DOCUMENTOS DE TRABAJO

Identification of Earnings Dynamics Using Rotating Samples Over Short Periods: The Case of Chile

Carlos Madeira

N.º 754 Abril 2015

BANCO CENTRAL DE CHILE

DOCUMENTOS DE TRABAJO

Identification of Earnings Dynamics Using Rotating Samples Over Short Periods: The Case of Chile

Carlos Madeira

N.º 754 Abril 2015

BANCO CENTRAL DE CHILE

Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

Documento de Trabajo
N° 754

Working Paper
N° 754

IDENTIFICATION OF EARNINGS DYNAMICS USING ROTATING SAMPLES OVER SHORT PERIODS: THE CASE OF CHILE[*]

Carlos Madeira
Banco Central de Chile

Abstract
Due to the absence of longitudinal data, empirical studies ignore labor income dynamics in developing economies, where earnings inequality is highest and social insurance is weakest. This work proposes a dynamic earnings process with two distinct shocks: unemployment spells and the wage of workers who stay employed. I then show this income process can be estimated from employment surveys with a rotating sample design, which are available for several countries. Applying this procedure to Chilean data I show wage volatility and unemployment rates are highly heterogeneous across workers. Unemployment spells are the most important source of earnings risk for workers.

Resumen
Debido a la ausencia de datos panel, los estudios empíricos ignoran la dinámica de ingreso laboral en economías en desarrollo, donde la desigualdad es más alta y la protección social tiene menor cobertura. Este trabajo propone un proceso dinámico de ingresos con dos choques distintos: periodos de desempleo y el sueldo de los trabajadores que permanecen empleados. Se muestra que este proceso dinámico puede ser estimado a través de encuestas de empleo con un esquema de muestras rotativo, lo que son datos disponibles en muchos países. Aplicando este procedimiento a Chile se muestra que la volatilidad de sueldos y las tasas de desempleo son altamente heterogéneas entre distintos trabajadores. Los periodos de desempleo son la fuente más relevante de riesgo de ingresos de los trabajadores.

[*] Todos los errores y omisiones son de exclusiva responsabilidad del autor. Email: cmadeira@bcentral.cl.

1 Introduction

Labor earnings represent the most important source of income for households. The volatility of labor earnings is important to study a variety of topics, including: earnings inequality and its persistence (Storesletten, Telmer and Yaron, 2004, Krueger, Perri, Pistaferri and Violante, 2010), the welfare costs of business cycles (Storesletten, Telmer and Yaron, 2001, De Santis, 2007), life-cycle consumption (Carroll and Samwick, 1997, Gourinchas and Parker, 2002, Attanasio and Weber, 2010), the optimality of households' asset portfolios (Viceira, 2001), and the welfare effects of social insurance schemes (Hubbard, Skinner and Zeldes, 1995, Low, Meghir and Pistaferri, 2010).

However, despite the importance of earnings and wage dynamic processes, the study of income risk and volatility is limited to a few developed countries (such as the USA, Germany, UK, Canada and Australia), which have published household longitudinal surveys (Schluter, 1998, Krueger, Perri, Pistaferri and Violante, 2010, Zhang, 2010).[1] Household and workers' longitudinal surveys are demanding in terms of budget and suffer high sample attrition over the years, which has made this kind of data unavailable in most countries. In the developing world income inequality is much larger than in developed economies (Sala-i-Martin, 2006), while social insurance nets such as unemployment benefits, food programs, disability support and health protection are weaker (Dethier, 2007, Cho, Margolis, Newhouse and Robalino, 2012). Therefore knowing the volatility of income and the persistence of negative shocks is of much higher importance in developing countries.

In this paper I propose a method for estimating the income volatility of individual workers and households by using rotating samples from consecutive income and employment surveys. Annual surveys with information on income and employment status are available for 61 countries, including all the 34 members of the OECD (OECD, 2013a,b) and 27 developing economies (UNU-WIDER, 2013).[2] Although being classified as cross-sectional datasets, employment surveys include a rotating sample design and sample the same housing units for a fixed number of quarters to save operating

[1] In the United States longitudinal household datasets include at least the PSID (1968-) and the SIPP (1984-). Other longitudinal household surveys exist for Australia (HILDA, 2001-), Canada (SLID, 1993-), Germany (SOEP, 1984-), the UK (BHPS, 1991-), and for 14 member countries of the European Union (ECHP, 1994-2001).

[2] This list includes at least Argentina, Armenia, Azerbaijan, Belarus, Bolivia, Brazil, Bulgaria, Costa Rica, Dominican Republic, Ecuador, El Salvador, Honduras, Kyrgyz Republic, Latvia, Lithuania, Macedonia, Moldova, Panama, Paraguay, Peru, Romania, Russia, Serbia, Taiwan, Ukraine, Uruguay, and Venezuela.

costs (International Labour Office, 1990, Eurostat, 2012, United Nations, 2008), therefore the procedure proposed in this paper can be applied to study income risk for a large area of the world. Rotating samples differ from panel or longitudinal data in the sense that panel surveys attempt to include all members of the initial sample, while rotating designs do not focus on following the entire initial sample. However, it is possible to create longitudinal data for a finite number of T periods (Kish, 1998) from rotating samples by matching the same housing units over different surveys and then matching individuals in the same housing unit based on their sex, age and education. These small T panel samples are much richer for studying idiosyncratic shocks than other alternatives such as pseudo-panel built from cohort data, since cohort averages only capture aggregate shocks and ignore the idiosyncratic risks faced by individual workers.

Since employment surveys keep their housing units for a period of 12 to 15 months (International Labour Office, 1990), then there is a sample overlap of 80% between consecutive quarters and 20% from year to year. Depending on whether the survey frequency is monthly, quarterly, or annual, then these overlapping samples allow the construction of several series of panel samples with 2 to 8 periods for each unit. These longitudinal samples are still informative about earning dynamics over short horizons. Obviously, small T panels fail to distinguish on whether shocks are of a permanent or temporary nature, since such difference is meaningless over a short period. Some studies discuss whether income shocks are persistent over an entire lifetime (Guvenen, 2007, Browning, Erjnaes, Alvarez, 2010), which is an important question that cannot be answered with a short panel.

I then propose a simple model of labor earnings dynamics, including three sources of fluctuations: wage volatility, loss of employment, and the replacement ratio of income benefits to labor earnings during the unemployment spell. The model accounts for both a continuous wage shock suffered by employed workers and a discrete earnings shock due to transitions into and out of unemployment (similar to Low, Meghir and Pistaferri, 2010). The income process can then be reliably estimated using a set of moments from the panel observations, as well as from the cross-sectional dataset. In particular, flows into and out of unemployment are re-weighted to match the stocks of workers, which adjusts for panel attrition and margin error (as in Abowd and Zellner, 1985, Shimer, 2012).

This estimation procedure is then applied to Chilean data using the National Employment Survey (ENE, 1990Q1-2012Q4). The ENE dataset covers a large sample of 35,000 households and their individual members at a quarterly frequency. Information is collected on the demographic characteristics, occupation and employment status of each member of the household, providing a cross-sectional sample around 50,000 workers in each quarter. Also, the survey collects earnings information for all household members in the fourth quarter of each year, allowing to estimate the annual earnings dynamics of different worker types according to their characteristics in terms of gender, age, education, industry and income level. The results show that Chile has a fluid labor market, with unemployment inflow and outflow rates similar to the United States and substantially higher than other OECD countries (Elsby, Hobijn and Sahin, 2013). Also, the average employed worker faces idiosyncratic income shocks with a standard deviation of 18% at an annual frequency, which is substantially higher than the 6% standard deviation estimated by previous studies using Chilean pseudo-panel data with cohort means (Huneeus and Repetto, 2005).

Unemployment, separation and job-finding rates, and wage volatility are heterogeneous across worker types, but there is a strong cyclical component to job creation and destruction which affects all groups. Workers transiting into unemployment suffer an average discrete earnings loss around 60% of their previous income. However, income loss during unemployment is also heterogeneous, with workers of the lowest income quintile losing 90% or more of their income, which is due to the individual accounts applied by the Chilean unemployment insurance scheme. Some workers have more stable income than others and the standard deviation of income shocks for employed workers can vary from as low as 5% to as high as 35%. Low income workers experience both a higher probability of unemployment and a larger wage volatility. However, workers of lower income have a higher job-finding rate and therefore face shorter unemployment spells, perhaps because their job matches involve less specific human capital. A decomposition of the income variance in terms of job loss and wage volatility finds that employment transitions explain more than 50% of the wage fluctuations across all income groups, especially among the middle class where finding or losing a job explains over 75% of the earnings variance. Also, the share of the income variance due to employment transitions is countercyclical and increases during recessions for all income groups.

This paper is organized as follows. In Section 2 I specify the stochastic income process of the workers. Section 3 shows how this dynamic process can be estimated using moments from panel data with two periods. Section 4 describes the Chilean survey data on income and employment. Section 5 shows how employment transitions and income risk vary across workers and over the years. Finally, section 6 concludes with implications for policy and future research.

2 Workers' stochastic earnings process

The specified model decomposes underlying earnings fluctuations between continuous shocks to working wages and a process of job destruction and arrival. Also, while unemployed each worker has differential access to unemployment insurance or informal sources of earnings. For simplicity I ignore shocks to hours worked and the transitions into and out of the labor force. Thus individuals face multiple sources of uncertainty: in each period employed individuals may enter unemployment or receive a shock to their wage earnings (due to, for instance, a wage raise, a change in hours worked or bonus payments); also, unemployed workers have different insurance coverage and may or not be offered a job. A large literature studying labor earnings focus on a decomposition into continuous transitory and permanent shocks while excluding unemployment spells and job transitions (Abowd and Card, 1989, Carroll and Samwick, 1997). However, as Low, Meghir and Pistaferri (2010) show, unemployment spells and job transitions explain a significant portion of annual earnings shocks, especially of the transitory component. In my framework, the income risk due to unemployment spells is specifically estimated by the separation rate (i.e., the flow into unemployment), the duration of an unemployment spell (i.e., the difficulty in finding a new job) and by the access to unemployment benefits (similar to Low, Meghir and Pistaferri, 2010).

Let $Y_{i,t+1}$ denote the total labor earnings at time $t+1$ of worker i, which has observable characteristics x_i. Worker i transits between employment states denoted by $U_{i,t+1}$, which can either be 0 if employed or 1 if unemployed. Since there are only two employment states, labor status dynamics can be summarized into unemployment to employment (UE) and employment to unemployment (EU) transitions. The stochastic process of the workers' earnings can be specified as the result of a continuous wage shock plus discrete employment transitions:

1) $Y_{i,t+1} = G_{t+1|x_i} Y_{i,t} \exp(\upsilon_{i,t+1}) R_{t+1|x_i}^{U_{i,t+1}-U_{i,t}}$,

where $G_{t+1|x_i}$ denotes the mean aggregate income growth of workers with characteristics x_i in period $t+1$ and $\upsilon_{i,t+1}$ is an idiosyncratic shock to the wage of worker i. $\upsilon_{i,t+1}$ is assumed to be independent across the agents i, with mean zero, $E\left[\upsilon_{i,t+1}\right] = 0$, and variance $Var(\upsilon_{i,t+1}) = \sigma^2_{t+1|x_i}$. Also, if entering unemployment each worker i receives a proportion of his wage income, $R_{t+1|x_i} \in (0,1)$, given by the replacement ratio of his unemployment benefits coverage plus informal earnings.

If workers' employment status is unchanged ($U_{i,t+1} = U_{i,t}$), then $Y_{i,t+1} = G_{t+1|x_i} Y_{i,t} \exp(\upsilon_{i,t+1})$, therefore only the continuous wage shock matters. However, the worker suffers an immediate income loss proportional to $R_{t+1|x_i}$ if separating out of his job ($U_{i,t+1} - U_{i,t} = 1$) and a relative income gain of $R^{-1}_{t+1|x_i}$ if transiting out of unemployment ($U_{i,t+1} - U_{i,t} = -1$). Labor status shocks follow a Markov transition process with job-separation and job-finding rates given by:

2) $\lambda^{EU}_{t+1|x_i} = \Pr(U_{i,t+1} = 1 \mid U_{i,t} = 0, x_i)$, and,

3) $\lambda^{UE}_{t+1|x_i} = \Pr(U_{i,t+1} = 0 \mid U_{i,t} = 1, x_i)$.

Given this Markov transition process, the unemployment rate $u_{t+1|x_i}$ of each worker type can be expressed as a dynamic equation of the previous period's unemployment probability plus the transitions into and out of unemployment (Shimer, 2012):

4) $u_{t+1|x_i} = \Pr(U_{i,t+1} = 1 \mid x_i) = u_{t|x_i}(1 - \lambda^{UE}_{t+1|x_i}) + (1 - u_{t|x_i})(\lambda^{EU}_{t+1|x_i})$.

All the model's parameters $\left\{G_{t|x} = \exp(g_{t|x}), R_{t|x} = \exp(r_{t|x}), \sigma^2_{t|x}, \lambda^{UE}_{t|x}, \lambda^{EU}_{t|x}, u_{t|x}\right\}$ are heterogeneous according to both the time period t and the characteristics of worker i. This earnings model can be more easily expressed in additive terms of log-earnings:

5) $y_{i,t+1} = \ln(Y_{i,t+1}) = g_{t+1|x_i} + y_{i,t} + (U_{i,t+1} - U_{i,t}) r_{t+1|x_i} + \upsilon_{i,t+1}$,

where the lower-case letters denote the natural logarithm of the equivalent upper-case parameter. In principle, the term $\upsilon_{i,t+1}$ may represent a mix of shocks to permanent income, transitory income, and measurement error. However, since the panel is short, it is not possible to distinguish how much of its volatility is due to transitory shocks or simply survey mis-report by the agents. We can therefore obtain the model's parameters as moments of the log-earnings process:

6) $E\left[y_{i,t+1} \mid x_i, y_{i,t}, U_{i,t+1} = U_{i,t}\right] = g_{t+1|x_i} + y_{i,t}$,

7) $E\left[y_{i,t+1} \mid x_i, y_{i,t}, U_{i,t+1} \neq U_{i,t}\right] = g_{t+1|x_i} + y_{i,t} + (U_{i,t+1} - U_{i,t}) r_{t+1|x_i}$, and,

8) $E\left[(y_{i,t+1} - E\left[y_{i,t+1} \mid x_i, y_{i,t}, U_{i,t+1}, U_{i,t}\right])^2 \mid x_i\right] = \sigma^2_{t+1|x_i}$.

Note that although this model specification is similar to a unit-root process in earnings, it is not the goal of this work to separate which wage shocks are permanent and transitory, since the data application does not allow to study income persistence after 18 months. Empirical studies have determined that labor earnings are notoriously persistent from year to year (Abowd and Card, 1989, Carroll and Samwick, 1997, Browning, Erjnaes, Alvarez, 2010), while others show the importance of heterogeneous life income profiles and that income shocks partially revert over periods of several years, rejecting a unit-root specification (Geweke and Keane, 2000, Guvenen, 2007, Browning, Erjnaes, Alvarez, 2010). Also, some studies have found it is important to take into account that workers are highly heterogeneous and that some workers may have highly volatile income while others do not (Meghir and Pistaferri, 2004, Browning, Erjnaes, Alvarez, 2010). The model summarized in equations 1)-8) is able to account for diverse levels of heterogeneity due to different workers's observable characteristics (x_i).

In this model I take the perspective of specifying the workers' stochastic income realizations and unemployment transitions as final outcomes. This implies that the probability of falling into unemployment or receiving a wage shock are taken to be pure statistical processes conditional on the workers' characteristics. It is possible to further decompose these outcomes of income and unemployment as a mix between a purely exogenous shock to the agents' environment plus an endogenous choice about how much effort to put working or in job search. However, this model focuses purely on measuring the separate components of earnings volatility in terms of wage shocks and unemployment spells, not on the endogeneity of its outcomes. Some empirical literature finds a low correlation between hourly wage shocks and hours of work (Abowd and Card, 1989), which favors the view that agents' labor income is almost an exogenous process to the worker and not the result of contemporary endogenous choices of effort[3].

[3] It is relevant to note that labor market outcomes may certainly be affected by deliberate choices prior to adulthood, such as the occupation chosen, the completion of college, or even the human capital accumulated during one's early childhood. Some reviews of the empirical literature on income, education and human capital suggest that one's labor income is largely determined by age 16 (Keane and Wolpin, 1997).

3 Model identification from data moments of short panels

3.1 Identification from Panel Moments with at least 2 periods

The specified model can be easily estimated for any economy with surveys that measure employment transitions and income for a panel of workers in consecutive periods. The unemployment transition parameters $\left\{\lambda^{UE}_{t+1|x}, \lambda^{EU}_{t+1|x}, u_{t+1|x}\right\}$ are easily identified by the three moments defined in equations 2), 3) and 4). Also, the income growth, wage volatility and replacement ratio during unemployment, $\left\{G_{t+1|x}, R_{t+1|x}, \sigma^2_{t+1|x}\right\}$, are identified from the conditional moments in equations 6), 7) and 8).

3.2 Use of further moments from Cross-sectional waves

Using 6 panel data moments to identify the 6 parameters of the model is an elegant and succinct approach for estimation. However, I am considering explicitly the case in which estimation is carried out on cross-sectional waves and that some of these cross-sectional waves have a common sample that can be used as a panel data with a small number of periods T. The panel data observations are therefore only a small sub-sample of a larger pool of cross-sectional samples and excluding the other observations available in these cross-sectional samples implies a substantial loss of efficiency.

One useful information in the pooled cross-sections is that employment surveys ask how many weeks the unemployed worker has been searching for a job or when was the last time the worker was at a job, allowing us to know both $\Pr(U_{i,t+1} = 1 \mid x_i)$ and $\Pr(U_{i,t+1} = 1 \mid U_{i,t} = 0, x_i)$. It is also possible to estimate the moment $\Pr(U_{i,t} = 1 \mid x_i)$ from the previous cross-sectional survey, then apply the labor market equilibrium condition (equation 4) to obtain a third moment and use these 3 moments to estimate all the employment transition parameters $\left\{\lambda^{UE}_{t+1|x}, \lambda^{EU}_{t+1|x}, u_{t+1|x}\right\}$ from the cross-sectional data, as in Juhn, Murphy and Topel (1991).

Also, given the specified income process then the evolution of the mean and variance of the income distribution of each cohort allows us to identify $\left\{G_{t+1|x}, R_{t+1|x}, \sigma^2_{t+1|x}\right\}$ as well. Therefore one may add cross-sectional moments of income to enhance the efficiency of estimation with:

9) $E\left[y_{i,t+1} \mid x_i, U_{i,t+1}\right] = g_{t+1|x_i} + E\left[y_{i,t} \mid x_i, U_{i,t+1}\right] + r_{t+1|x_i} E\left[\left(U_{i,t+1} - U_{i,t}\right) \mid x_i, U_{i,t+1}\right]$, and,

10) $Var\left[y_{i,t+1} \mid x_i, U_{i,t+1}\right] = \sigma^2_{t+1|x_i} + Var\left[y_{i,t} \mid x_i, U_{i,t+1}\right] + r^2_{t+1|x_i} Var\left[(U_{i,t+1} - U_{i,t}) \mid x_i, U_{i,t+1}\right]$,

where $Var[z]$ denotes the variance or demeaned second moment of z, i.e. $E[(z - E[z])^2]$. Equations 9) and 10) add 4 extra moments for the estimation, since $U_{i,t+1} \in \{0, 1\}$. However, equation 10) for unemployed workers provides little additional information on wage variability $\sigma^2_{t+1|x}$, therefore I use only the moment 10) for employed workers ($U_{i,t+1} = 0$). Some inputs of these expressions, such as $E[y_{i,t} \mid x_i, U_{i,t+1}]$, are moments of income at time t conditional on the employment status of time $t+1$. These moments can be derived using Bayes theorem, since:

11) $E[z_{i,t} \mid x_i, U_{i,t+1}] = \sum_{a=0}^{1} \Pr(U_{i,t} = a \mid x_i, U_{i,t+1}) E[z_{i,t} \mid x_i, U_{i,t} = a]$,

with $\Pr(U_{i,t} = a \mid x_i, U_{i,t+1} = b) = \dfrac{\Pr(U_{i,t+1} = b \mid x_i, U_{i,t} = a)\Pr(U_{i,t} = a \mid x_i)}{\Pr(U_{i,t+1} = b \mid x_i)}$. Therefore by replacing $z_{i,t}$ as $y_{i,t}$, $(U_{i,t+1} - U_{i,t})$, $(y_{i,t} - E(y_{i,t}))^2$, or $((U_{i,t+1} - U_{i,t}) - E(U_{i,t+1} - U_{i,t}))^2$, one solves all the unorthodox moments included in equations 9) and 10) as weighted sums from moments easily estimated from the cross-sectional sample of time t, where the weights are given by the employment transition probabilities $\Pr(U_{i,t+1} = b \mid x_i, U_{i,t} = a)$ times the ratio of the unconditional employment rates $\dfrac{\Pr(U_{i,t} = a \mid x_i)}{\Pr(U_{i,t+1} = b \mid x_i)}$. Expression 11) is intuitive if one considers the case in which unemployment at $t+1$ and t are independent, which gives $E[z_{i,t} \mid x_i, U_{i,t+1} = a] = E[z_{i,t} \mid x_i, U_{i,t} = a]$. Since unemployed workers are more likely to have been unemployed in the previous period, then one must correct the mean wage and the wage variance by special weights.

An alternative moment condition relative to equation 10) can be obtained by thinking of the term $y_{i,t} + (U_{i,t+1} - U_{i,t}) r_{t+1|x_i}$ as an initial condition problem for the variance of $y_{i,t+1}$, which is a standard framework in panel data models with a small T (Wooldridge, 2002). In this case, one can specify a functional approximation for the $Var\left[y_{i,t} + (U_{i,t+1} - U_{i,t}) r_{t+1|x_i} \mid x_i, U_{i,t+1}\right]$ and then use such a condition to obtain a moment for $\sigma^2_{t+1|x_i}$. Since $\sigma^2_{t+1|x_i}$ denotes the wage volatility of employed workers, perhaps it is easy to focus on the moment for employed workers at time $t+1$ and assume for simplicity the following functional form, $Var\left[y_{i,t} - U_{i,t} r_{t+1|x_i} \mid x_i, U_{i,t+1} = 0\right] = (1 - g(x_i)) Var\left[y_{i,t+1} \mid x_i, U_{i,t+1} = 0\right]$. This functional form can then be used to get the following moment for the volatility of the wage process, $\sigma^2_{t+1|x_i}$:

12) $Var\left[y_{i,t+1} \mid x_i, U_{i,t+1} = 0\right] = \sigma^2_{t+1|x_i} + Var\left[y_{i,t} - U_{i,t} r_{t+1|x_i} \mid x_i, U_{i,t+1} = 0\right] \Longleftrightarrow$

$\Longleftrightarrow \sigma^2_{t+1|x_i} = g(x_i) Var\left[y_{i,t+1} \mid x_i, U_{i,t+1} = 0\right]$.

Equation 12) clearly results from a strong assumption on the functional form of $Var(y_{i,t} - U_{i,t}r_{t+1|x_i} \mid x_i, U_{i,t+1} = 0)$. However, it is a simple condition to interpret, since it imposes that the wage volatility, $Var(\upsilon_{i,t+1}) = \sigma^2_{t+1|x_i}$, has a common component that is constant over all time periods and one time varying change that affects equally the variances of $y_{i,t+1}$ and of $\upsilon_{i,t+1}$[4].

3.3 Decomposition of the earnings' variance

To complete this section, I note it is possible to decompose the variance of the earnings innovations of the workers in terms of the role of the continuous wage shocks ($\upsilon_{i,t+1}$) and the discrete shock due to transitions into and out of unemployment ($\Delta U_{i,t+1} \equiv U_{i,t+1} - U_{i,t}$). From equation 5) one can easily derive the variance of the earnings innovations into three additive components:

13) $Var(y_{i,t+1} - y_{i,t} - g_{t+1|x_i}) = Var(\upsilon_{i,t+1}) + Var(r_{i,t+1}\Delta U_{i,t+1}) + 2COV(\upsilon_{i,t+1}, r_{t+1|x_i}\Delta U_{i,t+1})$.

If one assumes independence between the wage and unemployment shocks, this gives a simplified expression for the share of the income variance represented by unemployment dynamics:

14) $\theta_{t+1|x_i} = \dfrac{Var((U_{i,t+1} - U_{i,t})r_{t+1|x_i})}{Var(y_{i,t+1} - y_{i,t} - g_{t+1|x_i})}$.

It is questionable if unemployment transitions are independent of wage shocks, since a positive wage shock should decrease the probability that a worker and a firm would decide to terminate their relationship and a negative wage shock should increase the probability that the relationship is terminated. Therefore one could plausibly assume that $COV(\upsilon_{i,t+1}, r_{i,t+1}(U_{i,t+1} - U_{i,t})) \leq 0$. However, since there is no data on what was the last wage offered to the worker prior to the job termination, it is not possible to estimate the value of this covariance term and therefore I ignore the covariance term when computing the role of unemployment shocks in the overall wage variance, $\theta_{t+1|x}$. For this reason perhaps it is more appropriate to understand $\theta_{t+1|x_i}$ as an upper bound on the role of the unemployment transitions in terms of income shocks rather than its precise value.

[4] Obviously, one could add further constraints such as additivity in the vector x plus a time dummy, i.e. $\sigma^2_{t+1|x_i} = \exp(\theta x_i + d_t)$. This extra restriction would specifically impose that the wage variance is additive in the different characteristics of x and that the time shock to wage volatility is the same for all workers. However, such parametric forms seem too restrictive, therefore I will not apply such specific assumptions here.

3.4 Survey weights, attrition, asymptotic errors and confidence intervals

The panel data and cross-sectional moment conditions described above provide a system of 12 moments to estimate 6 parameters for each time period t and each worker type x. Without loss of generality, I consider the simple case in which x_i results in a set of discrete mutually-exclusive categories, which have a positive representation in the sample, therefore all the moments are fully non-parametric, without considering linear or kernel approximations. Since this system of moment conditions provides for over-identification of the parameters, I obtain estimates by equally weighting all the moment conditions. I choose the one-step GMM, since Monte Carlo studies show the optimal 2-step GMM suffers a severe finite-sample bias even for large samples and is often dominated by the equally weighted one-step GMM (Altonji and Segal, 1996). Finally, in survey data each household is selected with different probabilities and it is necessary to apply population weights or expansion factors, $f_{i,t}$, to each worker observation, therefore all moments z are computed with population weights, $E\left[f_{i,t}z_{i,t}\right]$. Cross-sectional surveys provide the population weights (or inverse selection probabilities) for each worker observation, $f_{i,t}$. However, matching rotating samples to create a panel data is affected by attrition rates. For this reason, I use the methodology proposed by Abowd and Zellner (1985) and adjust the population weights of each panel sample for heterogeneous attrition rates of workers according to their characteristics and employment status, $f_{i,t}^{Panel} = f_{i,t}\dfrac{\Pr(U_{i,t}, x_i \mid CS)}{\Pr(U_{i,t}, x_i \mid P)}$, where CS denotes the joint probability in the Cross-sectional sample and P its equivalent in the Panel sample.

The traditional GMM asymptotic variance matrix assumes that the selection probabilities for each worker are known with full certainty, which is incorrect in surveys where households have different rates of non-response and panel attrition (Rao and Wu, 1988). However, all the model's parameters are continuous statistics, therefore asymptotically consistent standard errors and confidence intervals can be obtained by bootstrap (Horowitz, 2001, Rao and Wu, 1988). I build 100 bootstrap replica samples for each one of the panel and cross-sectional datasets. One advantage of bootstrap is that each sample replica incorporates uncertainty about the true panel attrition process and the population weights $f_{i,t}$ by drawing with replacement inside each population strata. Finally, all the model's coefficients are re-estimated on each bootstrap replica.

4 Data Description

To estimate the earnings model I use the quarterly Chilean Employment Survey or Encuesta Nacional de Empleo (ENE, 1990Q1-2012Q4), which covers 35,000 home addresses at the national level every quarter. Since households can have either several working members or none, then this sample of homes corresponds to slightly more than 45,000 labor force participants per quarter. The ENE survey is implemented by the Chilean Institute of National Statistics, therefore responses are compulsory by law and non-response is low. Furthermore, during the 4th quarter of each year the ENE has an Income module, the ESI (1990Q4-2012Q4), which provides a measure of labor and non-labor income that can be used to estimate the workers' quarterly wage growth ($G_{t+1|x}$), wage volatility ($\sigma^2_{t+1|x}$), and replacement ratios ($R_{t+1|x}$). The definition of $y_{i,t}$ considers only labor earnings (whether from formal or informal work) plus unemployment benefits, excluding other forms of earnings such as asset income and implicit rental income from properties and real assets. All wages and earnings in the survey are reported as after-tax amounts.

The earnings dynamics' parameters are heterogeneous according to the workers' characteristics x_i. Since the ENE/ESI sample is large I consider a wide range of observable heterogeneity, by classifying workers in 540 mutually exclusive types. Each worker's type is given by a set of observable demographic and professional characteristics, x_i = {Santiago Metropolitan city or Outside, Industrial Activity (primary, secondary, service sectors), Gender, Age ($\leq$ 35, 35 $-$ 54, and $\geq$ 55 years), Education (less than secondary schooling, secondary school or technical education, college), and Worker's Income quintile}. The worker's income quintile is measured as a pre-determined variable based on the workers' reported income in the current period or the previous year, $Q_{i,t} = \arg \max_{q\in\{1,2,3,4,5\}} q$ subject to $\frac{100F(y_{i,j} \mid U_{i,j}, x_i^*)}{20} \geq q-1,\ j \leq t$, where $F()$ is the empirical cdf of income conditional on unemployment status and the workers' non-income characteristics x_i^*[5]. This income quintile classification assumes the proportions of each quintile in the employed and unemployed pools are different, due to distinct unemployment probabilities conditional on x_i^*.

As explained in the previous section, the estimation of the earnings dynamics and employment transitions is carried out on a combination of moments from panel and cross-sectional samples.

[5] The empirical cdf of the income distribution is obtained non-parametrically as $F(p \mid U_{i,j} = a) \equiv \frac{1}{n_j} \sum_{k=1,U_{k,j}=a}^{n_j} \frac{f_{k,j} 1(y_{k,j} \leq p)}{\Pr(U_{k,j} = a \mid x_k^*)}$, where $1()$ is the indicator function, $f_{k,j}$ the worker's expansion factor, x^* is the vector x without the quintile information (i.e., $x = \{x^*, q\}$), and n_j is the total labor force population in period j.

Therefore it is useful to distinguish the size of both sources of data and how much the estimation improves by adding the cross-sectional moments. Table 1 details the total number of observations available from all the sample sources over the period 1990-2012. In terms of the cross-sectional sample the ENE contains over 4 million observations of workers in the labor force over 92 quarters (1990Q1-2012Q4), which corresponds roughly to a sample size above 45,000 workers per quarterly period. Since the ESI Income Module of the ENE is only collected in the 4th quarter of each year, then there are fewer cross-sectional observations with reported labor earnings, corresponding roughly to a sample of 1 million observations of workers in the labor force. It is important to note that no ESI Income Module was collected in the year of 1994.

Table 1: Panel and Cross-sectional Labor Force Observations (1990-2012)

	Number of Cross-Sectional Observations
ENE(t)	**4,215,790**
ESI(t)	**999,291**
	Number of Panel Observations
ENE(t,t-1)	**Observations with employment data in consecutive quarters**
Q2-Q1	**516,009**
Q3-Q2	**518,284**
Q4-Q3	**526,516**
Q1-Q4	**420,578**
ENE(t,t-1)	**Observations with employment data in 2 quarters and past income**
Q2-Q1	**137,544**
Q3-Q2	**51,353**
Q4-Q3	**12,936**
Q1-Q4	**327,270**
ESI(t,t-4)	**Observations with income data in consecutive years (4th quarter)**
Q4-Q4	**163,187**

The ENE survey and its income module, the ESI, are published as cross-sectional samples of households and workers. The ENE follows a rotating sample scheme in which selected home addresses are kept for approximately 18 months, with all observations labelled with an unique home address identifier. In order to establish which home addresses incorporate the same worker in different quarters I use the criteria that the worker is the same if he/she has the same gender, education, and years of age (plus 0 to 1 additional year for matches in consecutive quarters or 0 to 2 additional years for matches between different years to allow for possible anniversaries between periods). A few workers in the same household have the same gender, education and age, therefore these are dropped from the sample to prevent incorrect matches. If there is no attrition in the non-rotating sample, then it should be possible to recover 5/6 of the total sample as panel observations between consecutive quarters and 1/3 between consecutive years (4 quarters). In practice the number of matches are lower for four reasons: i) the ENE samples were completely renewed in the 4th quarter of 1994, 2006 and 2007, therefore these 3 quarters imply no matches with previous periods; ii) the rotating sample design is designed as an operational cost-saving method and if no household in a given home unit is available to answer the survey after a few contacts, then the survey selects an equivalent home unit of similar location and characteristics to replace its place in the sample; iii) there are workers moving out of their home units between different time periods, especially in the primary industry (agriculture, fishing and mining) where workers alternate periods with their families and at living quarters close to their employment; and finally, iv) my analysis discards matches of workers who moved outside the labor force (since the model only studies transitions between employment and unemployment), which implies that only workers in the labor force in both periods are counted in the sample and some classes of seasonal or temporary workers with weak labor force attachment are dropped from the sample (for example, domestic workers may decide to work in one quarter and not in the following quarter).

After accounting for sample attrition one obtains around 2 million observations of workers observed in different quarters. The number of observations is more or less similar in all consecutive quarters, except for the transition between the 4th quarter and the 1st quarter of the following year, since this is the time when major sample changes of the ENE are implemented. However, the income information is collected only in the 4th quarter of each year. Therefore if one wants to study the employment transitions of workers conditional on their income, then it is necessary to use the panel matches of workers who had an income measured in the previous 4th quarter. Obviously, this implies there are fewer matches for periods with a larger time lag from the 4th quarter. Overall, this gives more than 500,000 observations of workers with employment transitions and observable past income. Finally, since income is measured only in the 4th quarter, then the panel sample with income measured in consecutive years includes only 163,000 workers.

The income quintile of each worker is a special variable, since it can be affected by past income shocks and employment transitions. Other variables such as age, education, and industry can plausibly be assumed to be predetermined. For this reason, I estimate the employment transition parameters, $\lambda_{t|x} \equiv \left\{\lambda^{UE}_{t|x}, \lambda^{EU}_{t|x}, u_{t|x}\right\}$, for the five worker income quintile in two-steps. Let us separate the vector of worker characteristics into a vector of predetermined characteristics x^* and its income quintile q, $x = \{x^*, q\}$. First, I obtain the employment transition parameters conditional on x^*, $\lambda_{t|x^*}$, using the non-parametric estimation of panel and cross-sectional moments described in the previous section. Since in some quarters there are few observations to estimate $\lambda_{t|x}$ with precision, I only estimate the employment transition parameters for x with the observations of all periods, $\lambda_{|x}$, instead of different transition rates for each quarter t. In a second-step I obtain the employment transitions for the worker type x, $\lambda_{t|x_i} = \lambda_{t|x^*_i} \exp(\theta_{x^*} q_i)$, as a function of a time-invariant parameter specific for each income quintile θ_{x^*} plus the quarterly employment transitions for the workers of characteristics x^*, $\lambda_{t|x^*}$. This parameter is obtained by minimizing for each worker type, x, the square-difference between the mean transition rates observed for all quarters and the transition rates of each quarter, $\theta_{x^*} = \arg\min_\theta (\lambda_{|x} - \frac{1}{T}\sum_{t=1}^{T} \exp(\theta q)\lambda_{t|x^*})^2$.

5 Results

5.1 The heterogeneity of income shocks and unemployment risk

All the parameters of the income and employment dynamics model are estimated for each different quarter (t) and for each separate worker type given by the multivariate vector $x_i = \{$Metropolitan Area, Industrial Activity, Gender, Age, Education, and Worker Income quintile$\}$. Since x_i contains all the possible interactions of its categorical variables, this results in 540 worker types with time-varying parameters for 92 quarters (in the case of the employment transition dynamics) and for 23 years (in the case of the income volatility parameters). Table 2 summarizes the proportions of the Chilean labor force (i.e., both employed and unemployed workers) in terms of sex, age, education and industry, at a few selected years between 1990 to 2010. This confirms that Chile has experienced significant demographic changes over the last 23 years. Basically, the proportion of Female, workers above 55, College educated, and Services workers, has increased substantially, especially since 2000, while younger workers and the primary sector have declined in importance.

Table 2: Demographics of the Chilean workers over time (% of the total labor force in each year)

Variable / Year	1990	1995	2000	2005	2010
Male	66.6%	66.7%	66.1%	63.3%	60.8%
Female	33.4%	33.3%	33.9%	36.7%	39.2%
age: <35 years	48.8%	45.6%	40.5%	36.8%	34.9%
36-54 years	38.7%	41.7%	46.1%	47.8%	47.9%
over 55 years	12.5%	12.7%	13.4%	15.3%	17.2%
Basic education	80.1%	78.6%	78.2%	76.1%	73.0%
Secondary or Technical	6.7%	8.4%	7.4%	8.4%	9.1%
College	13.2%	13.0%	14.4%	15.5%	17.9%
Primary sector	19.3%	17.6%	15.2%	13.9%	13.0%
Secondary sector	24.8%	25.1%	23.7%	22.8%	21.6%
Services sector	55.8%	57.4%	61.1%	63.4%	65.4%

Now I summarize the heterogeneity of the employment and income parameters in Chile over the complete 1990-2012 period. Table 3 shows the mean plus the percentiles 10, 25, 50, 75 and 90 of the estimates of each parameter. All means and percentiles are calculated applying the expansion factors (or population weights of each group) for all years. All the other tables and figures in this work are also computed with the appropriate population weights of the survey samples.

Unemployment and job transition flows $(u_{t|x}, \lambda_{t|x}^{EU}, \lambda_{t|x}^{UE})$ for Chile have been estimated before in several empirical papers (see Bravo, Ferrada and Landerretche, 2005, and Jones and Naudon, 2009, for a review of previous literature on Chilean job flows). The methodology in this paper for measuring job flows is similar to the one described in Bravo, Ferrada and Landerretche (2005) and Jones and Naudon (2009) and the related literature and the underlying datasource (the ENE dataset) is also the same, therefore the mean statistics for $(u_{t|x}, \lambda_{t|x}^{EU}, \lambda_{t|x}^{UE})$ in Table 3 are similar to those studies. However, there is a big difference in the measurement of job flows in this paper relative to the previous literature. Basically, the major difference is that I match the workers' income, which is reported only once every 4 quarters, to estimate employment transitions conditional on income (or quintile of income, more specifically) in addition to other characteristics. The previous papers on employment flows in Chile do not include income among the variables of interest, since income is only observed every year and not every quarter. Income dynamics and heterogeneous shocks for workers of different income levels will be the major point of interest in the following analysis.

The unemployment probability ($u_{t|x}$) for the population over the complete period has a mean of 6.2%, but its range goes from as low as 0.6% (the percentile 10) to as high as 14.4% (percentile 90). The unemployment rate is slightly lower than the average official unemployment rate for this period, since the employment transitions model does not consider workers moving from out of the labor force into unemployment. The quarterly separation rate or entry rate into unemployment ($\lambda^{EU}_{t|x}$) has an average of 3.4% over the whole population, but it ranges from as low as 0.4% (percentile 10) to as high as 7.3% (percentile 90). However, Chile also has a large degree of job creation with the mean quarterly job-finding rate ($\lambda^{UE}_{t|x}$) being slightly higher than 50%. In the list of 14 OECD countries studied by Elsby, Hobijn and Sahin (2013), there were 10 countries that had both inflow ($\lambda^{EU}_{t|x}$) and outflow ($\lambda^{UE}_{t|x}$) rates from unemployment of less than half of the Chilean values estimated in Table 3. Also, in their study only the United States had higher inflow and outflow rates from unemployment than Chile. Therefore Chile has a fluid labor market, with a substantial amount of both job creation and destruction between quarters.

Table 3: Estimates of the parameters across the population and over time

Variable	$u_{t\|x}$	$\lambda^{EU}_{t\|x}$	$\lambda^{UE}_{t\|x}$	$R_{t\|x}$	$\sigma_{t\|x}$	$g_{t\|x}$
	$x =$ {t, Area, Industry, Education, Age, Sex, Income-Quintile}. 540 worker types.					
Mean	.062	.034	.503	.278	.183	.054
Percentile 10	.006	.004	.224	.077	.045	.023
Percentile 25	.019	.011	.341	.153	.069	.027
Percentile 50	.044	.024	.483	.322	.177	.037
Percentile 75	.089	.048	.657	.397	.271	.073
Percentile 90	.144	.073	.819	.417	.341	.124

The heterogeneous estimates of the Replacement Ratio of Income during unemployment ($R_{t|x}$) have a population mean of 27.8% and a median of 32.2%, and show a range from as low as 7.7% (percentile 10 in the population) to as high as 41.7% (the percentile 90). These estimates are reasonable, since Chile has a complex unemployment insurance scheme that has both an individual account which depends on the worker's past contributions and a solidarity payment proportional to its past income (Superintendency of Pensions of Chile, 2010). The Chilean Superintendency of Pensions estimates that the maximum value of unemployment insurance in the first quarter of unemployment should represent around 43% of the worker's previous income, which is close to the 90th percentile reported in Table 3. Also, the Superintendency of Pensions of Chile (2010) reports an average value of the unemployment replacement ratio across workers with formal labor contracts of 30%, which is close to the 27.8% and 32.2% reported in Table 3 respectively as the mean and median values in the population. Finally, around 30% of the Chilean workers engage in temporary or informal contracts, which implies a substantial portion of the labor force has a low replacement income during unemployment (Superintendency of Pensions of Chile, 2010).

Table 3 also shows the wage shocks ($\sigma_{t|x}$) experienced by workers have a mean standard deviation of 18.3% in the population, but it ranges from as low as 4.5% (percentile 10) to as high as 34.1%. These estimates show that Chilean workers face substantial labor earnings risk from year to year even if they are not experiencing unemployment. In contrast for the United States, Low, Meghir and Pistaferri (2010) estimated a permanent income shock with a standard deviation of 0.10 plus a temporary income shock with a standard deviation of 0.09. Therefore the standard deviation of the wage shock estimated with Chilean data is comparable to the ones estimated for the USA.

Finally, Table 3 reports the annual growth rate of log-income (in real terms) across different types of workers in the Chilean population and different years from the 1990-2012 period. Real log-income growth ($g_{t|x}$) has a range from as low as 2.3% (percentile 10) to a median value of 3.7% and a high value of 12.4% (percentile 90). The average real income growth rate in the population is 5.4%, but note that this value is weighted across the population of workers and does not weigh their income, therefore it differs substantially from the total income growth rate in the economy. If some workers of low income experience high income growth rates then it can affect the population mean substantially (since low income and high income count in the same way), while having little impact on aggregate labor income or Chilean GDP which weighs workers by their income value.

Table 4 shows how unemployment risk and income risk are correlated. One important question is if the workers suffering more from unemployment risk are also the ones suffering more from other risks such as wage volatility. The search theory of unemployment predicts that workers and employers terminate a job relationship if the benefits of the job position fall below a certain cut-off or equivalently if the labor productivity falls below a reservation wage (see for instance, Shimer, 2012). If the search theory is valid, then one would expect that unemployment ($u_{t|x}$) and separation rates ($\lambda_{t|x}^{EU}$) to be positively correlated to wage volatility ($\sigma_{t|x}$). This is exactly what is observed in the Chilean data, with wage volatility having a correlation above 20% with separation and unemployment rates. Another relevant question is what causes some worker types to experience high unemployment rates, since unemployment can be caused either by lots of job separations (i.e., a high $\lambda_{t|x}^{EU}$) or long unemployment spells (i.e., a low job finding rate $\lambda_{t|x}^{UE}$). Separation and job-finding rates are positively correlated, which shows that workers more subject to job destruction also find new employment easily. Also, one observes that the workers with highest risk of unemployment and job separation are the ones with the lowest income replacement ratio during unemployment ($R_{t|x}$). This finding confirms that the unemployment insurance scheme in Chile does not reward moral hazard, since workers facing unemployment do not really enjoy high benefits.

Table 4: Correlation of employment and income risk measures

Correlation	$u_{t\|x}$	$\lambda_{t\|x}^{EU}$	$\lambda_{t\|x}^{UE}$	$R_{t\|x}$	$\sigma_{t\|x}$	$g_{t\|x}$
$\mathbf{u}_{t\|x}$	100%					
$\mathbf{separation}_{t\|x}$	61.8%	100%				
$\mathbf{jobfind}_{t\|x}$	-8.9%	28.8%	100%			
$\mathbf{R}_{t\|x}$	-21.0%	-19.0%	-8.5%	100%		
$\sigma_{t\|x}^2$	22.3%	20.7%	7.1%	-41.6%	100%	
$g_{t\|x}$	-2.2%	1.5%	5.9%	3.3%	-4.1%	100%

Finally, I show how much of the wage dispersion among workers ($Var\left[y_{i,t+1} \mid U_{i,t+1} = U_{i,t}, x_i, t\right]$) is represented by wage volatility ($\sigma_{t|x}^2$). A low value of $\frac{\sigma_{t|x}^2}{Var[y_{i,t}|U_{i,t}=U_{i,t-1},x_i,t]}$ denotes a labor market where workers start with heterogeneous productivities and the volatility of the income paths does not represent much in the overall lifetime heterogeneity. Table 5 shows how much of the unobserved wage dispersion across different types of workers is explained by annual wage volatility and shows this dispersion across different worker types in the population. Wage volatility represents less than 20% of the overall wage dispersion faced by more than 50% of the workers (i.e., the median value in the population). In fact, even for the workers facing highest wage volatility (i.e., the percentile 90), their wage shocks explain only 30% of the overall wage dispersion. Therefore Chilean workers have significant levels of both initially unobserved wage heterogeneity as well as different wage shocks.

Table 5: Share of new shock $\sigma^2_{t+1|x_i}$ in overall wage dispersion, $Var\left[y_{i,t+1} \mid U_{i,t+1} = U_{i,t}, x_i, t\right]$

Percentile	$\dfrac{\sigma^2_{t+1\|x_i} = Var\left[y_{i,t+1} \mid y_{i,t}, U_{i,t+1} = U_{i,t}, x_i, t\right]}{Var\left[y_{i,t+1} \mid U_{i,t+1} = U_{i,t}, x_i, t\right]}$
10	0.077
25	0.114
50	0.182
75	0.262
90	0.304
Mean	0.191

5.2 Computing the reliability of the estimates

With a large sample of worker observations (Table 1) it is expected that the model estimates should be fairly precise. Bootstrap standard errors were obtained for all the 540 worker types and all time periods from 100 sample replicas. Table 6 shows the distribution of standard errors of each parameter for all time-periods and worker types. Note that since each parameter has heterogeneity over time and over different types of workers, then there is the possibility that it has a low standard error for some types of workers while at the same time being imprecise for other types of workers. In addition I show how these standard errors decrease if the estimation considers less heterogeneity by using only 90, 30 and 5 types of workers in each time period.

It is worth noting, however, that these standard errors only account for sampling error (i.e., the estimation error caused by having a finite sample of data) and not for measurement error due to other sources such as respondents mis-reporting their answers (either deliberately by lowering their income or unintentionally by misinterpreting the survey question). This mis-reporting error is likely to be uncorrelated across observations at a given period. However, survey questionnaires change over time and different questions may have different kinds of mis-reports, which can cause some noise over different time periods that does not disappear even with large numbers of observations.

Looking at the standard errors estimated for the 540 worker types in all time periods it is easy to conclude that the median or the mean standard error for all parameters are quite low. However, there is a lot of heterogeneity in the standard errors for different types and for the percentile 90 of most uncertain parameters the standard errors are sizeable. Therefore there is a lot of noise if one analyzes some particular worker type with a small number of observations in each time period. However, Table 6 also shows that aggregating across different worker types is helpful and the standard errors for the most uncertain parameters (those in the percentile 90 or above) fall by about a half if one considers only 90 worker types. Therefore an analysis with only 90 worker types should be fairly accurate for even the worker types with the smallest number of observations. The analysis also shows that if one considers only 5 worker types (i.e., workers sorted across their income quintile), then the standard errors are negligible for all parameters.

Table 6: Standard error estimates of the parameters (100 bootstrap replicas)

Variable	$u_{t\|x}$	$\lambda^{EU}_{t\|x}$	$\lambda^{UE}_{t\|x}$	$R_{t\|x}$	$\sigma_{t\|x}$	$g_{t\|x}$	$u_{t\|x}$	$\lambda^{EU}_{t\|x}$	$\lambda^{UE}_{t\|x}$	$R_{t\|x}$	$\sigma_{t\|x}$	$g_{t\|x}$
	$x = \{$t, Area, Industry, Education, Age, Sex, Income-Quintile$\}$. 540 worker types.						$x = \{$t, Education, Age, Sex, Income-Quintile$\}$. 90 worker types.					
Mean	.018	.009	.046	.065	.014	.0026	.010	.005	.026	.036	.009	.0015
Percentile 10	.003	.0005	.008	.019	.002	.0002	.004	.001	.011	.013	.004	.0002
Percentile 25	.007	.003	.020	.029	.007	.0004	.005	.002	.017	.018	.006	.0002
Percentile 50	.014	.006	.039	.048	.012	.0008	.008	.004	.023	.033	.008	.0005
Percentile 75	.023	.012	.062	.090	.019	.0022	.012	.006	.032	.048	.012	.0013
Percentile 90	.037	.019	.091	.141	.029	.0056	.018	.009	.043	.067	.016	.0031
	$x = \{$t, Age, Sex, Income-Quintile$\}$. 30 worker types.						$x = \{$t, Income-Quintile$\}$ 5 worker types.					
Mean	.008	.004	.023	.030	.008	.0011	.004	.002	.012	.018	.005	.0006
Percentile 10	.004	.001	.011	.011	.004	.0002	.002	.001	.007	.006	.003	.0002
Percentile 25	.005	.002	.015	.016	.005	.0002	.003	.002	.009	.009	.003	.0002
Percentile 50	.007	.003	.021	.022	.007	.0004	.004	.002	.011	.012	.004	.0003
Percentile 75	.010	.005	.027	.043	.010	.0009	.005	.003	.014	.027	.005	.0006
Percentile 90	.014	.007	.036	.057	.013	.0027	.006	.003	.018	.034	.009	.0015

Another kind of uncertainty is the use of more moments than strictly necessary for estimating the earnings dynamics model. As explained in the previous section, the 6 parameters of the earnings and employment dynamics can be estimated purely from the panel data sample moments. However, I also use additional cross-sectional moments of earnings which give information on income growth, the earnings loss of unemployed relative to similar employed workers, and even the volatility of wage shocks (which can be inferred from changes in the wage dispersion between years). These additional cross-sectional moments add information which can reduce the uncertainty of estimates, but can also cause unintended sources of bias in case the assumptions of the model are incorrect and therefore the additional moments provide invalid information. Table 7 presents the correlation of the parameter estimates that use only the necessary 6 panel data moments and the estimates using the complete set of 12 data moments (6 panel data and 6 cross-sectional data moments) across the 540 different worker types and all available time periods (92 quarters or 23 years). All parameters show a high correlation with the parameters estimated purely from the panel data sample. For the case of the unemployment rate ($u_{t|x}$), the replacement ratio of income ($R_{t|x}$) and the earnings growth rate of each worker type ($g_{t|x}$), the estimates using the additional parameters are almost the same as the ones obtained purely from the panel data. For the employment transitions and wage volatility there are more changes after one uses the additional cross-sectional parameters to reduce estimation uncertainty, but the correlation of the estimates across time periods and different workers is still above 60%. Therefore the data analysis appears to be robust to whether one uses a small set of identifying moments versus a larger set of moments.

Table 7: Correlation of estimates using 6 and 12 moments

$u_{t\|x_i}$	$\lambda^{EU}_{t\|x_i}$	$\lambda^{UE}_{t\|x_i}$	$R_{t\|x_i}$	$\sigma_{t\|x_i}$	$g_{t\|x_i}$
85.2%	**77.7%**	**60.6%**	**96.3%**	**62.2%**	**97.1%**

5.3 Employment transitions in the population and over time

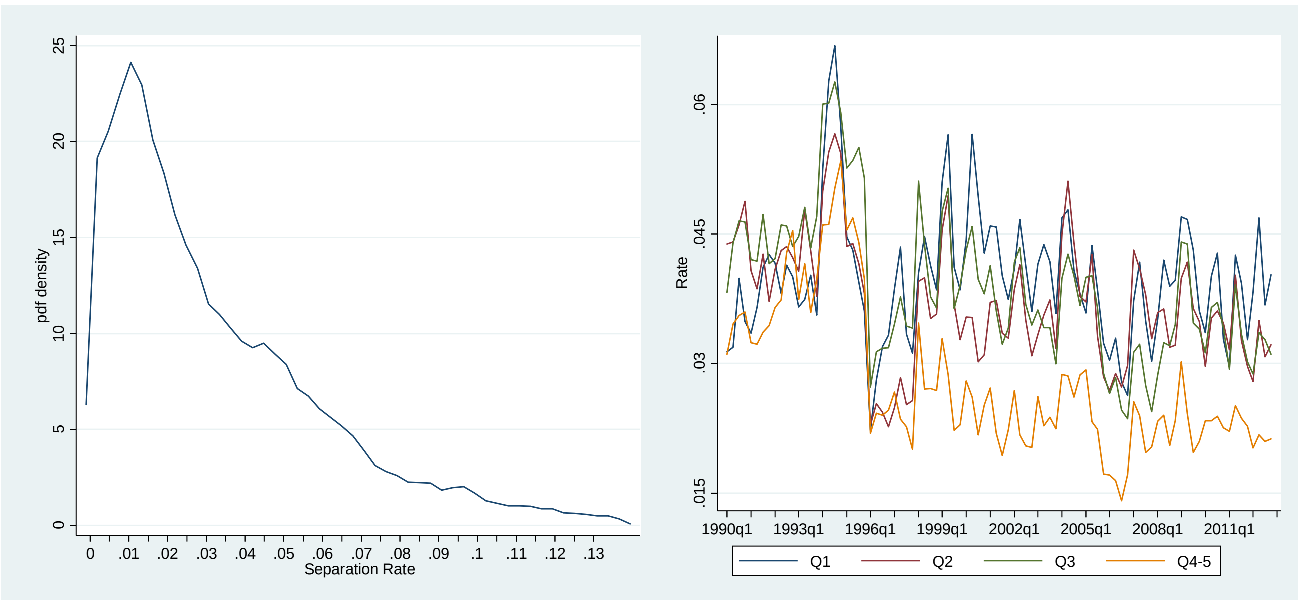
This section summarizes the empirical evidence on the heterogeneity and time changes for each parameter of the dynamic earnings model in two graphs: i) the probability density function of the different values of each parameter across different worker types (x) and time (t), and ii) how the mean parameters of workers of different income quintiles have changed over the last 23 years, with the first quintile, Q1, corresponding to the workers with the 20% lowest income and Q5 denoting the highest income level. For simplicity the two top quintiles are shown together, with Q4-Q5 denoting their mean. The probability density functions for each variable are estimated non-parametrically using a kernel estimator with data on all groups and time periods, $\hat{p}(x) = \dfrac{\sum_{t=1}^{T}\sum_{i=1}^{N_t} f_{i,t} K(\dfrac{x_{i,t}-x}{h})}{h\sum_{t=1}^{T}\sum_{i=1}^{N_t} f_{i,t}}$, where $K()$ is chosen to be the Epanechnikov function and the bandwidth $h = \frac{0.9IQR(x_{i,t})}{(TN_t)^{0.2}}$. This choice is asymptotically consistent and minimizes the sample mean square error (Pagan and Ullah, 1999). All density functions account for the population weights of the observations, $f_{i,t}$.

Figures 1, 2 and 3 show the heterogeneity and quarterly time series for the unemployment ($u_{t|x}$), separation ($\lambda_{t|x}^{EU}$) and the job finding ($\lambda_{t|x}^{UE}$) rates, respectively. The population mode of the unemployment and separation rates is around 1% to 2%, which can be explained by a significant number of workers being protected by job contracts. However, these parameters have a long tail and there are a significant proportion of workers with unemployment rates between 2% and 10% and separation rates between 2% and 6%. Figure 1 shows that unemployment rates were low for all income quintiles until Chile started being affected by the Asian crisis in 1998.

Figure 1: Probability density function of the Unemployment Rates in the population and Quarterly time series of the Unemployment rate by Income Quintile (1990-2012)

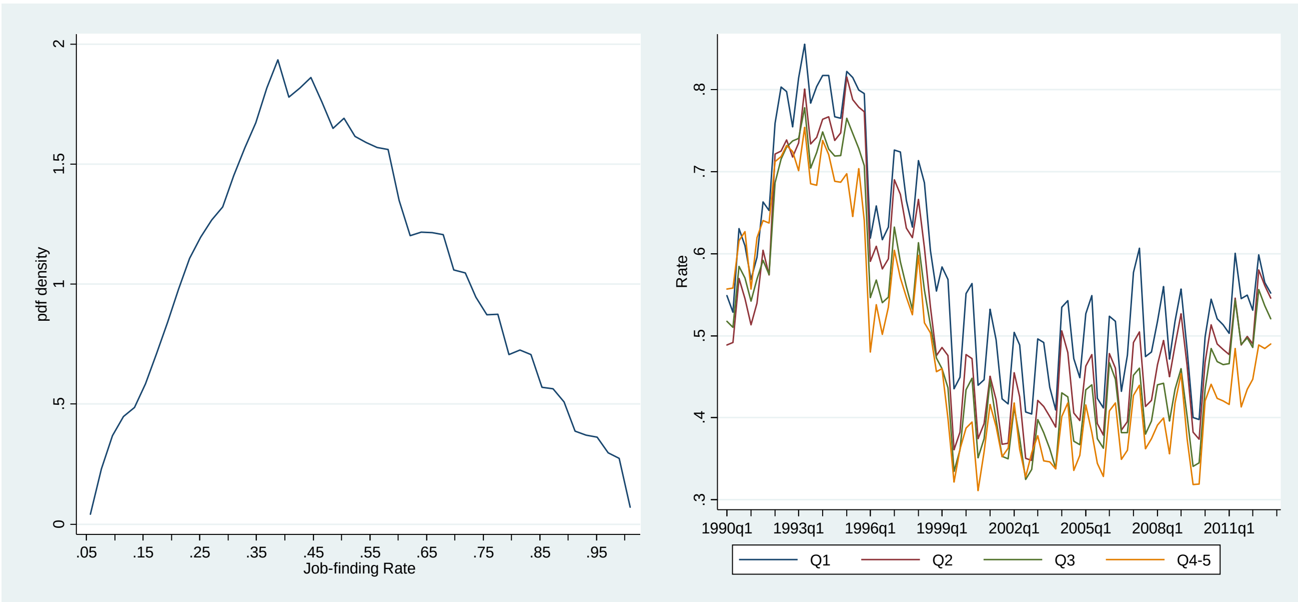
Unemployment rates remained high until 2003 when the economic expansion lowered unemployment for all groups. After a brief increase in unemployment during the international financial crisis in 2008, the recent years saw unemployment drop for all groups except the lowest income quintile. Separation rates (Figure 2) show a less clear cyclical pattern, although there are strong increases in 1995, 1998 to 2001, and again in 2008 and 2009. There is a clear monotonicity between workers' income and unemployment risk, with the lowest income quintiles experiencing higher unemployment and separation rates all over the business cycle. However, there is also a strong common component to booms and downturns which affects all groups.

Figure 2: Probability density function of the Separation Rates in the population and Quarterly time series of the Separation Rate by Income Quintile (1990-2012)



The job-finding rate has a population mode around 35% to 45%, but a significant number of workers have job-finding rates well above 50% which implies their unemployment spells last only one or two quarters. Workers of low income have a higher job-finding rate and therefore face shorter unemployment spells, which can be explained if workers of low income and education have less specific human capital and can be more easily matched to employers. In the early 90s the Chilean labor market was even more fluid than now and job-finding rates reached as high as 70% to 80% for all income groups up until 1996. However, job-finding rates declined a lot with the Asian crisis in 1998 and 1999 and never recovered their previous levels. Therefore Chile suffered not just a cyclical downturn in 1998, but also a structural change in the labor market with both job creation and destruction becoming less fluid than in the early 90s. The economic expansion in Chile after 2010 was expressed in terms of higher job finding rates, while the separation rates remained stable. Employment transitions also have significant seasonal changes between the quarters of a given year.

Figure 3: Probability density function of the Job-Finding Rates in the population and Quarterly time series of the Job-Finding Rate by Income Quintile (1990-2012)



5.4 The time varying Income Shocks

Figures 4, 5 and 6 summarize the empirical evidence of heterogeneity and the yearly time series for the mean log-income growth ($g_{t|x}$), the standard deviation of wage shocks ($\sigma^2_{t|x}$), and the replacement ratio of income during unemployment ($R_{t|x}$), respectively. Figure 4 shows that the modal growth rate of real log-income is positive but close to 0%, which is to be expected for many occupations and during periods of slow growth. However, a significant proportion of workers benefit from yearly income growth between 2% and 10%. The proportion of workers actually suffering real income drops or with income growth above 15% is clearly negligible. The workers in quintiles 2 to 5 had a steady income growth over the last 23 years, but the lowest income quintile has a more volatile income growth rate with bigger expansions and sharp downturns, perhaps because several workers of low income have temporary and more flexible labor contracts. The time series estimates show negative real wage growth during the recessions in 1998 and again in 2008. However, the series is noisy and there are periods of negative wage growth during economic expansions, perhaps because the survey occasionally changes its questionnaire and also its sample coverage over time. Another possible explanation for why real wages could drop in some expansionary periods could be if workers of low productivity are more likely to join the labor force and therefore reduce the overall wages (even if the wages for the more attached labor force members are actually growing).

Figure 4: Probability density function of the Mean-Log Income Growth Rates in the population and Yearly time series of the Real Log-Income Growth by Income Quintile (1990-2012)

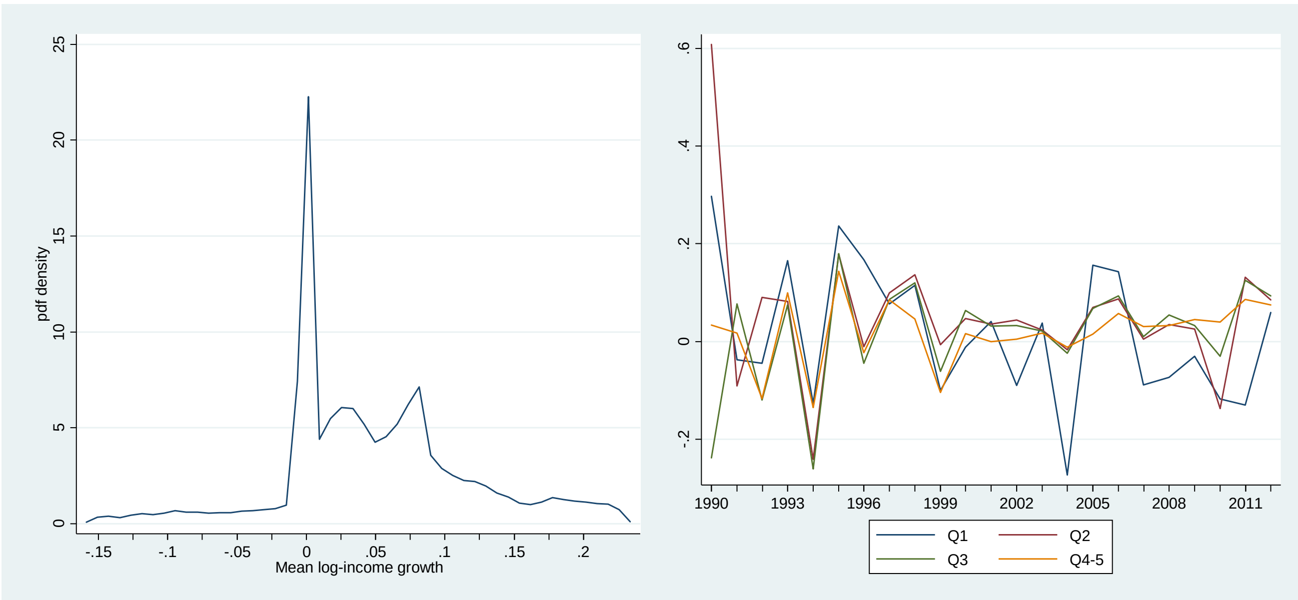
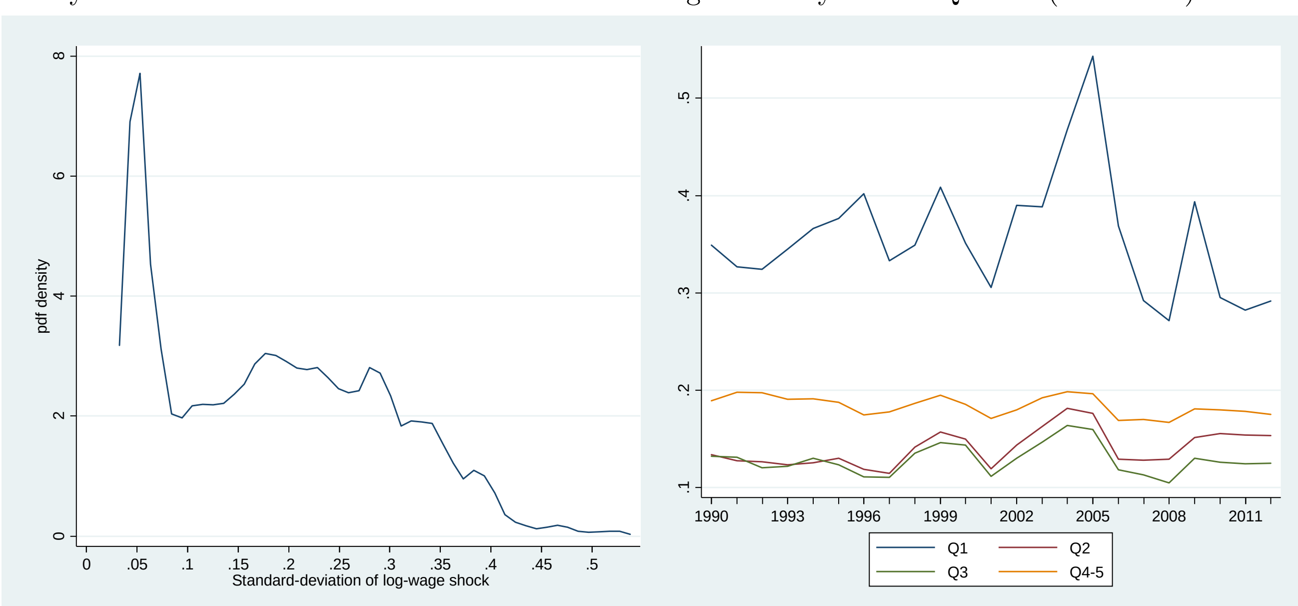


Figure 5: Probability density function of the Standard-deviation of Wage Shocks in the population and Yearly Time series of the Standard-deviation of the Wage Shock by Income Quintile (1990-2012)

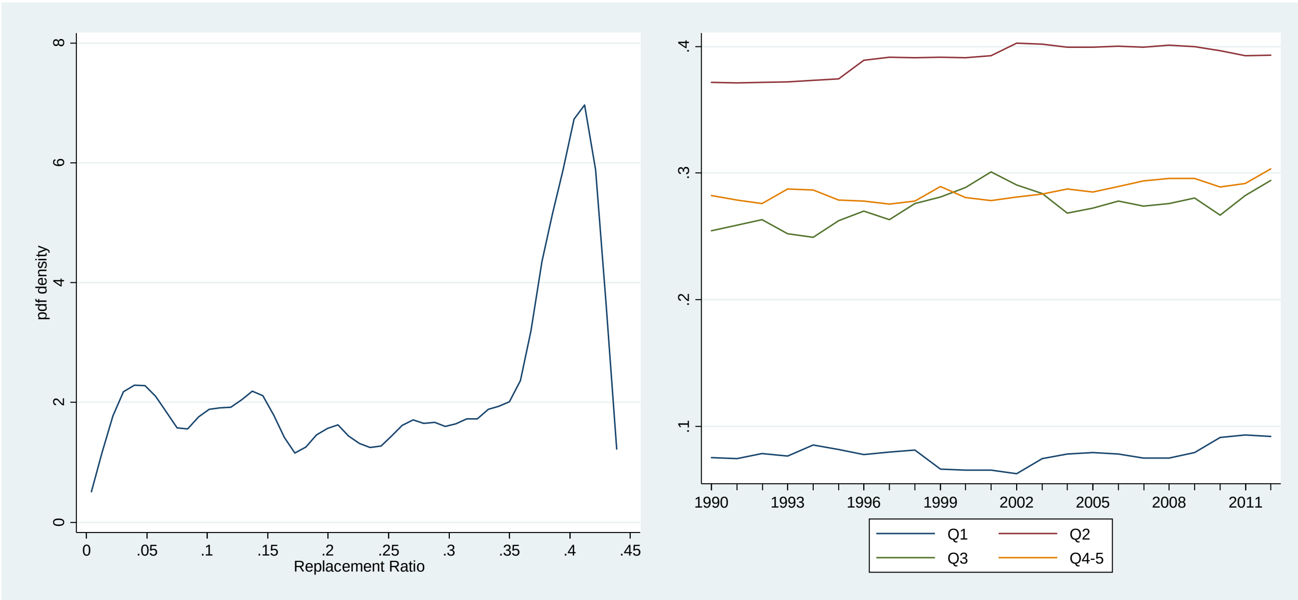


The population mode for the standard deviation of wage shocks is only 5%, which shows that a significant portion of workers may have contracts with fixed wages. However, a significant percentage of workers have an annual wage volatility of 10% to 35%. Curiously, it is the middle class of workers - the income quintiles 2 and 3 - that have the lowest annual wage volatility and this happens in all periods over the last 23 years. The workers in the top income quintile have a bit more wage volatility than the middle class, but their wage volatility is below 20% in most years, while the lowest income quintile has an annual wage volatility around 30% to 40%. Estimates of the standard deviation of wage shocks for other countries are around 30% to 32% for the United States, 27% to 34% for Germany, and 22% for Spain (see Table 7A in Krueger, Perri, Pistaferri and Violante, 2010). Therefore, except for the lowest income quintile, Chilean workers seem to have lower levels of wage volatility relative to their counterparts in developed economies.

Figure 5 also shows that the standard deviation of wage shocks increased during the recessions of 1998 and 2009 for all groups, especially for the lowest income quintile. However, it is not clear that wage volatility is countercyclical. The years of 2004 and 2005 saw high real GDP growth rates in Chile and yet this economic expansion coincided with the period of highest wage volatility for the workers of all quintiles (Figure 5). Perhaps wage volatility is associated with both recessions and expansions, since these are periods where the quality of the job matches change and firms could implement more restructuring. It is worth noting that international research on earnings dynamics has not studied how idiosyncratic income volatility fluctuates over the business cycle (Krueger, Perri, Pistaferri and Violante, 2010), therefore it would be interesting to know if the higher wage volatility during both economic recessions and expansions is also observed in other countries.

Most workers have a replacement ratio of income during unemployment around 30% to 40%. However, the replacement ratio is quite different across income quintiles, with workers in the lowest income quintile having a replacement ratio below 10% while workers in the quintile 2 have replacement ratios close to 40% and the top 3 to 5 income quintiles have replacement ratios around 30%. Therefore the workers in the lowest income quintile have both higher unemployment rates, higher wage risk and lower replacement ratios of income during unemployment. However, one can observe that the replacement ratio for the lowest income quintile increased after 2002, which could be caused by the legislation implemented that year with the goal of enlarging the coverage of unemployment insurance and increasing its benefits (Superintendency of Pensions of Chile, 2010).

Figure 6: Probability density function of the Replacement Ratio of Income in the population and Yearly time series of the Replacement Ratio of Income by Income Quintile (1990-2012)



5.5 Log-Income Variance Decomposition

Finally, I report the graphical evidence on the heterogeneity and time changes of the importance of employment transitions in the overall income variance. Figures 7 and 8 summarize the empirical evidence of heterogeneity and time changes for the standard deviation of total income ($Var(y_{i,t+1} - y_{i,t} - g_{t+1|x})$) and the share of employment transitions in total income shocks ($\theta_{t+1|x}$), respectively. The overall standard deviation of earnings - that is, a measure who counts both wage shocks and earnings fluctuations caused by transitions into and out of unemployment - is around 15% to 45% for most workers. Workers in the income quintiles 2 to 5 have a standard deviation of earnings around 25% to 40%. However, the lowest income quintile suffers fluctuations in income far above the rest of the population with a standard deviation of log-earnings above 85% in most years.

Employment transitions represent 45% to 95% of the earnings variance for most workers, although for some worker types this share can be as low as 15%. For middle class workers (quintiles 2 and 3) unemployment transitions represent 75% to 85% of the wage variance, while in the lowest and highest income quintiles this share is around 45% to 65%. The time series of the share of the income variance explained by employment transitions is counter-cyclical and it increased for workers of all income groups during the downturns of 1998 to 2003 and again from 2008 to 2009.

Figure 7: Probability density function of Standard-deviation of Total Income in the population and Yearly Time series of Standard-deviation of Total Income by Income Quintile (1990-2012)

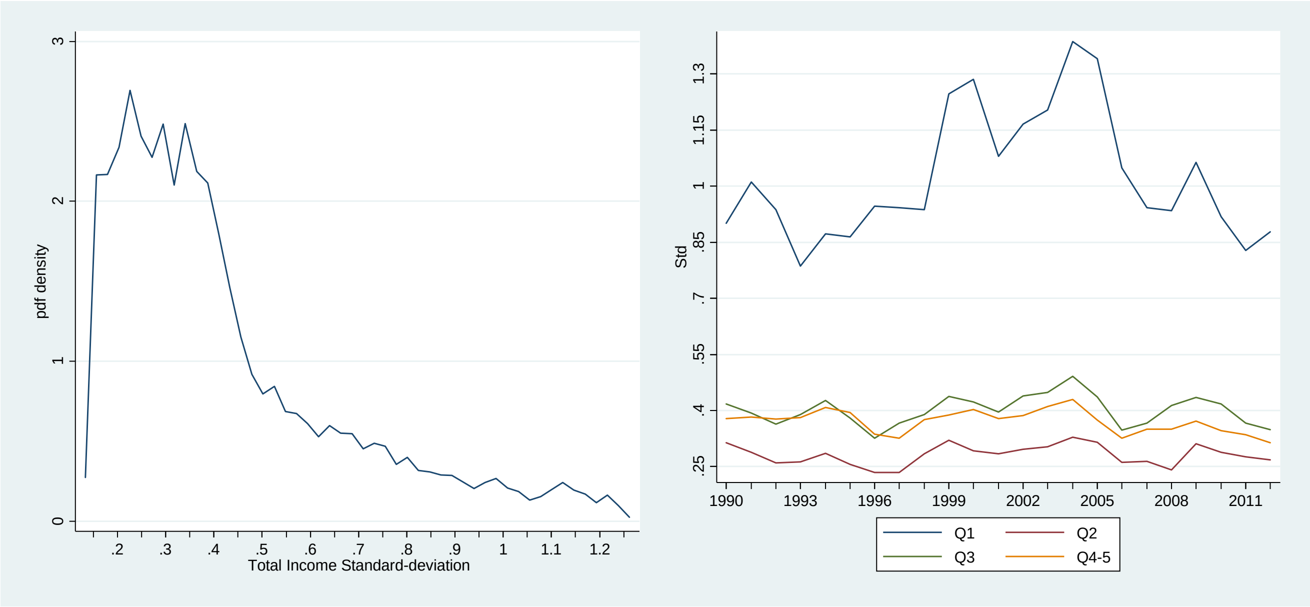
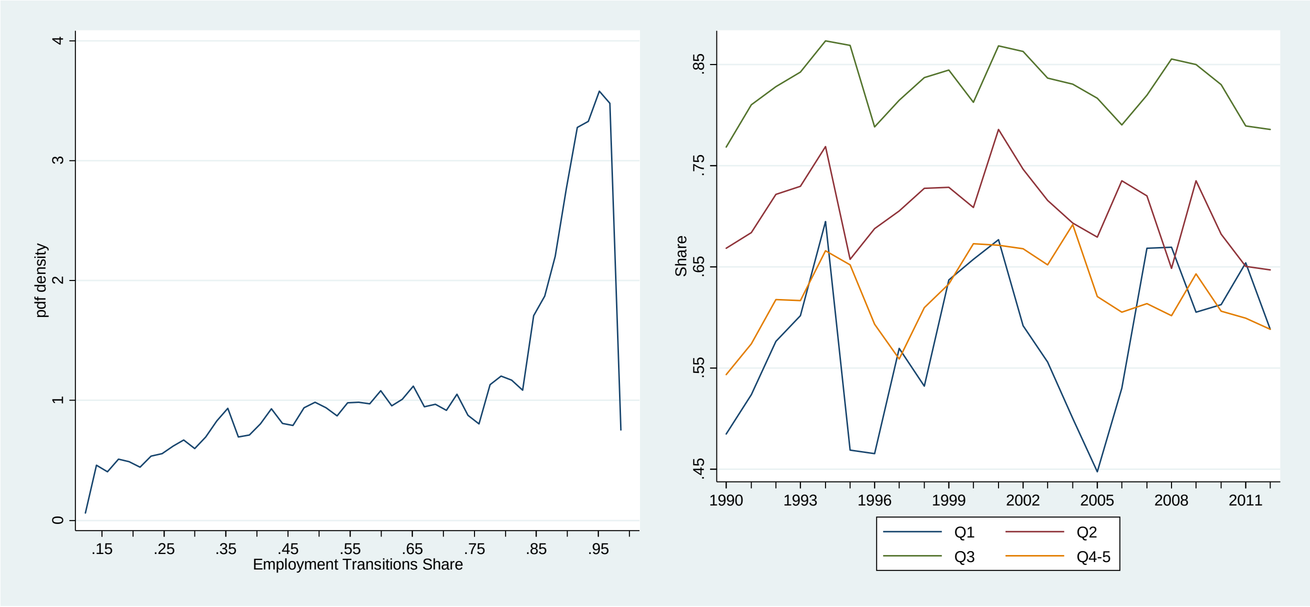


Figure 8: Probability density function of Share of Employment Transitions in Total Income Shocks in the population and Yearly Time series by Income Quintile (1990-2012)



6 Conclusions

The research agenda in terms of how life-cycle income and income shocks affect economic decisions such as consumption, saving and retirement is quite large (Attanasio and Weber, 2010), but it is still mostly limited to developed economies who benefit of panel datasets of households and working individuals. This paper shows it is possible to estimate a large range of statistics on income dynamics and volatility by using matched observations from cross-sectional employment and income surveys which have rotating samples. These surveys are available in many countries and can be used to build effective panel datasets with a large number of individuals (large N), but a short time sample (small T). Therefore research on income dynamics in developing countries can stretch the data analysis a lot forward and I present one such example with Chilean data.

After building a series of short T panels with income and employment information in Chile at a quarterly frequency for the years 1990 to 2012 (92 quarters), I estimate a rich income dynamics model that considers three sources of fluctuations: i) the quarterly transitions into and out of unemployment, ii) the severity of the income loss during the unemployment spell, and iii) a continuous wage shock that is experienced by employed workers even if they do not change their employment status. I then show that the model's parameters can be estimated from the available sample with a strong degree of statistical precision even if one considers a large degree of workers' heterogeneity. In particular, one can separate the sample into different worker types based on their characteristics of sex, age, education, industry and income quintile, and still get reliable estimates for the time varying model parameters for a large proportion of the worker types. For simplicity the model ignores hours worked, but further research could study how much of the earnings volatility is due to fluctuations in hours and how flexible hours are over the business cycle.

Job separations across worker types are highly correlated with wage volatility, which is evidence that a substantial proportion of separations are efficient. The idiosyncratic wage volatility of employed workers and the income variance due to flows into and out of unemployment are both high, but the share of volatility caused by unemployment increases much more during recessions. Evaluations of the welfare costs of recessions (De Santis, 2007) must therefore account for worker heterogeneity, since it matters who suffers the increases in unemployment. Further research should study if Chilean households can adequately use the financial system to smooth earnings shocks.

References

[1] Abowd, John and David Card (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411-445.

[2] Abowd, John and Arnold Zellner (1985), "Estimating Gross Labor-Force Flows," *Journal of Business and Economic Statistics*, 3(3), 254-283.

[3] Altonji, Joseph and Lewis Segal (1996), "Small-sample bias in GMM estimation of covariance structures," 14(3), 353-366.

[4] Attanasio, Orazio and Guglielmo Weber (2010), "Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy," *Journal of Economic Literature*, 48(3), 693-751.

[5] Bravo, David, Christian Ferrada and Oscar Landerretche (2005), "The labor market and economic cycles in Chile," mimeo, Universidad de Chile.

[6] Browning, Martin, Mette Ejrnaes and Javier Alvarez (2010), "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77(4), 1353–1381.

[7] Carroll, Christopher D. and Andrew A. Samwick (1997), "The nature of precautionary wealth," *Journal of Monetary Economics*, 40(1), 41-71.

[8] Cho, Yoonyoung, David Margolis, David Newhouse and David Robalino (2012), "Labor markets in low and middle income countries: trends and implications for social protection and labor policies," Social Protection Discussion Papers N° 1207, The World Bank.

[9] De Santis, Massimiliano (2007), "Individual consumption risk and the welfare cost of business cycles," *American Economic Review*, 97(4), 1488-1506.

[10] Dethier, Jean-Jacques (2007), "Social security: What can developing countries learn from developed countries?," in 2020 Focus briefs on the world's poor and hungry people, Chapter 21, International Food Policy Research Institute (IFPRI).

[11] Elsby, Michael, Bart Hobijn and Aysegul Sahin (2013), "Unemployment dynamics in the OECD," *Review of Economics and Statistics*, 95(2), 530-548.

[12] Eurostat (2012), "Labour force survey in the EU, candidate and EFTA countries: Main characteristics of national surveys, 2010", European Commission.

[13] Geweke, John and Michael Keane (2000), "An Empirical Analysis of Income Dynamics Among Men in the PSID: 1968–1989", *Journal of Econometrics*, 96, 293–356.

[14] Gourinchas, Pierre-Olivier and Jonathan A. Parker (2002), "Consumption Over the Life Cycle," *Econometrica*, 70(1), 47-89.

[15] Guvenen, Fatih (2007), "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?", *American Economic Review*, 97 (3), 687–712.

[16] Horowitz, Joel L. (2001), "The Bootstrap," in Handbook of Econometrics, edition 1, vol. 5, chapter 52, 3159-3228, Elsevier.

[17] Hubbard, R. Glenn, Jonathan Skinner and Stephen P. Zeldes (1995), "Precautionary Saving and Social Insurance", *Journal of Political Economy*, 103(2), 360-399.

[18] Huneeus, Cristóbal and Andrea Repetto (2005), "The Dynamics of Earnings in Chile," in Jorge Restrepo and Andrea Tokman R. (Editors), Labor Markets and Institutions, vol. 8, chapter 12, 383-410, Central Banking, Analysis, and Economic Policies Book Series.

[19] International Labour Office (1990), "Survey of Economically Active Population, Employment, Unemployment and Underemployment: ILO Manual on Concepts and Methods," Geneva.

[20] Jones, Ingrid and Alberto Naudon (2009), "Dinamica laboral y Evolucion del Desempleo en Chile," Economia Chilena, 12(3), 79-87.

[21] Juhn, Chinhui, Kevin M. Murphy and Robert H. Topel (1991), "Why Has the Natural Rate of Unemployment Increased over Time?," *Brookings Papers on Economic Activity*, 22(2), 75-142.

[22] Keane, Michael and Kenneth Wolpin (1997), "The career decisions of young men," *Journal of Political Economy*, 105(3), 473-522.

[23] Kish, Leslie (1998), "Space/Time Variations and Rolling Samples," *Journal of Official Statistics*, 14(1), 31-46.

[24] Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri and Giovanni L. Violante (2010), "Cross-sectional facts for macroeconomists," *Review of Economic Dynamics*, 13(1), 1–14.

[25] Low, Hamish, Costas Meghir, and Luigi Pistaferri (2010), "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, 100(4), 1432–1467.

[26] Meghir, Costas and Luigi Pistaferri (2004), "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

[27] OECD (2013a), "Labour force statistics in OECD countries", OECD, Paris.

[28] OECD (2013b), "OECD Database on the Earnings distribution", OECD, Paris.

[29] Pagan, Adrian and Aman Ullah (1999), "Nonparametric Econometrics," Cambridge University Press.

[30] Rao, J. and C. Wu (1988), "Resampling Inference With Complex Survey Data", *Journal of the American Statistical Association*, 83, 231-241.

[31] Sala-i-Martin, Xavier (2006), "The World Distribution of Income: Falling Poverty and...Convergence, Period," *Quarterly Journal of Economics*, 121(2), 351-397.

[32] Schluter, Christian (1998), "Income dynamics in the USA, Germany, and the UK - Evidence from Panel Data", London School of Economics, CASE working paper N^a 8.

[33] Shimer, Robert (2012), "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15(2), 127-148.

[34] Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron (2001), "The Welfare Cost of Business Cycles Revisited: Finite Lives and Cyclical Variation in Idiosyncratic Risk," *European Economic Review*, 45(7): 1311–39.

[35] Storesletten, Kjetil, Chris I. Telmer and Amir Yaron (2004), "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, 112(3), 695-717.

[36] Superintendency of Pensions (2010), "Unemployment insurance in Chile", Solange Berstein (ed.), Superintendency of Pensions, Chile.

[37] United Nations (2008), "Designing Household Survey Samples", Studies in Methods, Series F.

[38] UNU-WIDER (2013), "World Income Inequality Database," United Nations.

[39] Viceira, Luis (2001), "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income," *Journal of Finance*, 56(2): 433-470.

[40] Wooldridge, Jeffrey M. (2002), "Econometric analysis of cross section and panel data", MIT Press.

[41] Zhang, Sisi (2010), "Recent Trends in Household Income Dynamics for the United States, Germany and Great Britain", *Economics Bulletin*, 30(2), 1154-1172.

Documentos de Trabajo Banco Central de Chile	**Working Papers Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES
La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.	Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC – 753
The Impact of Commodity Price Shocks in a Major Producing Economy. The Case of Copper and Chile
Michael Pedersen

DTBC – 752
Nominal Term Structure and Term Premia: Evidence from Chile
Luis Ceballos, Alberto Naudon y Damián Romero

DTBC – 751
The Labor Wedge: New Facts Based on US Microdata
David Coble

DTBC – 750
El Rol de las Asimetrías en el Pass-Through: Evidencia para Chile
Lucas Bertinatto y Diego Saravia

DTBC – 749
Dissent in FOMC Meetings and the Announcement Drift
Carlos Madeira y Joao Madeira

DTBC – 748
Post-crisis Financiera y Expansión de las Exportaciones: Micro-Evidencia para Chile
Roberto Álvarez y Camila Sáez

DTBC – 747

Exchange Rate Pass-Through to Prices: VAR Evidence for Chile
Santiago Justel y Andrés Sansone

DTBC – 746

A New Liquidity Risk Measure for the Chilean Banking Sector
Sebastián Becerra, Gregory Claeys y Juan Francisco Martínez

DTBC – 745

Sensibilidad de las Exportaciones al TCR: Un Análisis Sectorial y por Destino
Samuel Carrasco, Diego Gianelli y Carolina Godoy

DTBC – 744

Agrupación de Instituciones Bancarias a Partir del Análisis de Cluster: Una Aplicación al Caso de Chile
Alejandro Jara y Daniel Oda

DTBC – 743

Bailouts and Prudential Policies-A delicate Interaction
Ernesto Pasten

DTBC – 742

Facing Volatile Capital Flows: The Role of Exchange Rate Flexibility and Foreign Assets
Rodrigo Cifuentes and Alejandro Jara

DTBC – 741

Risk Matters: The Impact of Nominal Uncertainty in Chile
Luis Ceballos y Damián Romero

DTBC – 740

Credit Stabilization through Public Banks: The Case of BancoEstado
Luis Felipe Lagos y Matías Tapia

DTBC – 739

Trade with Asymmetric Information
Robert Hall

BANCO CENTRAL
DE CHILE