small anyway) does not contribute to the variance. Then our formulas give the value 0.4674 for the theoretical annual standard deviation and 0.1514 and 0.1431 for the theoretical annual first and second autocovariances, respectively. Simulating 1,000,000 years of data, we obtain the simulated moments 0.4890, 0.1505 and 0.1422. Thus the effect of Jensen’s inequality is negligible for the first and second autocovariances but for the standard deviation the error is about 4.6 percent.

### 2.4 Estimation results

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_{\alpha}^2$</th>
<th>$\rho$</th>
<th>$\sigma_{\varepsilon}^2$</th>
<th>$\sigma_{\gamma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual</td>
<td>0.0073</td>
<td>0.9528</td>
<td>0.0160</td>
<td>0.0553</td>
</tr>
<tr>
<td>Monthly</td>
<td>0.0215</td>
<td>0.9953</td>
<td>0.0015</td>
<td>0.7382</td>
</tr>
</tbody>
</table>

Figure 3: Monthly model. Empirical and estimated theoretical yearly autocovariances.
In Table 1 we report the estimation results for the stationary versions of the annual and the monthly models, respectively. The key thing to notice is that the relative importance of the three components is quite different in the two cases. In the monthly model, about 80 percent of the variance is accounted for by transitory shocks and about 17 percent by persistent shocks. In the annual model, only about 23 of the total variance is accounted for by the transitory shocks and 87 percent by the persistent component, leaving almost nothing for the permanent component to account for. What explains this difference? Our findings are consistent with significant monthly noise that is to some extent washed out at the annual frequency. In any case, our first result is that it makes a difference whether we use an annual or a monthly model to measure the relative importance of transitory and persistent shocks.

We have just said that, in the monthly model, 80 percent of the variance is accounted for by transitory shock. That is the monthly variance. What about the yearly variance? Clearly much more of the variance of the transitory component washes out than that of the persistent component. As it turns out, in the annual model, about 40 percent of the annual variance of earnings is accounted for by the persistent component and about 55 percent by the transitory component. So the result remains that transitory shocks are more dominant than the persistent shocks, in terms of overall variance.

3 Application: Implications for Risk-Sharing

3.1 General-Equilibrium Incomplete-Market Model

In our computational experiments we use the canonical infinite-horizon model of Huggett (1993). Each period, households choose the level of consumption $c_t$ and saving $a_{t+1}$ in a real risk-free bond, given their current earnings and asset states. The earnings state is stochastic and idiosyncratic, and consists of three components as described in the estimation section: the permanent component $\alpha$, the persistent component $z_t$, and the transitory component $x_t$. The
permanent component is known to the household and does not change period to period. The persistent income state \( z_t \) is discrete and evolves according to a Markov process with associated transition function \( \Gamma_z(z'|z) \). The transitory shock \( x \) is likewise discrete and i.i.d. Denote by \( \mathbb{P}_x \) the probability of realization of a given shock \( x \).

The household problem in recursive formulation is

\[
V(\alpha, z_t, x_t; a_t) = \max_{c_{t+1}, a_{t+1}} \left( u(c_t) + \beta \sum_{z'} \sum_{x'} \Gamma_z(z'|z_t) \mathbb{P}_x V(\alpha, z_{t+1}, x_{t+1}; a_{t+1}) \right)
\]

s.t. \( c_t + a_{t+1} = \exp(\alpha + z_t + x_t) + a_t (1 + r_t) \)

\[
c_t \geq 0 \\
a_{t+1} \geq a
\]

Here, the household expectation of future shock realizations is written in terms of discretized shocks \( z_t \) and \( x_t \). The risk-free bond pays real return \( r_t \). The transition matrices give probabilities of future shock realizations conditional on current realizations. In the budget constraint, the current earnings realization is given by the product of the three components of (log) earnings. Households can borrow against the asset \( a \) subject to the borrowing constraint \( a_{t+1} \geq a \) where \( a < 0 \).

We will consider the stationary equilibrium of this economy, where the distribution of agents along the earnings and assets dimensions is constant over time. Define the transition function \( \Pi(s, B) \) from state \( s_t = (\alpha, z_t, x_t; a_t) \in S \) to the subset of state space \( B \) in the standard way. Denote by \( \Psi(s_t) \) the distribution of agents across the state space. Denote by \( g_a(s_t) \) and \( g_c(s_t) \) the decision rules with respect to saving and consumption in some period \( t \).

The stationary equilibrium for this economy is the set of functions \((g_c(s), g_a(s), \Psi(s), V(s), r(s))\) such that: (a) \( g_c(s) \) and \( g_a(s) \) are optimal decision rules for the household given the price \( r(s) \); (b) consumption and asset markets clear, so that \( \int g_c(s) d\Psi = \int \alpha z x d\Psi \) and \( \int g_a(s) d\Psi = 0 \); (3) the distribution \( \Psi \) is a stationary probability measure, so that \( \Psi(B) = \int \Pi(s, B) d\Psi \forall B \).

\footnote{Formally, \( B \) is a subset of the Borel \( \sigma \)-algebra on the state space \( S \).}
See, e.g., Huggett (1993) for a detailed discussion of this equilibrium concept, which is standard in the literature.

### 3.2 Calibration and Computation

We calibrate those parameters of the model that pertain to our earnings process from the estimation that we described above. The estimation yields parameters of the AR(1) processes for the persistent and transitory components of earnings which we then discretize using the method of Rouwenhorst (1995).

Following Huggett (1993), we test a variety of borrowing limits, from 0.5 to 5 times average annual earnings, computing the equilibrium interest rate in each case. We choose the standard CRRA functional form for the utility function: $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$. The remaining parameters are calibrated within the range in Huggett (1993). For the annual calibration, the discount factor $\beta^a$ is set at 0.97, which at monthly frequency yields $\beta^m = 0.9975$. The coefficient of risk aversion is $\sigma = 2$.

The algorithm to compute the model is standard: given a guess of the interest rate $r$, we solve the household problem, then compute the stationary distribution of agents across the state space, and then check market clearing conditions. We iterate on the price until the market clears given household optimization. To solve for the decision rules of the household, we use the endogenous grid method of Carroll (2006). The search for the market-clearing price is done using a bisection method.

### 3.3 Measures of Risk-Sharing

We devise several measures of risk sharing, all of which are inspired by considering two polar opposite cases: perfect risk-sharing and autarky. The measures are:
1. The regression coefficient $\beta$ of log-consumption changes on log-earnings changes:

$$\Delta \log(c_{it}) = \beta \Delta \log(e_{it}) + \epsilon_{it},$$

where $e_{it} = \alpha z_{it} x_{it}$. (See Krueger and Perri (2004).)

2. Variance of log-consumption relative to variance of log-earnings.

3. Autocorrelation of log-consumption.

The first two measures share the property that they are one in autarky and zero under perfect risk sharing. All three form a convenient metric through which model predictions are summarized and compared across the models of differing frequencies. For both models, the measures are computed for annual consumption and earnings, which in the monthly case requires adding up monthly data to annual frequency.

### 3.4 Results

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$a$</th>
<th>$r$</th>
<th>$\beta$</th>
<th>$\frac{\text{var}(\ln(c))}{\text{var}(\ln(e))}$</th>
<th>$\rho(\ln(c))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual</td>
<td>0.5</td>
<td>1.4</td>
<td>0.75</td>
<td>0.99</td>
<td>0.78</td>
</tr>
<tr>
<td>Monthly</td>
<td>1.3</td>
<td></td>
<td>0.72</td>
<td>0.96</td>
<td>0.76</td>
</tr>
<tr>
<td>Annual</td>
<td>1.5</td>
<td>1.6</td>
<td>0.67</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>Monthly</td>
<td>1.5</td>
<td></td>
<td>0.63</td>
<td>0.96</td>
<td>0.78</td>
</tr>
<tr>
<td>Annual</td>
<td>2.5</td>
<td>1.8</td>
<td>0.57</td>
<td>0.97</td>
<td>0.84</td>
</tr>
<tr>
<td>Monthly</td>
<td>1.6</td>
<td></td>
<td>0.58</td>
<td>0.95</td>
<td>0.80</td>
</tr>
<tr>
<td>Annual</td>
<td>4</td>
<td>1.9</td>
<td>0.46</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td>Monthly</td>
<td>1.7</td>
<td></td>
<td>0.52</td>
<td>0.94</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Note: $a$ is the borrowing limit given as a factor multiplying average annual earnings. $\beta$ is the regression coefficient of log consumption changes on log earnings changes. $\rho(\ln(c))$ is the autocorrelation of log consumption. The interest rate $r$ is the annual rate.
In Table 2, we report the results from the computational experiment. First, notice that as the borrowing limit increases, the interest rate increases as well, since the bond becomes less valuable at the margin as a hedge against earnings risk when the borrowing constraint is relaxed. This results is consistent with the findings of Huggett (1993). Second, as the borrowing limit increases, the regression coefficient falls, suggesting increasing ability to self-insure against risk, as we would expect (see, e.g., Krueger and Perri (2004)); moreover, our regression coefficients are in line with theirs. The other measures of risk-sharing confirm this tendency as well.

If we now compare the models of differing frequency, the main result is that frequency of decision-making on its own does not make a dramatic difference. All of our measures of risk-sharing are close in the two models, as are implied equilibrium interest rates. Generally, the implication is that with monthly decision-making frequency, agents are able to self-insure slightly more. Moreover, the interest rate and self-insurance measures largely suggest that the difference between the two models increases as the borrowing limit increases. However, the difference is quantitatively small. There is also some non-monotonicity in the regression coefficient result: for low borrowing limits, such as 0.5 or 1.5, the models imply that monthly decision-making allows better self-insurance than annual decision-making. But for the borrowing limit at or above 2.5, the regression coefficient implies better risk-sharing at annual frequency. The overall message is that the model predicts similar degree of risk-sharing regardless of the frequency.

4 Conclusion

In this paper, we used the standard annual longitudinal household-level income data from PSID to study the nature of income risk that faces households at monthly frequency. In particular, we estimated an annual income process typical of the literature, but also posed and estimated an underlying monthly income process. We view the results of our estimation as interesting in their own right, as they shed light on the properties of risk that households face in the data between the times that we observe them in our surveys. We find that monthly earnings, though
quite persistent, have a very significant transitory component.

We used our estimates to test whether frequency of decision-making is important in deriving implications of incomplete-market models for risk-sharing between households. In the context of the Huggett (1993) model, we find that frequency alone does not lead to significant differences in risk-sharing. We view this as encouraging for the literature that has studied risk-sharing predominantly in annual models.

We believe our results to be valuable to anyone who is interested in studying household decisions that are not usefully modeled at low frequency, so that an annual model would be too restrictive, and even uninformative, for the question of interest. For example, aspects of portfolio allocation may be best studied at the frequency at which households are paid their labor income, especially if the interest might be in money demand or liquid assets more generally. The same goes for applications pertaining to household decisions to borrow or revolve debt. Our results can be used directly as an input to such calibrated models, or our methodology can be applied more broadly to study high-frequency risk of other types or from other data sets.
References


