VULNERABILITY TO POVERTY IN CHILE: AN APPROACH USING CROSS SECTION DATA[1].

Patricio Domínguez Rivera[2]

Goldman School of Public Policy, UC Berkeley

Pontificia Universidad Católica, Chile

ABSTRACT

This article analyzes the vulnerability structure of households in Santiago, Chile, during the 1990s. For that purpose, we use Vulnerability as Expected Poverty approach, developed by Chaudhuri (2002) that allows us to assess the vulnerability to poverty at the household level using cross sectional data. Basically, we define vulnerability as the risk or probability that a household could fall or remain in poverty in the future, considering the surrounding associated conditions. Using the Chilean National Socio-Economic Survey (CASEN, 1996) we show that there is an unequal distribution of the risk of being poor in Santiago. In a context were only 12.39% of the population was below the poverty line we find that due to a large proportion of vulnerable households, more than 36% of the households should be take into account for a social policy design that aim to reduce poverty in a sustainable way. Those results were validated with the available sample of the CASEN Survey Panel Data (CASEN) 1996-2001-2006. Finally, regarding a more suitable design of policies to overcome poverty, we propose a scheme that incorporates as the beneficiary group the total of vulnerable households. Thus, instead of considering only those households who are below the poverty line in a specific moment in time, we identify at least three relevant groups to focus on: chronic poor (10.09%), non-vulnerable or transient poor (2.3%) and vulnerable non-poor (23.79%),

[1] I am grateful to Professor Alain De Janvry for his guidance on this work. Any error or omission is the sole responsibility of the author.

[2] Patricio Domínguez is a graduate student at the Goldman School of Public Policy at the University of California, Berkeley; and he is an Associate Researcher and Lecturer at Pontificia Universidad Catolica of Chile (PUC). He holds a M.A in Sociology and a B.S in Industrial and Transportation Engineering both from PUC. Author´s email: pdominguez@berkeley.edu, pdomingr@uc.cl

1. INTRODUCTION

In the last decades an increasing number of scholars have been demanding a better understanding of poverty. This article focuses on a crucial aspect of it: its dynamics. The main point that this article aims to convey is the idea of vulnerability to poverty. Vulnerability is defined as the risk of being poor regardless of the current observed situation of a specific individual or household.

In that sense, we aim to analyse the structure of the vulnerability to poverty in Chile. We will focus on the particular region of Santiago, Chile at the end of the 1990s when this country was experiencing both strong economic growth and a dramatic reduction in poverty. During those years the main analysis of poverty highlighted the fact that poverty was decreasing every year. However, later when the first longitudinal data became available, besides the number of people who were exiting poverty every year it was found a considerable number of people who also fell into poverty every year. The important point is that a significant part of the population was permanently facing the huge risk of becoming poor. Considering the data that was actually available at that time those people could have been easily identified. Therefore, this article focuses on the distribution of the risk of being poor among the population as well as the characterization of the people who were vulnerable and their differences from the observed poor.

The article is organized in the following way. First, we are going to introduce the concept of vulnerability in the particular context of Chile in the 1990s. This concept will then be measured and empirically analyzed. The next chapter is devoted to explaining the method chosen to estimate vulnerability at the household level using cross-sectional data. The method chosen has an important advantage in terms of data availability to many other developing countries. The results are divided between the analysis of how the risk of being poor is distributed among the population in Santiago in 1996, and the characterization of the vulnerable population and their differences with the poor in terms of observable characteristics. We also validate the results analyzing the predictive power of the model with the Panel-CASEN available. Finally, we will analyze some policy implications of this measurement that may lead to a more suitable classification to target the vulnerable population for different social policies.

2. VULNERABILITY TO POVERTY IN CHILE IN THE 1990s

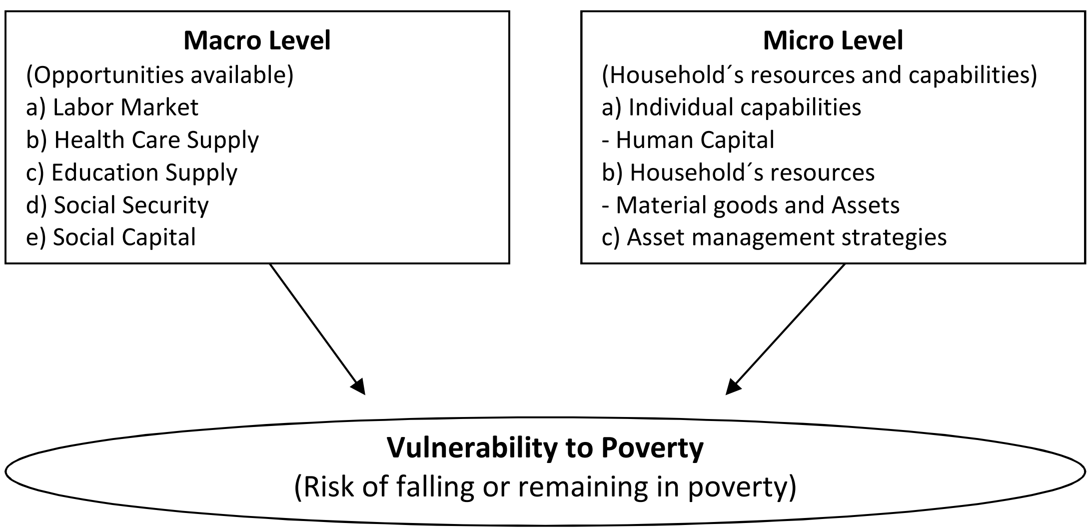
The starting point of this research is the idea of vulnerability to poverty. This approach has presented a new poverty paradigm in the last 15 years since it focused more on the assessment of the capacities and opportunities rather than the observed lack of income or consumption of a specific household. According to Chaudhuri et al (2002) what really distinguishes poverty and vulnerability is the idea of risk: “the fact that the level of future well-being is uncertain. The uncertainty that households face about the future stems from multiple sources of risk”. Thus, he states that “vulnerability is a forward looking or ex-ante measure of household`s well-being; (while) poverty is an ex-post measure of a household`s well-being (or lack thereof)”.

We have divided this section in three parts. First, as a theoretical approach we adapt the idea developed by Kaztman& Wormald (2002) and Moser (1998) to describe vulnerability to poverty. Then we describe some basic assumptions of the VEP approach which we used to estimate vulnerability at the household level. Finally, we want to highlight some interesting facts about the Chilean context in the 90s which was the period we choose to analyze the structure of vulnerability to poverty.

a. The idea of vulnerability to poverty

According to Chaudhuri (2003), vulnerability "depends on the complex dynamic interlinkages between the environment macroeconomic, institutional, sociopolitical and physical in which households operate; the resources, human, physical and financial it commands and its behavioral responses". A similar definition was previously proposed by Moser (1998) in her asset-vulnerability approach and Kaztman & Wormald (2002) later adapted to the Chilean case. The following figure summarizes that idea.

Figure 1. Social Vulnerability Framework[3]



According to Kaztman (2002) the asset-vulnerability approach opened the black-box of the poor since it considers them as an active agent to overcome poverty. Kaztman & Wormald (2002) also offered an extension of the model that more broadly emphasizes the structural vulnerability factors.

b. Chilean context in the 90s

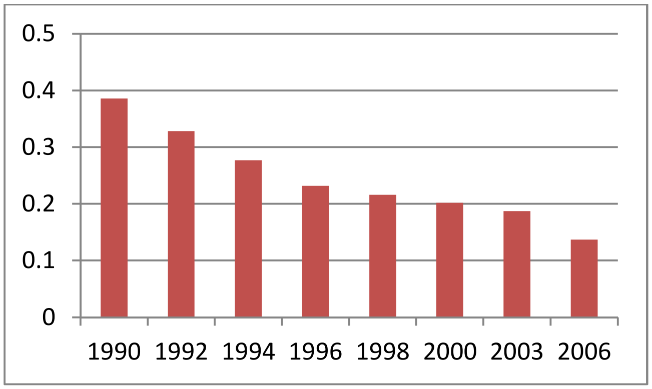
During the 90s Chile reduced poverty significantly as you can see in the Figure 2. According to Contreras & Larrañaga (1999), during the 1990s poverty was reduced on its three common measures[4]. There are a

[3] Adapted from Kaztman (1999), Wormald (2003) y Kaztman & Wormald (2002).

[4] After Sen (1976) published his seminal article criticizing the poverty measure as the percentage of the population that has an income/consumption level below a certain threshold, many articles have highlighted some other possible measures that address different issues related to the aggregation problem. The main idea is to measure not only the percentage of the population that is below a certain cutoff of the poverty line but also to measure how poor are those poor and how unequal are they. An extension of that measure is proposed by Foster et al.

number of articles that highlight the relationship between this huge poverty reduction and the accelerated economic growth in the 1990s[5].

Figure 2. Poverty evolution in Chile in the 1990s (Poverty Rate P_0). National Level



Source: CASEN (National Socioeconomic Characterization Survey) series. Own Elaboration

The tremendous success in terms of absolute poverty reduction was lessened when the national representative panel data became available. The Social Development Ministry decided to re-survey a subsample data of CASEN 1996 in 2001 and the result was a Panel data 1996-2001. That kind of data makes the reality of the evolution of poverty visible. The dynamics of poverty are revealed and that was not possible to see with just the series of cross section data in the CASEN surveys.

We highlight two main features of the Chilean poverty dynamics that are presented in the following table.

Table 1: Poverty Transition Matrix 1996-2001

1996/2001	Poor	Non-Poor	Total Row
Poor	45.16%	54.84%	**22.36%**
Non-Poor	11.36%	88.64%	**77.64%**
Total Column	**18.92%**	**81.08%**	**100%**

Source: Panel CASEN 1996-2001. Own Elaboration

Table 1 shows that although the poverty rate decreased from 22.36% to 18.92% in 5 years there was a lot of movement across the poverty line. More than 50% of the people who were poor in the first year (54.4%) exited poverty in 2001, but, on the other hand, a significant percentage of the non-poor (11.36%) fell into poverty in the same years. We could say that a conservative estimation of poverty dynamics should report that at least 32% of the population was in poverty in those 5 years. This estimation is conservative because we are just observing two points over time.

(1984). For the Chilean case in the 1990s, Contreras & Larrañaga (1999), show that there were significant reductions in all those poverty measures.

[5] See Valdés (1999); Contreras et al. (2001); Contreras & Larrañaga (1999); Olavarria-Gambi (2003); Katzman & Wormald (2002); and Meller (2002). For a broader or global perspective on this issue you can see Ravallion & Datt (1999); Lustig, Arias & Rigolini (2002); and Perry et al. (2006).

Besides the movement across the poverty line, an important point related to the notion of vulnerability is about the original position in the income scale of those people who fell into poverty in the second period of observation. This measure may give us an idea of the degree of vulnerability of the population. Contreras et al. (2004) illustrates this in the following table.

Table 2: Original Decile (1996) of per capita Income of those who fell into poverty in 2001

Per Capita Income Decile in 1996	Feel into Poverty in 2001
1	-
2	4.2
3	22.24
4	25.85
5	14.09
6	15.69
7	7.77
8	3.8
9	2.03
10	4.33
Total	**100**

Source: Contreras et al. (2004)

Contreras et al. (2004) argue that even people who were originally in the 6th decile of income showed a high proportion of poverty in the second period. Consequently, we could consider that Chile had a high proportion of people, regardless of their observed status in terms of the poverty/non-poverty situation, who were vulnerable to falling into poverty. Therefore, a more accurate description of the Chilean poverty evolution in the 1990s should take into account that regardless of the huge reduction in poverty there was also a huge vulnerability that was evident observing large households movements across the poverty line.

3. THE MODEL

The main challenge in assessing vulnerability to poverty is that we cannot directly observe it. Unlike poverty status which is observable for a specific household (you can easily tell whether a household has an income below the poverty line or not), vulnerability is a more complex phenomenon which requires some inferences to make about the prospects of future outcomes. A plausible way to define it can be the probability of the household to have a future income below the poverty line (z). This can be easily estimated if we assume that differences in terms of vulnerability are due to observable household characteristics (X_H). In that sense a general[6] specification of vulnerability for a household H at the moment t can be formally defined as:

[6] This general expression depicts the main intuition behind the estimation of vulnerability to poverty. As a general representation, it allows multiple model specifications that might be constrained by the available data. One of them is the eventual contribution of aggregate shocks in the macrostructure that each household faces (time variant coefficient β_t). As we will see above that coefficient is not possible to estimate using only one cross-sectional data.

$$(1)\quad v_{Ht} = \Pr\{y_{H,t+1} = y(X_H, \beta_{t+1}, \alpha_H, \epsilon_{H,t+1}) < z \mid X_H, \beta_t, \alpha_t, \epsilon_{Ht}\}$$

It is clear that a household's vulnerability is related to some stochastic properties for the inter-temporal income variation. Therefore, we need at least an estimation of the expected income and its variation over time. Ideally, we can use longitudinal panel data[7] to estimate both parameters for each household, but that kind of data is rarely available in many developing countries. Chaudhuri (2003) proposes a specific method to estimate vulnerability to poverty at the household level using only one cross-sectional data base. The method has three steps:

First, we assume that a household's income is a stochastic process determined by:

$$(2)\ \ln(y_H) = \beta X_H + \epsilon_H$$

Where y_H is the household total income and X_H represents several household characteristics. The first term in the right side of the equation refers to the systematic estimation of log-income, whereas $\mathcal{E}_H$ is a mean zero error term that captures idiosyncratic factors (shocks) that affect income at the household level. The idea is that $\mathcal{E}_H$ contributes to different income levels for households that are equivalent in terms of observational variables (X_H). It is assumed in the model that idiosyncratic shocks to income are identically and independently distributed over time for each household (inter-temporal variation), however, that assumption does not imply that $\mathcal{E}_H$ are identically distributed across households (as we typically want in OLS regression under homoscedasticity assumption).

The second step is to find an estimation of the household variability of income. As Chaudhuri (2003) proposes we assume a parametric estimation based on observable household characteristics[8].

$$(3)\ \sigma_H = X_H\theta$$

Therefore, for both, the log-income and income variation we have the following parametric equations:

$$(4)\ E[\ln(y_H) \mid X_H] = X_H\hat{\beta}$$
$$(5)\ Var[\ln(y_H) \mid X_H] = \hat{\sigma}^2_{e,H} = X_H\hat{\theta}$$

Chaudhuri (2003) assumes that the income distributes log-normal, so the household's vulnerability to poverty can be estimated as the probability that given some observed characteristics, a household will have a future income below the poverty threshold (z).

$$(6)\ \widehat{v_H} = Prob[\ln(y) < \ln(z) \mid Xh] = \phi\left(\frac{\ln(z) - X_H\hat{\beta}}{\sqrt{X_H\hat{\theta}}}\right)$$

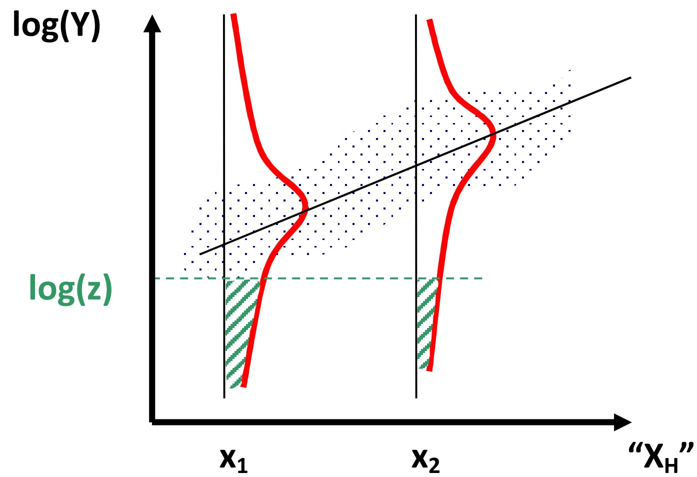
In (6) ϕ is the cumulative density of the standard normal function that gives a number between 0 and 1 for each household (0 represents the lowest probability to be below the poverty line in the future while

[7] See Pritchett et al. (2000) for an application of this method.

[8] Following Chadhuri (2003) we propose to estimate β and θ using three-step feasible generalized least square (FGLS) procedure suggested by Amemiya (1977).

1 represents the highest one). We can summarize the procedure in the Figure 3. For simplicity we show only one explanatory variable.

Figure 3: Log-Income volatility in a cross section data over the attribute X



In cross-sectional data we can estimate for each level of the explanatory variable X (say education) not only a point that belongs to the OLS line (as we usually do in regression analysis), but also a distribution due to income dispersion for each educational level. That distribution should not necessarily be the same for each kind of household[9], and we can actually calculate the part of that distribution that has an income below the poverty line (z). The shadowed area represents, for a household of such characteristics in terms of educational level, the probability of having an income below the poverty line –regardless of the actual income that it shows. We can extrapolate the bivariate case for a multilevel one using several explanatory variables. The main concern of this method is related to the assumption of the cross-section income variation as a proxy of the household income variation over time.

We analyzed several model specifications for the 2 steps. The more general one was the following:

$$(7)\ \ v_i = f(\mathrm{Charact}_H, \mathrm{Assets}_H, \mathrm{Transfers}_H, \mathrm{Context}_H)$$

"Charact" represents several variables related with the household composition (dep.ratio, head of household's age, Proportion of children/adults); "Assets" is a set of variables related to household assets (Head of HH schooling, home ownership, land, water access, etc.); "Transfers" is a dummy variable that shows whether the household receives a government transfer or not, and finally "Context" is a set of variables that characterizes the neighborhood of the household (average schooling and unemployment rate). Using that model, the main findings are shown in the following section.

[9] Actually, households with zero educational level may present large log-income variability based on the different working conditions of the parents (employed with at least the minimum wage vs. unemployed with no wage), rather than an average estimation based on the entire population.

4. RESULTS

Based on the methodology applied we can present a general description of the structure of the vulnerability to poverty in Chile in the 1990s. There are many issues that can be analyzed using this measure of vulnerability. First, using both the observed poverty and vulnerability estimation we can create a new classification of households in terms of the combination of both categories. Then, we can also show a general distribution of the risk of being poor among the population. A special section will assess the predictive power of the model using the observed movements around the poverty lines in two later periods available in the Panel Survey of CASEN 1996-2001-2006.

a. Distribution of the risk of being poor among the population

The estimation of the vulnerability to poverty at the household level allows us to characterize the population in terms of the risk of being poor rather than their current observed state of being poor or non-poor. By construction, the average level of the vulnerability will be approximately equal to the poverty ratio; however, since the risk of being poor is spread among the entire population we will be able to make a more complex description of the population that may inform social policies in a more appropriate way.

Typically the design of social policies takes into account two relevant groups among the population: poor and non-poor. In the case of Santiago in 1996 the proportion of urban households below the poverty line was 12.39%. However, that number does not take into account the heterogeneity of the people under the poverty line and it also misses another important group in terms of poverty reduction: those who are currently not poor but vulnerable to becoming poor. The combined use of the poverty and vulnerability measures will allow us to identify those groups.

It is not clear how to define the threshold to which a household may be considered vulnerable. Within the literature it can be identified by two different measures. The first definition called extremely vulnerable identifies those households that face a probability higher than 50% of being poor. The goal is to identify as vulnerable those who have an equal chance of being in or out of poverty. A more moderate definition of vulnerability focuses on those households that have a higher probability than the greater society of being poor (i.e. in this case, those with a probability higher than 12.39%). The idea in this case is to focus on those households that exhibit the main risk of being poor among the population. Table 3 shows a classification of the population considering both poverty and vulnerability measures.

Table 3. Groups Comparison using Vulnerability Classification and Poverty Rate

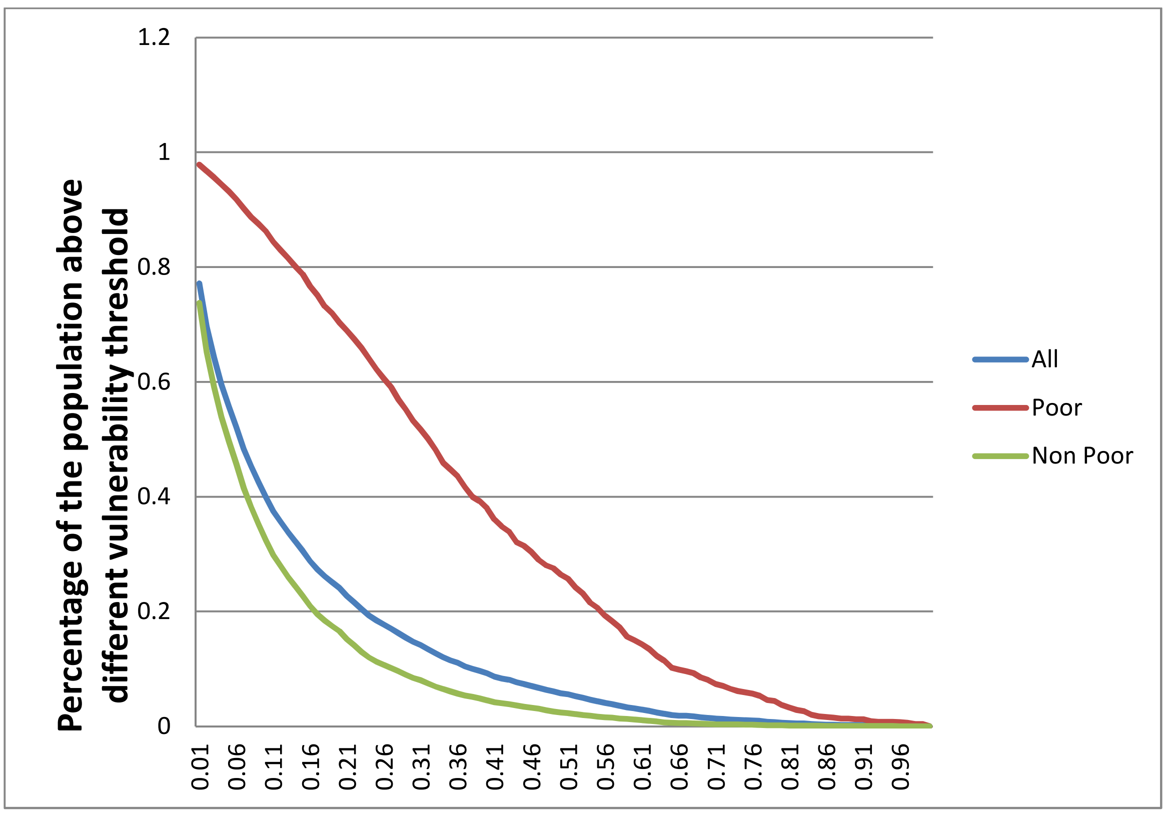
Groups	V mean	v >0.5 (Ext.Vuln)	Pov.R <v <0.5 (Mod.Vuln)	v<Pov.Ratio (Non Vulnerable)	Total
Non-Poor	0.091	3.83%	19.96%	63.82%	87.61%
Poor	0.3328	3.43%	6.66%	2.30%	12.39%
Total	0.1209	7.26%	26.62%	66.12%	100.00%

Source. CASEN Chile, 1996. Special Run.

The description of the population is richer when considering the risk of being poor that each household faces. In that way, among those who are currently below the poverty line, we can further distinguish those with higher vulnerability (3.83% or 10.09% depending on the cutoff point that we use to identify vulnerability: we can call them chronically poor). That segment of the population, with their current low income level, has a higher probability of remaining in poverty in the future. We can then think about the suitable policies for this group. A second group is composed of those who besides of being below the poverty line have a low probability of remaining in poverty in the future (8.96% or 2.3%, again depending the cutoff point that we use to identify vulnerability: we can call them transient poor. Finally, another social policy group of interest arises: those who in spite of their current level of income above the poverty line, based on their vulnerability estimation, present a high probability to be below the poverty line in the future (3.83% or 23.79%: we can call that group the non-poor vulnerable ones). Therefore, the interest groups, in terms of social policies, rather than being only the 12.39% whose incomes are below the poverty line, actually represent almost 37% of the population. Among other facts that can be highlighted, we can mention that among the extremely vulnerable (over 50% chance of being poor) there are more people above (observed-non-poor) than below (observed-poor) the actual 1996 poverty line.

A more general analysis of the distribution of the risk of being poor among the population requires relaxing the threshold definition of vulnerability. An attempt to do that is shown in the Figure 4:

Figure 4: Aggregate incidence of vulnerability to poverty under different vulnerability thresholds.



In the figure above we can directly see what percentage of each group (all population, poor, and non-poor) would be classified as vulnerable for each defined threshold. For example, if you want to consider vulnerable those households that face a risk of being poor higher than 50% (highly vulnerable) that would lead you to identify as vulnerable almost the 20% of the poor, around 5% of the non-poor and around 7% of the entire population. At first glance, we can see that the observed-poor are notably the group that concentrates the highest risk of being poor to all vulnerability cutoffs. Considering that they are just the 12.39% of the population, the curves that show the risk of the non-poor and the entire population are very similar since those groups share an important part of their members.

b. Characterization of the vulnerable population

As we mentioned earlier there is a strong correlation between actual poverty and vulnerability since the latter is defined as the probability of being poor. In this section we want to characterize the vulnerable group and compare them to the poor. The Table 4 offers a first approach to that comparison.

Table 4: Group average on characteristics related to income

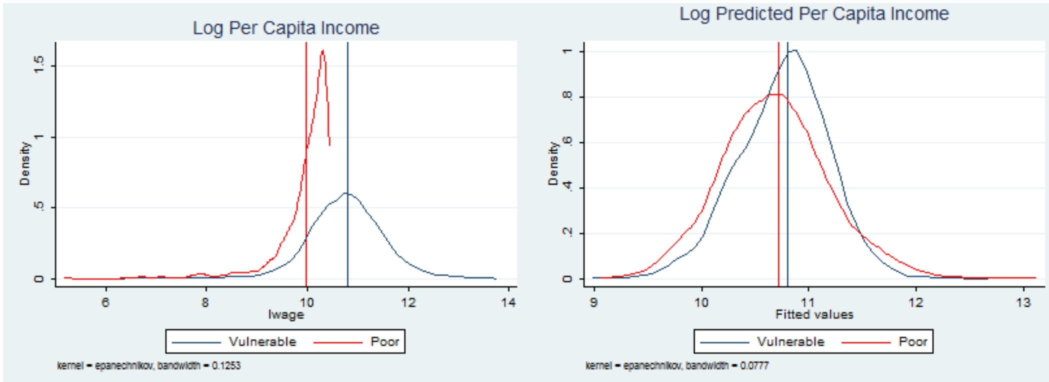
	Poor	Vulnerable	Non Poor	Non Vulnerable
% Population	12.39%	33.88%	87.61%	66.12%
Log Per Cap Income*	9.980767	10.80037	11.68021	11.81553
Predicted Log PC Inc*	10.71828	10.80516	11.5841	11.81109

Source. CASEN Chile, 1996. Special Run.

Table 4 compares the predicted and actual income (in log scale) for different groups. First we can see that the vulnerable and the poor are very different groups in terms of observed income. However, considering the size of each group: poor (12.30% vs. vulnerable: 33.88%); there is a surprising similarity in terms of predicted income for both groups, though the differences are statistically significant as compared in a mean test. In purely statistical terms we could say that income also accounts for many unobserved factors, both random and systematic. However, in terms of policy implications we could argue that since there is an important part of income that is not explained by household characteristics, at least for those variables that are typically observable. In that sense there is strong reason to suspect of the income as a unique reliable measure for some policy decisions as targeting on mean-tested programs.

Although the means test for the average predicted income was statistically significant, the gap in terms of observed income is dramatically reduced when we compare the predicted income instead. A general depiction of that gap reduction can be observed in the Figure 5 comparing the kernel density distribution of the two groups.

Figure 5: Kernel Density Distribution of Predicted and Observed (Log) Income for different groups.



The main explanation for this gap reduction or similarity in terms of predicted income may be the following: although there are some particular differences, both groups present a similar portfolio of assets, demographic and context characteristics that overall predict a similar income for both groups (poor and vulnerable). That idea, to some degree, may be corroborated looking at the main observable variables that affect incomes. In the Table 5 we can observe that.

Table 5. Group averages on observable characteristics

	Poor	Vulnerable	Non Poor	Non Vulnerable
Schooling*	7.935799	8.31357	10.35631	10.92397
Dep.Ratio	4.68052	4.62275	3.80913	3.55552
Female (D)	0.258786	0.231716	0.242814	0.251495
Age*	41.2056	43.7688	48.3218	49.3212
Neigh.Unemp. ratio	0.389141	0.399312	0.369491	0.357893
Neigh Schooling	6.610995	6.883233	8.553651	9.045527
Prop. kids*	0.262675	0.229577	0.139387	0.116277
H Owner (D)	0.319939	0.313051	0.510502	0.575965

Source. CASEN Chile, 1996. Special Run.

In the example where we compared the predicted income, in terms of each of the characteristics presented in Table 5, we can observe that the poor and vulnerable groups are very similar to each other and particularly different from the non-poor and the non-vulnerable group. However, it is important to highlight that the groups in Table 5 are not excludable. The idea in that case is to compare the observable characteristics of quite different groups in terms of population size.

A comparison considering excludable groups is presented in Table 6. We can easily show that among the poor/non-poor groups the distinction among the vulnerable/non-vulnerable is closely associated with significant differences in terms of observable household attributes. In other words, we can say that those groups can be easily identified using the available cross section information of households. Table 6 shows that.

Table 6. Group averages on observable characteristics

	Poor		Non Poor	
	Non Vulnerable	Vulnerable	Non Vulnerable	Vulnerable
Schooling	8.811647**	7.731306	11.00025**	8.564188
Dep.Ratio	3.99751**	4.83656	3.53956**	4.53212
Female (D)	0.2259732	0.266282	0.2524161	0.217062
Age	47.4966**	39.7684	49.3871**	45.4646
Neigh.Unemp. ratio	0.3769949**	0.3919153	0.3572034**	0.4024473
Neigh Schooling	7.625986**	6.379117	9.096786**	7.09694
Prop. kids	0.1878058**	0.2797792	0.1136943**	0.2082944
H Owner (D)	0.5984429**	0.2563143	0.5751535**	0.3371024

Source. CASEN Chile, 1996. Special Run.
**Ttest: Significantly different at 5%

c. Validation of the methodology

As we discussed earlier, a robust estimation of vulnerability requires many panel waves that may allow you to observe the current dispersion of income for a specific household over time. In this case we have used as a proxy of the inter-temporal variance of income the cross-section variances of households with similar observable characteristics. Although it is likely to be a smart strategy to measure vulnerability to poverty as the probability of being poor observed in a cross section data which is the available data for many countries, a validation of the predictive value method is an interesting issue to be analyzed.

Based on our findings using CASEN (Urban households of Santiago in 1996) we estimate the vulnerability to poverty in each observation of the subsample of CASEN that was resurveyed later (the PANEL-CASEN 1996-2001-2006). In that sense we got an idea of the predictive value of the method based on the actual observations of the households in terms of entry/exit to poverty. The Table 7 summarizes the main results for the different poverty transitions between 1996 and 2001.

Table 7. Imputed vulnerability average for different poverty trajectories observed between 1996 and 2001.

1996\2001	*EP*	*PNE*	*NP*
Extremely poor (EP)	0.633	0.468	0.340
Poor, but not extremely poor (PNE)	0.523	0.503	0.316
Non poor (NP)	0.317	0.321	0.149

Source: Panel-CASEN 1996-2001.

As you can see in Table 7 there is a clear correlation between the direction of the trajectory (upwards or downwards the poverty and the extreme poverty line) and the estimated vulnerability in 1996. As an example, you can see that those who were identified as extremely poor in 1996 and 2001 were also the group with the highest vulnerability level in 1996 (on average a 63.3% probability of being poor). The extreme poverty line is based on an ad-hoc definition (z/2) that in a sense represents those households that do not have enough resources to even pay for basic food goods[10]. Similarly, those who were poor in 1996 and also fell into extreme poverty in 2001 were already revealed as the more vulnerable ones in 1996. The table shows many other trajectories that may be interpreted as the predictor power of the vulnerability estimations in the sense that they may predict movements around the poverty line.

Finally, in Table 8 we can observe an extension of the argument presented above in terms of the predictive power of the estimation using a longer observable trajectory looking at the transition of the households also in 2006.

Table 8. Imputed vulnerability average for different poverty trajectories actually observed between 1996, 2001 and 2006.

$t_0 \backslash t_1$	*1996-2006*		*2001-2006*	
Poverty Transitions	*P*	*NP*	*P*	*NP*
Poor (P)	0.439	0.356	0.479	0.350
Non-poor (NP)	0.200	0.144	0.173	0.148

Source: Own Elaboration using Panel-CASEN

d. Methodological caveats

The idea of this section is to discuss many issues that may arise in the methodology that we adopted, specifically for the Chilean case in the 90s.

First, the assumption about the households' income variation seems to be a relevant aspect to discuss. As we discussed in the appendix, ideally we should estimate these models using a large panel data that may allow us to empirically observe how households' income varies over time. However, panel data does not directly solve all the problems that may arise and is also rarely available[11]. Thus, using the cross-sectional income variance as a proxy for the inter-temporal variation of the households' income seems to be a reasonable strategy in order to identify those households who face riskier situations in terms of the income predictions for the future. In that sense, we should also say that since we are estimating not only income differences but also an income distribution for each level on the explanatory variables, a very clear caveat is sample size of the data available. This aspect should not be an issue due to the large sample data that we are using with CASEN which is the largest socioeconomic data available in Chile.

Second, given that we are not controlling for macro structural changes, we should be aware that this model offers a good estimation only for stationary periods. The main point here is that our estimation is based on shocks that may affect a household in a specific period of time but it does not offer any specification for aggregate shocks that may affect a number of households at the same time. As a result, we cannot be sure of the quality of the prediction in periods of macro changes (Ex. Economic crisis).

[10] The poverty line in Chile is defined as having enough resources (income from different sources, labor, subsidies, imputed rent, etc.) to acquire the basket of basic foods. It is also assumed that a household spends half of its income on food and half on non-food goods. Therefore, to be identified as poor, a household has to present an income lower than z that compounds 2 baskets of basic foods.

[11] As an example, we could mention that without panel data we cannot control for unobserved household-level effects which might bias the coefficient of the observed variables. However, as we discuss here, panel data as well as cross-sectional data does not control for macro variables that may change over time and directly affect household's incomes.

However, in our case, there is no evidence of important changes in the Chilean economy in 1996 or other changes related to macro variables that may have affected household incomes and their volatility over time.

A third issue that might be discussed is related to the heteroskedasticity assumption that typically arises in OLS analysis. In cross-section regression, heteroskedasticy implies efficiency loss in the coefficients (though we know that it does not affect them in terms of bias). In order to avoid that problem, in most cases we try to find a model specification with the same variance for all households (homoscedasticity). This seems to be quite restrictive for our purposes since we are trying to find a proxy of the inter-temporal variance of the household income. For that reason, instead of assuming homoscedasticity, where all households have the same income variance, we followed Chaudhuri (2003) who proposes a more general model in which the variation of income could also be related to some observed household characteristics (equation 3).

Finally, in terms of validation, we assessed the predictive value of the model using a simple panel data. We compared the vulnerability levels of those families who were poor in the baseline of the panel (1996) and they showed different trajectories in 2001 (poor-poor; poor-non-poor). As we would expect from a model with good predictive power, those who qualified as chronic poor (poor-poor) had a higher probability to actually remain in poverty five years later (1996). The opposite was also true since the group that showed a lower vulnerability was more likely to exit poverty and not fall into poverty in the second period of observation. As a result, this methodology has proven to be a useful tool for this specific period that might also encourage a better design of social policies that could lead to a better strategy in terms of sustainable poverty reduction for many developing countries.

5. CONCLUSIONS

We have been able to analyze the structure of vulnerability in Chile using cross sectional data, particularly for the case of urban households in Santiago, which represents almost 90% of the population of the main city in Chile. The VEP (vulnerability as expected poverty) approach has allowed us to estimate at the household level the probability of certain families to fall into or remain in poverty in the future. The main weakness of this approach is the assumption of the cross-section variance as a proxy of the household inter-temporal variance of income. However, taking into account the lack of a large sample of longitudinal data which is commonly the case in most developing countries, this approach seems to be a reasonable way toward a more dynamic analysis of poverty. In any case, we validated the model in terms of its predictive power using an available panel data and we saw that our vulnerability measure was strongly related to the direction of the poverty trajectories that were actually observed. In other words, those families who followed a downward (upward) poverty trajectory after 1996 were the ones that actually showed a higher (lower) vulnerability level in the first period (1996).

In that sense, this article aims to introduce a more comprehensive analysis of poverty that needs to overcome the simple distinction between poor/non-poor. There are two main arguments to support that idea. First, the acknowledgment that the distinction between poor/non-poor misses an important group that in the case of Chile in the 1990s, has directly affected the evolution of poverty: the non-poor people who are likely to be poor in the future. In 1996, 23.79% of the population belonged to this category, much more people than the observed 12.39% who were below the poverty line in that year. Moreover, that classic definition of poverty usually sees people in poverty as a homogeneous group and that conclusion results in a similar set of policies. As we have seen in the Chilean case, among those who were in poverty, there was a clear distinction in terms of their characteristics and productive assets at least between two relevant groups: the ones who are likely to remain in poverty in the future due to the low level of their productive assets (10.09%), and those who are likely to exit poverty precisely due to their characteristics and productive assets (2.39%). Different sets of policies are recommended for those different groups. Therefore, besides the analysis of the structure of vulnerability in Chile in a context of poverty reduction, there are two further contributions in terms of policy implications that are presented below.

a. An approach to define vulnerability line

In a practical sense vulnerability has been defined as a broader percentage of the population that is above the poverty line. In that sense, the World Bank has defined vulnerability as 2.5 times the traditional poverty line. With the model estimated here we could have a sense of that ad-hoc definition in terms of the proportion of the vulnerable population who might fit into that definition.

Table 9. Comparison between Our Definition and World Bank ad-hoc definition of Vulnerability

Our Definition / World Bank Definition	Non-Vulnerable	Vulnerable
Non-Poor nor Vulnerable	45.39	19.11
Poor and Vulnerable	0	9.95
Poor and Non-Vulnerable	0	2.33
Non-Poor and Vulnerable	5.94	17.27
Total	51.33	48.67

Source: Own Elaboration using CASEN 1996

As you can see in the Table 9, according to the World Bank classification of vulnerable groups (poverty line as 2.5z) 48.67% of the population would be considered as vulnerable in 1996. Besides the traditional 12% of people below the poverty line, this definition adds as a targeting group for social policies an extra 19.11% who, according to our classification, were non-poor and vulnerable and also 17.27% of people who were estimated as vulnerable. In that sense we can see that at least 25% (5.94 over 23.21) of the Non-Poor Vulnerable group are still not considered part of the targeted population.

Following the idea of the World Bank to define vulnerability just as an extension of the poor, perhaps an interesting question to be addressed is how to set the vulnerability line. In that sense, a more appropriate definition of vulnerability should consider at least 90% of the Non-poor vulnerable group. We calculated that based on CASEN 1996 and we found that we have to use a vulnerability line that is almost 4 times (Z_v= 3.966Z_p) the traditional poverty line. In that estimation the all targeted population would be 67% of the population. In that case, the characteristics of the different groups are described in the Table 10.

Table 10. Comparison between Our Definition and a new ad-hoc definition of Vulnerability

	Non Poor		Non Poor		
	Non Vulnerable (y>4z)	Vulnerable (y<4z)	Non Vulnerable ($v>P_0$)	Vulnerable ($v>P_0$)	Poor
% Population	**32%**	**55%**	**63.80%**	**24%**	**12.39%**
Schooling	12.7136	8.966681	11.00025	8.564188	7.935
Dep.Ratio	3.35765	4.07787	3.53956	4.53212	4.68
Female (D)	0.2460565	0.2402793	0.2524161	0.217062	0.2587
Age	49.0544	47.8663	49.3871	45.4646	41.2
Neigh.Unemp. ratio	0.3667729	0.3714234	0.3572034	0.4024473	0.389
Neigh Schooling	9.932751	7.746044	9.096786	7.09694	6.61
Prop. kids	0.0982738	0.1634236	0.1136943	0.2082944	0.2626
H Owner (D)	0.5464675	0.4888577	0.5751535	0.3371024	0.3199

Source: Own Elaboration using CASEN 1996

In Table 10 you can compare the differences between the observable poor and the vulnerable defined as both, an income threshold (4z) that accounts for the 90% of the non-poor vulnerable, and a probability to fell in poverty greater than P_0 (12.39%). As you can see with our definition of vulnerability ($v>P_0$) we found a much more similar to the observable poor than this alternative vulnerable group (y<4z).

b. A new classification of social policy groups

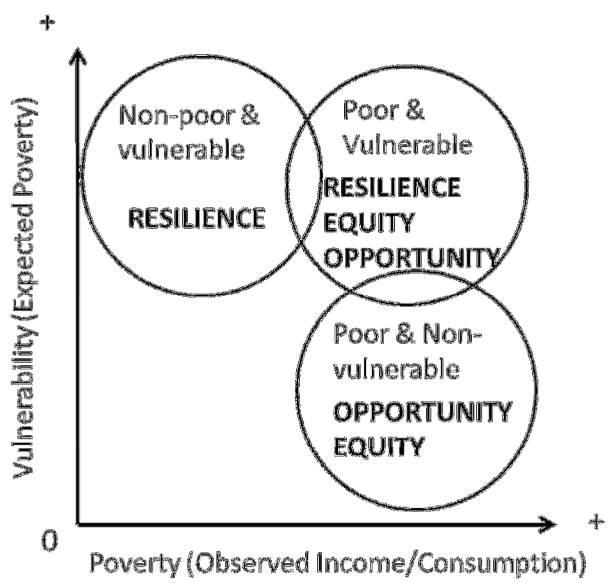
The methodology presented here shows a practical way to characterize the population of a country in order to implement suitable policies that may reduce poverty in a sustainable way. The identification of different groups may encourage the design of suitable policies for the particularities of each group. In this case, the addition of vulnerability as a relevant dimension of a household's wellbeing may allow us to identify three relevant groups in terms of policy design: (i) current poor and vulnerable, (ii) current poor but non-vulnerable, (iii) current non-poor but vulnerable. Although there is no "one-size-fits-all" kind of policies for each identified group, this framework aims to provide some guidance in terms of what kind of policies are more appropriate to strengthen in order to increase the chances of each group to exit or avoid poverty in the future.

On the other hand, the World Bank proposes resilience, equity and opportunity as the main goals of its current social protection and labor strategy (Robalino et al., 2012). The World Bank report states that a major challenge of effective access to Social Protection and Labor is to "ensure that programs –and ultimately the whole SPL system in a country- are responsive to the needs of various groups and risks, drawing a "portfolio" of programs that together provide resilience, equity and opportunity to all who need them" (Robalino et al., 2012, p.27). To some extent that policy classification fits very well with the major needs of each group identified in this article, so it could define an implicit guidance of what kind of policies strengthen each group.

The first group of programs, which is called the equity programs, is designed to protect against destitution and promote equality of opportunity to the entire population. Social assistance programs as safety nets that include transfers and in-kind transfers such as school feeding and targeted food assistance are presented as clear examples of this kind of policy that aim to ensure a basic level of well-being to everyone. Similarly, the World Bank defines those programs that aim to increase opportunity among the population as another pillar of its social protection and labor strategy. The opportunity programs are those that promote better health, nutrition, education, and skills development and also help women and men secure better jobs. Finally, the resilience programs are those that insure the population against the effects of decreased well-being from a range of possible shocks, e.g. unemployment compensation, disability insurance, old-age pensions. Particularly, to those who have a certain level of well-being these kinds of programs are more appropriate in a contributory basis, but due to their general set of assets also face a risk to their well-being, although they do show a certain level of income above the poverty line.

Based on that classification of social policies we propose to emphasize different set of policies for each identified group. They are described in the following scheme for a suitable design of social policy.

Figure 7. General description of a suitable policy for identifiable groups



6. REFERENCES

Amemiya, T. (1977). “The maximum likelihood estimator and the non-linear three stage least squares estimator in the general nonlinear simultaneous equation model”, Econometrica 45, 955-968.

Calvo, C (2005) The measurement of vulnerability to poverty. Mimeo, Department of Economics, University of Oxford.

Calvo, C. & S. Dercon. (2005). Measuring individual vulnerability. Discussion paper series, Department of Economics, University of Oxford, number 229.

Chaudhuri, S. (2003). Assessing vulnerablity to poverty: concepts, empirical methods and illustrative examples. School of International and Public Affairs, Columbia University, NewYork. Unpublished manuscript.

Chaudhuri, S., J. Jalan, and A. Suryahadi (2002), Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia, Columbia University, Discussion Paper 0102-52.

Christiaensen, L. & K. Subbarao (2004), Toward an Understanding of Household Vulnerability in Rural Kenya. World Bank Policy Research Working Paper 3326.

Contreras, D., Larrañaga, O. Litchfield & A.Valdés (2001). Poverty and Income Distribution in Chile, 1987-1998. New evidence. Cuadernos de Economía 38(114): 191-208.

Contreras, D. Cooper, R. Herman, C. Nelson, C. (2004). Dinámica de la Pobreza y movilidad social: Chile 1996-2001. Documento de trabajo, Universidad de Chile, Escuela de Economía.

Contreras, D & Osvaldo Larrañaga. (1999). Los activos y recursos de la población pobre en América Latina: El caso de Chile. Banco Interamericano de Desarrollo, Documento de Trabajo R358.

Dercon, S. (2005). Vulnerability: A micro perspective, in Insurance against Poverty Oxford University Press.

Haughton, Jonathan, and Shahidur R. Khandker. Handbook on poverty and inequality. World Bank-free PDF, 2009.

Hoddinot, J. & Agnes Quisumbing. (2003a). Data sources for microeconometric risk and vulnerability assessments. Social Protection Discussion series N0323, Washington DC. Human Development Network, The World Bank.

Hoddinot, J. & Agnes Quisumbing. (2003b). Methods for microeconometric risk and vulnerability: A review for empirical examples Social Protection Discussion series N0324, Washington DC. Human Development Network, The World Bank

Kamanou, G & J.Morduch (2002) Measuring vulnerability to poverty. Paper No 2002/58, World Institute for Development Economics Research, United Nations University.

Kaztman, R. (2001). Seducidos y Abandonados: El aislamiento social de los pobres urbanos. Revista de la CEPAL, Nº 75, LC/G.1969-P, Santiago de Chile, Comisión Económica para América Latina y el Caribe (CEPAL).

Kaztman, R. (2005). Notas sobre la medición de la vulnerabilidad. En BID-Banco Mundial-CEPAL-IDEC, 5to Taller Regional. La medición de la pobreza: Métodos y Aplicaciones. Santiago de Chile.

Kaztman, R. & G.Wormald. (2002). Trabajo y Ciudadanía. Los cambiantes rostros de la integración y exclusión social en cuatro áreas metropolitanas de América Latina.

Kaztman, R. (1999). Activos y estructura de oportunidades: estudio sobre las raíces de la vulnerabilidad social en Uruguay. Montevideo: CEPAL.

Larrañaga, O. (2005) Incorporating social (minimum) services in the measurement and assessment of poverty: Conceptual and policy issues with an application to Chile. Conference paper, Arusha Conference,“New Frontiers of Social Policy” – December 12-15

Ligon, E. (2003). Estimating vulnerability: Static welfare measures in a dynamic world. Mimeo, Stanford University and the Department of Agricultural and Resource Economics, University of California, Berkeley

Ligon, E. & Laura Schechter. (2004). Evaluating different approaches to estimating vulnerability. World Bank, Social Protection Discussion Paper No. 0410.

Ligon, E. & Laura Schechter. (2003). Measuring vulnerability: The Director´s cut. The Economic Journal, Volume 113 Issue 486 Page C95

Lustig, N. Arias, O & J.Rigolini. (2002) Reducción de la Pobreza y Crecimiento Económico: La doble causalidad. Banco Interamericano de Desarrollo. Disponible en www.iadb.org/sds/pov

Meller, P.(2002). El Modelo Económico y la Cuestión Social. En Revista Perspectivas, Departamento de Ingeniería Industrial, Universidad de Chile.

Moser, C. (1998). The asset-vulnerability framework: reassessing urban poverty strategies. World Development 26 (1): 1-19.

Olavarría-Gambi, M. (2003). Poverty Reduction in Chile: has economic growth been enough?. Journal of Human Development, Vol 4, #1.

Perry, G., Arias, O., López, H., Maloney, W. & Server, L. (2006). Poverty Reduction and Growth: Virtuos and Vicious Circles. The World Bank. Disponible en www.worldbank.org/lacpublications

Pritchett L., Suryahadi A. & S. Sumarto (2000), Quantifying Vulnerability to Poverty: A Proposed Measure, with Application to Indonesia. SMERU Working Paper, Social Monitoring and Early Response Unit (SMERU), World Bank, Washington.

Ravallion, M & G.Datt (1999). When is Growth Pro-Poor? Evidence from the Diverse Experiences of India's States, World Bank Policy Research Working Paper 2263, Banco Mundial, Washington, D.C.

Robalino, D. A., Rawlings, L., & Walker, I. (2012). Building social protection and labor systems: concepts and operational implications. Background Paper prepared for the Social Protection and Labor Strategy, 2022.

Sen, A. (1976). Poverty: an ordinal approach to measurement. Econometrica 44(2), 219-231.

Valdés, A. (1999). Pobreza y Distribución del Ingreso en una Economía de Alto Crecimiento: Chile 1987-1995. Estudios Públicos 75.

Wormald, G. (2003). Algunas reflexiones sobre pobreza y vulnerabilidad social. Revista CIS, Fundación Un Techo para Chile.

7. APPENDIX

7.1. Descriptive Statistics of CASEN (1996)

Variable	Obs	Mean	Std. Dev.	Min	Max
Head of HH' schooling {years}	7,128	9.648148	4.395515	0	21
Number of HH's individuals	7,228	3.94383	1.78904	1	20
Female HH's head {D}	7,228	0.247925	0.431838	0	1
Head of HH's age {years}	7,228	47.83107	15.28026	19	98
Head of HH's unemployed {D}	7,228	0.241976	0.428309	0	1
HH's infants proportion	7,228	0.077918	0.132711	0	1
HH's adult proportion	7,228	0.660641	0.267969	0	1
Head of HH's Work condition {D}	7,228	0.400664	0.490067	0	1
Home ownership {D}	7,228	0.511206	0.499909	0	1
No Electricity {D	7,228	0.003182	0.056324	0	1
Health Shock {D}	7,228	0.178196	0.382704	0	1
Neighborhood Schooling	7,228	8.040349	2.175143	3.166667	19
Neighb's unemp.ratio	7,228	0.377795	0.12157	0	1
Log of HH's total income	7,202	11.39262	0.96221	5.170484	16.03

7. 2. Different Vulnerability to Poverty Approaches

Basically we can identify three main approaches to measure vulnerability: VEP (vulnerability as expected poverty, in which we focused in this article), VEU (vulnerability as expected utility) y VER (vulnerability as uninsured exposure to risk).

a. VEP[12]: It considers vulnerability to poverty as a household/individual probability to fell/remain in poverty in the future. A formal description is the following:

$$v_{Ht} = P(c_{Ht+1} < z)$$

Other scholars[13] expand the time horizon using a similar analysis.

The main advantage of this approach is the great availability of cross sectional data. As a drawback we can mention the strong assumption of using the cross sectional variance as a proxy of the inter-temporal income variation.

Some extensions of this model can be found using pseudo panels based on a series of cross sectional data[14]. Ligon & Schechter (2004) evaluating different approaches to measure vulnerability highlight the one proposed by Chaudhuri (2002) as th

---

[12] Tos me applications: Chaudhuri et al. (2002); Chaudhuri (2003); Christiaensen & Subbarao (2001), (2004); Kamanou & Morduch (2002); McCulloch & Calandrino (2001)

[13] Ver Pritchett, Suryahadi & Sumarto (2000),

[14] Bourguignon & Goh (2004) use a series of repeated cross sectional data to recover the main parameters of individual income dynamics. They found that those results may fit well what they observe in a panel data. According to them that measure may be pretty useful to estimate other relevant measures as the vulnerability to poverty at the individual level.

e best for stationary periods with income/consumption measurement without significant error.

b. VEU[15]: This approach asume that vulnerability i associated with a well-being loss that may have two different cuases: income/consumption or the level of risk or uncertainty that a household faces. Therefore, it defines vulnerability to poverty as the difference between the utility derived from minimum consumption (poverty line) and the expected consumption value.

$$V^i(C) = U^i(z) - E[U^i(c^i)]$$

Likewise, if a household has an income (or a consumption level) of C which is certainly above the poverty line it has zero vulnerability level. However under the assumption that consumption utility is a growing and concave function that satisfies Jensen´s inequality which considers individuals as risk averse.

On the other hand, different model extensions distinguish the different risk in two levels. Thus, vulnerability can be defined as a triple component as:

$$V^i(c) = [U^i(c) - U^i(E[c])] + [U^i(E[c]) - E[U^i(E[c \mid x])]] + [E[U^i(E[c \mid x])] - E[U^i(c)]]$$

The first expression –deterministic- is called vulnerability related to household´s poverty, then the second term is related with the aggregate risk that household face in the simple and, finally, the last term represents the household idiosyncratic risk.

The main advantage of this method has to do with the possibility of differentiate the effect between aggregate and idiosyncratic shocks. However, some disadvantages are related with the estimation method which requires a large sequence of panel data. Additionally, this method requires an a priori definition for the utility function.

c. VER[16]: This method distinguishes among different shocks that a household may face. Some of them can be aggregate shocks that affect the entire population (Economic crisis, drought), while others may directly affect a specific household (illness). It requires panel data since you need a complete sequence of the same individuals over time. Finally, it doesn´t allow to estimate vulnerability at the household level.

Overall, we can say that VEP as well as VEU do refer to a basic measure (poverty line) to then attempt to estimate the probability of being below that defined threshold. Similarly, VEP and VEU can build a vulnerability measure at the household level On the other hand, VER does not estimate a probability as before and as a result it does not allow you to vulnerability at the household/individual level. The main VER´s focus is on distinguish different observed shocks that make wellbeing decrease.

7.3. Estimating Vulnerability with different data

[15] See Ligon & Schechter (2003), Ligon (2003), Kurosaki (2006),

[16] Algunas aplicaiones se pueden encontrar en: Dercon & Krishnan (2000), Tesliuc & Lindert (2002).

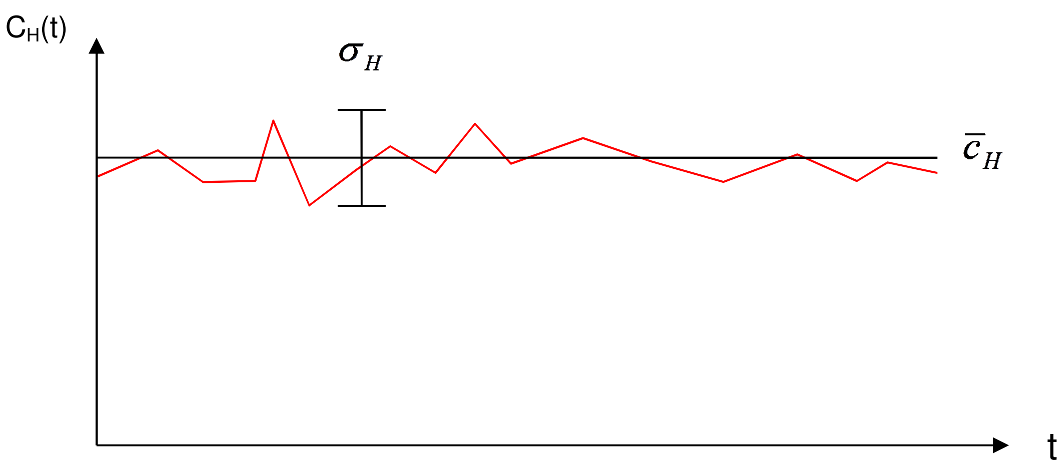
This section aims to distinguish the advantages and disadvantages of at least three methods and to discuss its implication for the Chilean case in terms of data availability.

a.- Panel of T waves (T>>2)

This case might be the best to measure vulnerability at the household/individual level. A sequence of observations over time for the same individuals may allow you to estimate their income volatility and their elasticity to idiosyncratic of aggregate shocks. The main drawback of this method is its scarce availability[17].

Three steps may allow you to estimate vulnerability at the household/individual level as you can see in the following figure:

Figure 7.3. Estimating Vulnerability using a panel data of T waves (T>>2).



In the Figure 7.3you can see the income/consumption of a specific household/individual over time. You can easily estimate two relevant measures here: $\bar{c}_H$ (Income Temporal Mean) and its variance over time σ_H.

- Income Temporal Mean

$$\bar{c}_H = \frac{1}{T}\sum_{t=1}^{T} c_{Ht}$$

[17] Una interesante aplicación de este método se encuentra en Baulch & McCulloch (1998), (2000), Christiaensen et al. McCulloch & Calandrino (2003).

- Income variance over time

$$\sigma_H^{\ 2} = \frac{\sum_{t=1}^{T}(c_{Ht} - \bar{c}_H)^2}{T-1}$$

Thus the vulnerability to poverty for a specific household can be estimated by:

$$v_H = \Pr(c_{HT} < z \mid X_H) = \Phi(\frac{\bar{c}_H - z}{\sigma_H})$$

b.- Two-period panel data

In this case we have a panel data of just two periods for each individual/household. There is no easy way to estimate vulnerability at the household/individual level; however you can certainly compare the differences between those who fell/exit poverty in those periods. Therefore, you can find what factors might be related with the different poverty transitions observed (poor - non poor; poor – poor)18.

Thus we can estimate a model as the following:

$$y_{t+1} = \beta X_t^H + \varepsilon$$

c.- Cross Section data

This method is mainly explained in the article.

7.4. Model specification

[18] This is the idea that Contreras et al. (2004) used to exploit a two-period panel data to analyze the vulnerability structure in Chile.

Given that we are estimating the model using 2 steps it is hard to find clear criteria to decide which model offer the best specification. We tested several model specifications and we compared the results in terms of coefficient significance level and other criteria that we describe here.

The next table summarizes the set of variables that each of them took into account.

Table 7.4. Model specifications

Vul M1	v(Y, R2)=f1(Charact, Assets, Transf, Context)
Vul M2	v(Y, R2)=f2(Charact, Assets, Transf)
Vul M3	v(Y, R2)=f3(Charact, Assets)
Vul M4	v(Y, R2)=f4(Charact)
Vul M5	Y=f3(Charact, Assets); R2=g(Comp, Assets, Transf, Context)

In terms of the household group classification the 5 models gave very similar results. That idea is supported by the next STATA tables.

a. Statistics summary of vulnerability measures using different model specifications.

Variable	Obs	Mean	Std.Dev	Min	Max
Vul M1	7096	0.14719	0.1901	0	0.99986
Vul M2	7094	0.1444	0.1848	2.26E-15	0.958248
Vul M3	7100	0.1464	0.18324	1.25E-12	0.994547
Vul M4	7101	0.1444	0.1708	1.21E-10	0.9682
Vul M5	7097	0.1427	1.707	1.10E-12	0.9996

b. Correlation table between different household estimation of vulnerability to poverty (7,090 obs)

Correlations	Vul M1	Vul M2	Vul M3	Vul M4	Vul M5
Vul M1	1				
Vul M2	0.9682	1			
Vul M3	0.9534	0.9827	1		
Vul M4	0.8965	0.9227	0.94	1	
Vul M5	0.91	0.9441	0.926	0.9579	1

As a result, all specifications models show a correlation of at least 90% which is really high.

7.5. Regression Coefficients for the more general model (Vul1)

Log_Per Capita Income	Indep. Variable	Coef.	Std. Err.	t
Assets	Head of HH Schooling	0.063132	0.00258	24.47
	No Sewerage (D)	-0.23035	0.03287	-7.01
	No Electricity (D)	0.029464	0.143271	0.21
	No Water (D)	-0.22574	0.110895	-2.04
	2nd House Ownership (D)	0.380379	0.002347	162.07
	Home Ownership	0.177543	0.018041	9.84
HH´s Characteristics	Dep.ratio	-0.05536	0.00561	-9.87
	Female (D)	-0.11552	0.02099	-5.5
	Head of HH Age	0.006347	0.000914	6.95
	Prop. Infants	-0.06436	0.071867	-0.9
	Prop.Adults	0.976717	0.052456	18.62
	Prop.Elderly	0.898332	0.069179	12.99
Labor	Self Employ (D)	0.332351	0.026038	12.76
	Gov Employ (D)	-0.20972	0.04745	-4.42
	Comp Employ (D)	-0.12164	0.030954	-3.93
	Housekeeper (D)	-0.04083	0.054663	-0.75
	Indefinite Contract (D)	0.042547	0.045082	0.94
	Definite Contract (D)	0.118464	0.022596	5.24
	Big Company (D)	0.131079	0.001617	81.05
Context	Unemp_rate	-1.28547	0.069105	-18.6
	Neighborhood Schooling	0.126933	0.005099	24.89
Transf	Gov.Transf (D)	-0.20258	0.017858	-11.34
Intercept	Constant	9.37439	0.07177	130.62
	N Obs	7,102		
	r2	0.5268		

Residuals2	Indep. Variable	Coef.	Std. Err.	t
Assets	Head of HH Schooling	0.0115	0.004691	2.47
	No Sewerage (D)	0.188446	0.059771	3.15
	No Electricity (D)	-0.17215	0.260526	-0.66
	No Water (D)	0.563815	0.201653	2.8
	2nd House Ownership (D)	0.181271	0.059604	3.04
	Home Ownership	-0.11339	0.032805	-3.46
HH´s Charact	Dep.ratio	-0.0118	0.010202	-1.16
	Female (D)	-0.04433	0.038169	-1.16
	Head of HH Age	-0.00561	0.001661	-3.38
	Prop. Infants	0.42917	0.130684	3.28
	Prop.Adults	0.11845	0.095386	1.24
	Prop.Elderly	-0.12501	0.125797	-0.99
Labor	Self Employ (D)	-0.55811	0.047347	-11.79
	Gov Employ (D)	-0.44884	0.086283	-5.2
	Comp Employ (D)	-0.49652	0.056286	-8.82
	Housekeeper (D)	-0.41954	0.099399	-4.22
	Indefinite Contract (D)	-0.15627	0.053235	-2.94
	Definite Contract (D)	-0.13864	0.081978	-1.69
	Big Company (D)	0.042138	0.041088	1.03
Context	Unemp_rate	0.002323	0.125661	0.02
	Neighborhood Schooling	0.025892	0.009272	2.79
Transf	Gov.Transf (D)	-0.16592	0.032474	-5.11
Intercept	Constant	1.013951	0.130508	7.77
	N Obs	7,102		
	r2	0.0553		

7.6. FGLS method of Amemiya (1977)

We summarize the procedure that Chaudhuri (2003) recommends in order to get asymptotically efficient estimates of $\beta \; and \; \theta$ known as three-step feasible generalized least square (FGLS). As we showed, we start assuming a stochastic process that generates income for a household H and also estimate its variance as we showed in equations (2) and (3):

First, we estimate residuals from (2) using OLS:

$$(9)\ \varepsilon^2_{OLS,H} = X_H\theta + \xi_H$$

Using the predicted values for (9), we can then transform the equation to:

$$(10)\ \frac{\hat{\varepsilon}^2_{OLS}}{X_H\hat{\theta}_{OLS}} = \left(\frac{X_H}{X\hat{\theta}_{OLS}}\right)\theta_{FGLS} + \frac{\xi_H}{X_H\hat{\theta}_{OLS}}$$

This transformed equation is estimated using OLS to obtain asymptotically efficient FGLS estimates. As Chaudhuri (2003) notes, $X\hat{\theta}_{FGLS}$ is a consistent estimate of $\sigma^2_{e,H}$ (the variance of the idiosyncratic component of household income). On the other hand, we can use the estimate $\sigma_H = \sqrt{X_H\hat{\theta}_{FGLS}}$ to then transform the first equation as follows:

$$(11)\ \frac{\ln y_H}{\hat{\sigma}_H} = \left(\frac{X_H}{\hat{\sigma}_H}\right)\beta_{FGLS} + \frac{\varepsilon_H}{\hat{\sigma}_H}$$

Thus, we can directly estimate (4) and (5).

A caveat: It should be mentioned that this procedure in context of measurement error in the dependent variable can lead to an overestimation of the variance of the log income. Besides, there is no guarantee that the estimate $\hat{\sigma}^2_{e,H}$ will be positive; however in none of the models specifications the number of negative estimate was significant for our purposes (around only 100 observations dropped).

Regarding the model specification there are other issues that can be discussed. Many of them may induce biased estimation on the coefficients of the model and as a result a later misidentification of the relevant groups for social policy purposes. First, as we usually do in observational studies we need to find a model specification that minimizes the risk of relevant omitted variable bias. Determinants of income coefficients (the first step in our model) can be biased if we are omitting a relevant variable that also covariates with some of the observed explanatory variables. A common omission in this relationship is the omission of the person's ability as an explanatory variable of income that could also covariates with the educational level with may induce to an overestimation of the education effect over incomes. What we could do in those cases is to find a proxy variable that could prevent as of the bias on our explanatory variables but a good proxy of that does not to seem available in our current data set.